PROCEEDINGS OF THE ROYAL SOCIETY B

Proc. R. Soc. B (2007) **274**, 809–817
doi:10.1098/rspb.2006.0281
Published online 19 December 2006

Stability of genetic polymorphism in host–parasite interactions

Aurélien Tellier* **and James K. M. Brown**

Department of Disease and Stress Biology, John Innes Centre, Colney, Norwich NR4 7UH, UK

Allelic diversity is common at host loci involved in parasite recognition, such as the major histocompatibility complex in vertebrates or gene-for-gene relationships in plants, and in corresponding loci encoding antigenic molecules in parasites. Diverse factors have been proposed in models to account for genetic polymorphism in host–parasite recognition. Here, a simple but general theory of host–parasite coevolution is developed. Coevolution implies the existence of indirect frequency-dependent selection (FDS), because natural selection on the host depends on the frequency of a parasite gene, and *vice versa*. It is shown that polymorphism can be maintained in both organisms only if there is negative, direct FDS, such that the strength of natural selection for the host resistance allele, the parasite virulence allele or both declines with increasing frequency of that allele itself. This condition may be fulfilled if the parasite has more than one generation in the same host individual, a feature which is common to most diseases. It is argued that the general theory encompasses almost all factors previously proposed to account for polymorphism at corresponding host and parasite loci, including those controlling gene-for-gene interactions.

Keywords: natural selection; host–parasite interactions; coevolution; frequency-dependent selection; victim–exploiter dynamical systems; gene-for-gene relationship

1. INTRODUCTION

Infectious disease limits reproductive fitness, while resistance to disease increases reproductive success and is thus a target for natural selection. In both plants and animals, defences are induced on recognition of parasite molecules, while parasites avoid detection by loss or mutation of those molecules (Dangl & Jones 2001). There is allelic diversity at parasite-recognition loci such as those controlling the major histocompatibility complex (MHC) in vertebrates (Apanius *et al.* 1997; Hill 2001) or gene-for-gene (GFG) relationships in plants (Thrall *et al.* 2001) and at parasite loci encoding proteins detected by the host (Thrall *et al.* 2001).

Current theories for the maintenance of polymorphism at these loci are diverse and complex, with a limited range of biological applicability, because they involve interactions between many genetic, epidemiological and ecological factors (Hughes & Nei 1992; Bergelson *et al.* 2001*a*). This has given rise to a view that complex interactions between many factors may be required for polymorphism (Bergelson *et al.* 2001*a*; Brown 2003*a*,*b*; De Meaux & Mitchell-Olds 2003).

This paper simplifies and generalizes the theory of coevolution of host–parasite specificities. We derive a simple, general condition for stability in a two-component system of an exploiter and a victim, such as a parasite and its host (§2). We then investigate its relevance to coevolution by analysing the GFG relationship, a paradigm for host–parasite interactions (Flor 1971; Thompson & Burdon 1992; Dangl & Jones 2001; Holub 2001). By showing that the classic GFG theory does not fulfil this condition, we demonstrate the need to include epidemiological factors in models of coevolution (§3). We then show that GFG interactions are stabilized by features of epidemiology which are common to almost all diseases of animals as well as plants (§4). Finally, we argue that features of genetics, epidemiology and ecology proposed to account for stable polymorphism in GFG models are special cases of the general condition (§5).

* Author for correspondence (aurelien.tellier@bbsrc.ac.uk).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rspb.2006.0281 or via http://www.journals.royalsoc.ac.uk.

2. GENERAL THEORY OF STABLE POLYMORPHISM IN COEVOLUTION

A condition for stable polymorphism in host–parasite interactions is derived from principles of linear algebra (Kot 2001). We focus on stability because it predicts situations in which polymorphism may be maintained in both species and thus be detectable. If a system is unstable, polymorphism will be lost in one or both species and therefore will not be detected.

A host–parasite interaction is a victim–exploiter (V–E) interaction, in which increased growth of the exploiter (E) reduces the growth rate of the victim (V) and increased availability of the victim increases growth of the exploiter, such that

$$\frac{\partial(\mathrm{d}V/\mathrm{d}t)}{\partial E} < 0 \quad \text{and} \quad \frac{\partial(\mathrm{d}E/\mathrm{d}t)}{\partial V} > 0. \tag{2.1}$$

Conditions for stability in this dynamical system are obtained by analysis of its Jacobian matrix. In continuous time, V and E evolve to stable values if

$$\frac{\partial}{\partial V}\left(\frac{\mathrm{d}V}{\mathrm{d}t}\right) + \frac{\partial}{\partial E}\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right) < 0. \tag{2.2}$$

(see section 1 of electronic supplementary material).

In host–parasite interactions, host resistance (H) and parasite infectivity (P) coevolve. For an interior equilibrium point to be stable, the rate of increase of H, P or

Received 19 September 2006
Accepted 20 November 2006

both must decline as their values increase (2.2). Organisms with seasonal reproduction can be regarded as existing in discrete time. In this case, host resistance at generation $t+1$ (H_{t+1}) is a function Ω_H of host resistance (H_t) and parasite infectivity (P_t) in the previous generation (Ω_P is the function for parasite infectivity), such that recursion equations are of the form

$$H_{t+1} = \Omega_H(H_t, P_t) \quad \text{and} \quad P_{t+1} = \Omega_P(H_t, P_t). \tag{2.3}$$

Here, an exact condition for stability of an interior equilibrium point is that the absolute value of both eigenvalues of the Jacobian matrix must be less than 1 (Jury condition; Kot 2001).The system can also be written as difference equations. The changes (Δ) in H and P between generations t and $t+1$ are functions ω_H and ω_P respectively,

$$\begin{aligned} \Delta H &= H_{t+1} - H_t = \omega_H(H_t, P_t) \quad \text{and} \\ \Delta P &= P_{t+1} - P_t = \omega_H(H_t, P_t). \end{aligned} \tag{2.4}$$

An exact condition for stability of an interior equilibrium of this dynamical system (2.4) is that the eigenvalues ($\lambda_1 = \alpha + i\beta$, $\lambda_2 = \alpha - i\beta$) of its Jacobian matrix ($\boldsymbol{J}$) must lie within a unit circle centred on $(-1, 0)$ in the complex plane (Roughgarden 1996)

$$-2 < \alpha < 0 \quad \text{with} \quad \alpha = \left(\frac{\partial \Delta H}{\partial H}\right) + \left(\frac{\partial \Delta P}{\partial P}\right) \quad \text{and} \quad -1 < \beta < 1. \tag{2.5}$$

A necessary but not sufficient condition for stable equilibrium is therefore that the rate of increase of H, P or both must decline as their values increase (2.5), a more stringent condition than that in continuous time (2.2).

Conditions (2.2) and (2.5) can be written in the same form in terms of host susceptibility or parasite non-infectivity. H may also represent quantitative variation in, for example, the strength of resistance, the frequency of a single resistance gene or the numbers of host individuals (likewise P; Nuismer & Otto 2005).

3. COEVOLUTION AND FREQUENCY-DEPENDENT SELECTION

We relate the general condition for stability (equations (2.2) and (2.5)) to the GFG relationship in plants, on which much of the theory of host–parasite coevolution is based (Leonard 1977; Frank 1993; Jeger 1997; Damgaard 1999; Sasaki 2000; Bergelson *et al.* 2001*a*; Thrall & Burdon 2002; Salathe *et al.* 2005; Segarra 2005). Resistance is induced if the plant has a resistance (*RES*) gene enabling recognition of specific parasite avirulence (AVR) protein. The parasite is not detected by the host and resistance is not induced if the host has a susceptibility allele (*res*) or the parasite has a virulence allele (*avr*).

The asymmetry of the GFG interaction suggests the possibility of an 'arms race', as successive pairs of *RES* and *AVR* alleles are driven to fixation (Holub 2001). In nature, however, there is substantial polymorphism at *RES* and *AVR* loci (Stahl *et al.* 1999; Thrall *et al.* 2001; Tian *et al.* 2002). Some polymorphisms are ancient, which are not consistent with the arms race model. It has been proposed that constitutive fitness costs of *avr* and *RES* alleles may account for GFG polymorphism (Leonard 1977) in a 'trench warfare' model of quasi-stable polymorphism (Stahl *et al.* 1999). Some experiments have detected such costs (Vera Cruz *et al.* 2000; Thrall & Burdon 2003; Tian *et al.* 2003) but others have not (Bergelson & Purrington 1996; Vera Cruz *et al.* 2000; Brown 2003*a*). We show here, however, that these costs are not sufficient to maintain GFG polymorphism.

We first investigate a simple GFG model (model A; figure 1*a*). This is essentially the classic model of GFG coevolution (Leonard 1977), on which subsequent models have been based. Model A illustrates the requirement for two forms of frequency-dependent selection (FDS). One of these is indirect FDS, in which host allele frequencies affect those of the parasite and *vice versa* (2.1), while the other is direct FDS, in which an allele's fitness effect decreases as it becomes more common (equations (2.2) and (2.5)).

In model A, the plant host is annual with discrete generations and there is one parasite generation per plant generation. At the start of each generation, each plant is attacked by one parasite of a genotype in proportion to its frequency. One host *RES* gene matches one parasite *AVR* gene. Although many more loci may be involved in GFG interactions in nature, one-locus models allow general principles of coevolution to be elucidated. The frequencies of the *AVR* and *RES* alleles at the start of generation g are A_g and R_g, respectively. The frequency of *avr* is $a_g = 1 - A_g$ and that of *res* is $r_g = 1 - R_g$. Recursion equations are obtained from host and parasite fitnesses (table 1)

$$\begin{aligned} \frac{A_{g+1}}{a_{g+1}} &= \frac{A_g(1 - cR_g)}{a_g(1 - b)} \quad \text{and} \\ \frac{R_{g+1}}{r_{g+1}} &= \frac{R_g(1-u)(1 - s + sA_g)}{r_g(1-s)}. \end{aligned} \tag{3.1}$$

At equilibrium $(R_{g+1}, A_{g+1}) = (R_g, A_g) = (\hat{R}, \hat{A})$. In this and subsequent models, there are four trivial equilibria, where $(\hat{R}, \hat{A}) = (0, 0)$, $(0, 1)$, $(1, 0)$ or $(1, 1)$. There is also one non-trivial, interior equilibrium, where

$$\begin{aligned} \frac{\hat{A}}{\hat{a}} &= \frac{u(1-s)}{s-u}, \quad \Rightarrow \hat{A} = \frac{u(1-s)}{s(1-u)} \quad \text{and} \\ \frac{\hat{R}}{\hat{r}} &= \frac{b}{c-b} \Rightarrow \hat{R} = \frac{b}{c}. \end{aligned} \tag{3.2}$$

Hence, $\hat{R}$ is largely a function of the cost of virulence (b) as $c \approx 1$ (c: cost to pathogen of being unable to infect a *RES* plant) while $\hat{A}$ depends mainly on the cost of resistance (u) and the cost to a plant of being diseased (s) (Frank 1992; Leonard 1977). The dynamics of the system is determined by analysis of its Jacobian matrix, $\boldsymbol{J}_A$. A logit transformation simplifies considerably the analysis of the difference equations

$$f_A = \log\left(\frac{A}{a}\right) \quad \text{and} \quad f_R = \log\left(\frac{R}{r}\right). \tag{3.3}$$

When this transformation is applied, Q_A, the discriminant of the characteristic equation of $\boldsymbol{J}_A$, is

$$Q_A = \frac{-u(s-u)b(c-b)}{s(1-u)c(1-b)}. \tag{3.4}$$

(section 2 of electronic supplementary material). Making the reasonable assumptions that $s > u$, avirulent parasites rarely infect host plants ($c \approx 1$) and the cost of virulence is small ($c \gg b$), Q_A is negative. This implies that a resistance–virulence (R, a) plot is an anticlockwise rotation (figure 1*a*) because R increases when a is small (*RES* is selected when *avr* is rare) and a increases when R is large (*avr* is selected when *RES* is common).

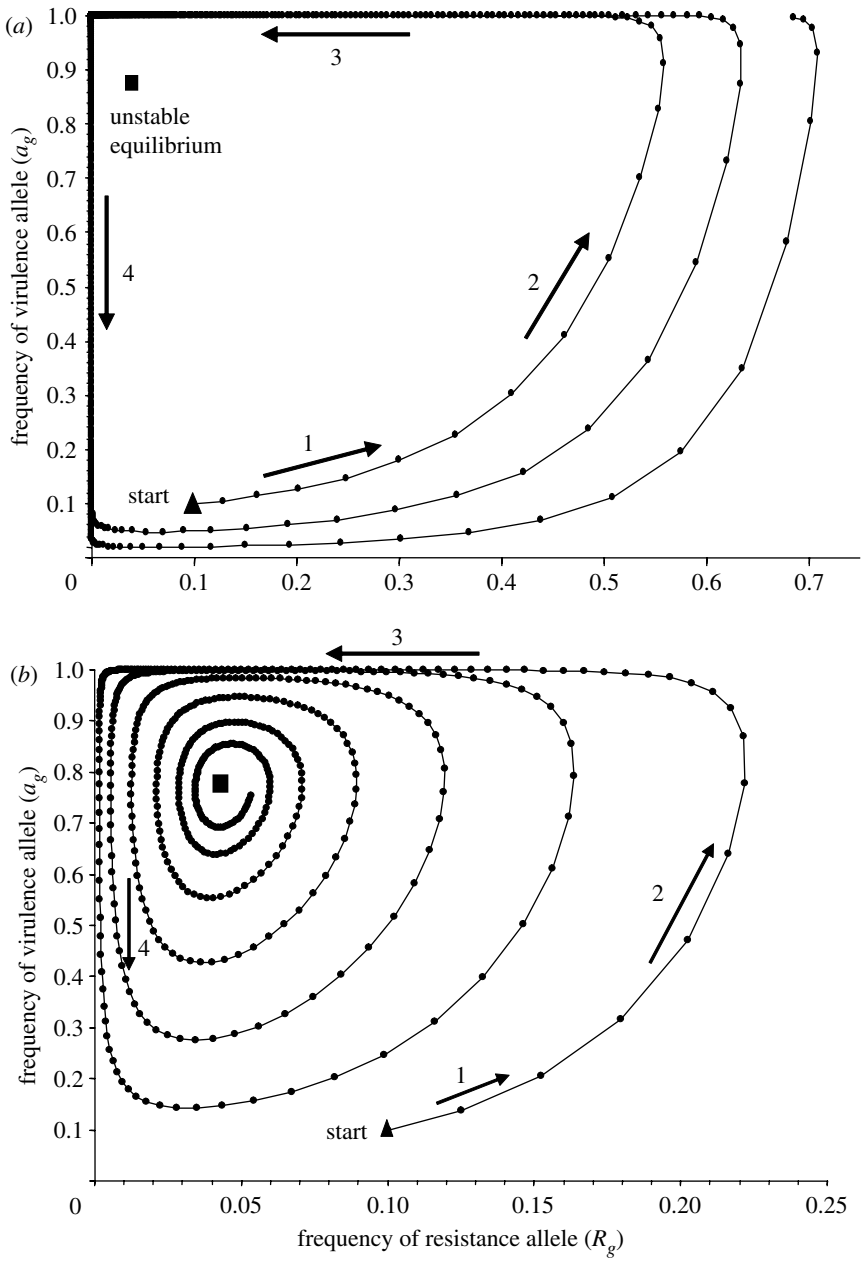

Figure 1. (*a*) Dynamics of allele frequencies in a model of single-locus gene-for-gene interactions with synchronous plant and parasite generations (model A). Avirulent parasites cannot cause disease on resistant plants ($c=1$). The cost to the plant of being diseased (s) is 0.3. The costs to the host of having the resistance (*RES*) allele (u) or the parasite, the virulence (*avr*) allele (b) are 0.05. The model was run for 1210 generations.If parasite avirulence (AVR) is common, the *RES* allele has a selective advantage because it protects the host against most parasites (arrow 1). As its frequency rises, it selects *avr* in the parasite (arrow 2). Constitutive costs then remove *RES* alleles which are no longer effective (arrow 3), then *avr* alleles which are no longer required to overcome host resistance (arrow 4). Gene frequencies spiral around and away from an unstable equilibrium point (equation (3.2)).(*b*) Dynamics of allele frequencies with two parasite generations per host generation (model B). $u=b=0.05$, as in figure 1*a*. The maximum cost of disease for plants infected by $G=2$ consecutive parasite generations is $\phi=0.3$ (with $z=1.4$ in equation (4.1)). Although the nature of the selection pressures is the same as in figure 1*a* (arrows 1–4), the allele frequencies spiral towards a stable equilibrium.

The interior, non-trivial equilibrium is unstable. Δf_R and Δf_A are constant over all values of f_R and f_A, respectively, $\partial\Delta f_R/\partial f_R=0$ and $\partial\Delta f_A/\partial f_A=0$). This implies that both diagonal elements of $\boldsymbol{J}_A$ are zero, so the eigenvalues of $\boldsymbol{J}_A$ have zero real parts. The imaginary parts are non-zero, however, implying that the condition for stable equilibrium in discrete time (equation (2.5)) is not fulfilled. Hence, stable polymorphism cannot be achieved in model A and the graph of allele frequencies spirals outwards from the interior equilibrium (figure 1*a*).

4. STABLE GENE-FOR-GENE POLYMORPHISM

The GFG system may be stable, however, if $\partial\Delta f_R/\partial f_R + \partial\Delta f_A/\partial f_A < 0$ (equation (2.5)), implying there must be direct FDS of *RES*, *avr* or both. Any factor which causes direct FDS, therefore, has the potential to allow polymorphism to be maintained. A distinctive feature of almost all parasites is that they have shorter life cycles than their hosts. This means that several parasite generations elapse per host generation and that the parasites may be dispersed between host individuals. The relative length of host and parasite life cycles has not

Table 1. Fitnesses of hosts and parasites in model A (parameters defined in §3).

host genotypes (frequencies)	parasite genotypes (frequencies)	fitness parasite	fitness host
RES (R_g)	*AVR* (A_g)	$1-c$	$1-u$
	avr (a_g)	$1-b$	$(1-u)(1-s)$
res (r_g)	*AVR* (A_g)	1	$1-s$
	avr (a_g)	$1-b$	$1-s$

been included in previous GFG models, but this basic feature of epidemiology is capable of stabilizing polymorphism.

(a) *Polycyclic disease*

In model B, the disease is polycyclic, with $G>1$ asexual generations of the parasite per host generation. In other respects, it is identical to model A. For simplicity, we assume that *AVR* parasites cannot infect *RES* plants ($c=1$). At the start of host generation g, each plant is attacked by one parasite of a genotype in proportion to its frequency (A_g for *AVR* and a_g for *avr*). $A_{g,1}$ and $a_{g,1}$ are the *AVR* and *avr* allele frequencies at the end of the first parasite generation (g, 1) during host generation g. In a polycyclic disease, the outcome of the second parasite generation (g, 2) depends on the result of parasite generation g, 1 (table 2). Three types of interaction can occur. First, *RES* plants attacked by *AVR* parasites in g, 1 can be attacked by any parasite genotype in g, 2, but *AVR* parasites still cannot do so successfully (table 2). Second, *RES* plants attacked by *avr* in g, 1 remain infected by *avr* in g, 2. Third, on *res* plants, new leaves produced after g, 1 are infected by the same parasite genotype that infected the older leaves. As a result, only the final parasite generation contributes to the population at the start of host generation $g+1$.

The loss of plant reproductive output caused by disease increases disproportionately with γ, the number of successful parasite generations ($\gamma \leq G$), because the parasite grows multiplicatively, corresponding damage is done to the host (Campbell & Madden 1990). Plant fitness (Y) is a decreasing function of γ

$$Y = 1-\phi\left(\frac{\gamma}{G}\right)^{z}, \tag{4.1}$$

where z defines the shape of the disease curve ($z>1$, section 3*a* of electronic supplementary material) and ϕ is the cost to a plant of being diseased by $\gamma=G$ parasite generations.

For simplicity, we assume that the parasite's reproductive fitness does not depend on γ (table 2). Equations (4.2)–(4.4) are obtained from the fitnesses in table 2 with $G=2$. For example, the *avr* population in g, 2 is the sum of:

(i) *avr* parasites infecting *RES* plants in g, 2 after an unsuccessful *AVR* attack in g, 1 ($R_g A_g a_{g,1}$),
(ii) *avr* parasites produced in g, 1 re-infecting the same *RES* plant in g, 2 ($R_g a_g$), and
(iii) *avr* parasites produced in g, 1 re-infecting the same *res* plant in g, 2 ($r_g a_g$)

$$\frac{A_{g,1}}{a_{g,1}} = \frac{A_g(1-R_g)}{a_g(1-b)}, \tag{4.2}$$

$$\frac{A_{g+1}}{a_{g+1}} = \frac{r_g A_g}{(1-b)\left[R_g(A_g a_{g,1}+a_g)+r_g a_g\right]}, \tag{4.3}$$

$$\frac{R_{g+1}}{r_{g+1}} = \frac{(1-u)R_g[A_g A_{g,1}+A_g a_{g,1}(1-\varepsilon)+a_g(1-\phi)]}{r_g(1-\phi)}, \tag{4.4}$$

where ε is the decrease of plant fitness after $\gamma=1$ ($\varepsilon=\phi(1/2)^z$: equation (4.1), $G=2$, $\gamma=1$). Note that equation (4.2) is identical to equation (3.1) in model A when $c=1$.

It was not possible to find exact solutions of the nonlinear system of equations (4.2)–(4.4) but approximate solutions were found by neglecting quadratic terms of the fitness penalties u and b, which are assumed to be small, in accordance with experimental data (Bergelson & Purrington 1996). At the interior equilibrium, virulence ($\hat{a}$) and resistance ($\hat{R}$) frequencies are approximately as follows:

$$\hat{a} = \frac{\varepsilon+\phi-\sqrt{(\varepsilon+\phi)^2-4\varepsilon(\phi-u)}}{2\varepsilon(1-u)}, \tag{4.5}$$

$$\hat{R} = \frac{b}{2-b-\hat{a}} = \frac{2b\varepsilon(1-u)}{\varepsilon(3-4u-2b(1-u))-\phi+\sqrt{(\varepsilon+\phi)^2-4\varepsilon(\phi-u)}}. \tag{4.6}$$

As in model A (equation (3.2)), $\hat{a}$ increases with increasing cost of being diseased (s in Model A, ε and ϕ here) and decreasing u but is not affected by b. $\hat{R}$ depends strongly on b, as in model A (equation (3.2)). We compared theoretical values of $\hat{a}$ and $\hat{R}$ (equations (4.5) and (4.6)) with the average values of a and R after 10 000 generations, when the system has stabilized. For realistic u and b ($<10\%$), equations (4.5) and (4.6) are indeed quite accurate (section 3*b* of electronic supplementary material).

Δf_R is constant for all values of f_R ($\partial\Delta f_R/\partial f_R=0$), while Δf_A is negatively correlated with Δf_A ($\partial\Delta f_A/\partial f_A<0$; Jacobian matrix $\boldsymbol{J}_B$ in section 3*b* of electronic supplementary material). Hence, the sum of the diagonal coefficients of $\boldsymbol{J}_B$ is negative, which is a necessary but not sufficient condition for stability (substituting H by R and P by A in equation (2.5)). In biological terms, direct FDS occurs ($\partial\Delta f_A/\partial f_A<0$) because *RES* plants infected by *avr* in g, 1 remain infected in g, 2. If the outcome of infection in g, 2 was independent of interactions in g, 1, *avr* parasites which infected the host in g, 1 might be replaced by *AVR* in g, 2 (Leonard & Czochor 1980; Kesinger *et al.* 2001). In model B, however, when a is high, most *RES* plants are *avr*-infected in g, 1 and remain so in g, 2. *avr* parasites therefore persist on *RES* plants and are not replaced by *AVR* parasites. In the case of independent infections, when A is high (and a is low), most *RES* plants infected by *avr* parasites in g, 1 encounter *AVR* in g, 2. In model B, however, these plants continue to support *avr* parasites. Hence the strength of selection for *avr* and

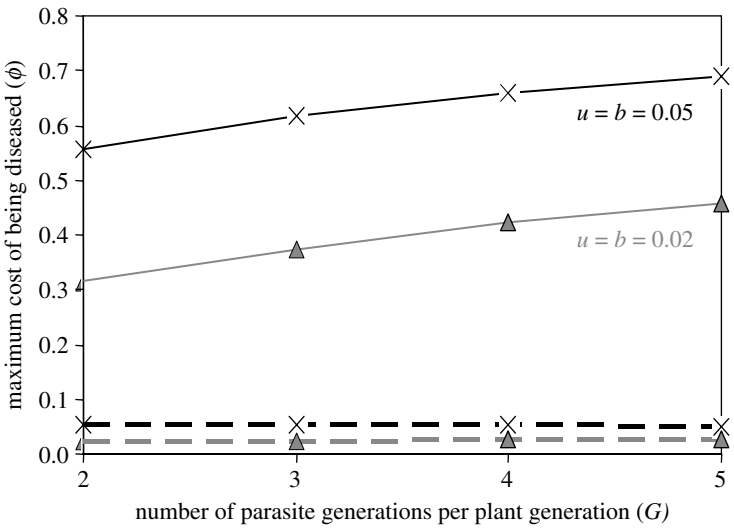

Figure 2. Outcomes of gene-for-gene coevolution in relation to the number of parasite generations per host generation (G) and the maximum cost of disease (ϕ). $z=1.4$. Constitutive costs of resistance (u) and virulence (b) are $u=b=0.05$ (black lines) or 0.02 (grey lines). Solid and dashed lines are the upper and lower limits of ϕ for coevolution to stable polymorphism, respectively. (i) When $\phi<u$ (dotted lines), virulent parasites and resistant plants are lost. The cost of disease is smaller than that of resistance, so resistance alleles are detrimental and virulence alleles are unnecessary. (ii) At intermediate ϕ, there is stable polymorphism, as allele frequencies spiral towards equilibrium as in figure 1*b*. (iii) For ϕ above an upper limit (solid lines), allele frequencies spiral away from the interior equilibrium, as in figure 1*a*, so avirulent parasites and susceptible plants are lost.

Table 2. Fitnesses of hosts and parasites in model B with two parasite generations ($G=2$) (parameters defined in §4).

	parasite genotypes (frequencies) within host generation g			fitness at beginning of host generation $g+1$	
	first generation	fitness of first parasite generation	second generation	fitness of second parasite generation	host fitness
host genotype *RES* (R_g)	*AVR* (A_g)	1	*AVR* ($A_{g,1}$)	0	$1-u$
			avr ($a_{g,1}$)	$1-b$	$(1-u)(1-\varepsilon)$
	avr (a_g)	$1-b$	(*AVR* negligible)	—	—
			avr (a_g)	$1-b$	$(1-u)(1-\phi)$
host genotype *res* (r_g)	*AVR* (A_g)	1	(*avr* negligible)	—	—
			AVR (A_g)	1	$1-\phi$
	avr (a_g)	$1-b$	(*AVR* negligible)	—	—
			avr (a_g)	$1-b$	$1-\phi$

against *AVR* is greater when A is higher, fulfilling one of the conditions for stability in equation (2.5).

A more convex shape of the disease curve, with higher z and smaller ε, increases selection for *RES* ($\partial\Delta f_R/\partial f_A$ increases) and diminishes the parameter space in which polymorphism is stable. With $G=2$, *res* plants are always infected twice and have a fitness reduction of ϕ (table 2). *RES* plants which encounter first *AVR*, then *avr*, only have a fitness reduction of ε ($\varepsilon<\phi$). Consequently, diminishing ε compared with ϕ increases the selection for *RES* (increasing $\partial\Delta f_R/\partial f_A$) and reduces the likelihood of stable polymorphism (section 3*c* of electronic supplementary material).

As coefficients of $\boldsymbol{J}_B$ could not be obtained analytically, we ran simulations for $2\leq G\leq 5$ (figure 2). An algorithm to construct equations equivalent to equations (4.3) and (4.4) for $G>2$ is described in section 3*c* of electronic supplementary material. Simulations were run with $0\leq\phi\leq 1$ for 20 000 host generations and various initial allele frequencies. The system was considered to be stable when the amplitude of the fluctuations of allele frequencies decreased over time and converged to equilibrium value $(\hat{R},\hat{a})$, calculated as the average frequencies over the last 1000 generations, for any of the initial allele frequencies tested (figure 1*b*). Allele frequencies close to $(\hat{R},\hat{a})$ were then perturbed to check that they would converge to $(\hat{R},\hat{a})$. If these criteria were not met, the system was considered to be unstable. The behaviour of the system did not depend on the initial frequencies, implying that there is no limit cycle (Leonard 1977; Kesinger *et al.* 2001).

There is a clear difference between stable and unstable behaviour in model B. Polymorphism is most likely to be stable at intermediate values of ϕ (figure 2; section 3*b* of electronic supplementary material). When ϕ is high, *RES* alleles are strongly selected, increasing $\partial\Delta f_R/\partial f_A$, preventing stable polymorphism and causing *RES* (and therefore *avr*) to become fixed (top of figure 2). At low ϕ ($\phi<u$), the cost of resistance outweighs its benefit so *res* and *AVR* alleles are fixed. As G increases, the system is stable over a broader range of ϕ (figure 2) because *avr* parasites are favoured, $\partial\Delta f_R/\partial f_A$ decreases and $\partial\Delta f_A/\partial f_R$ becomes larger

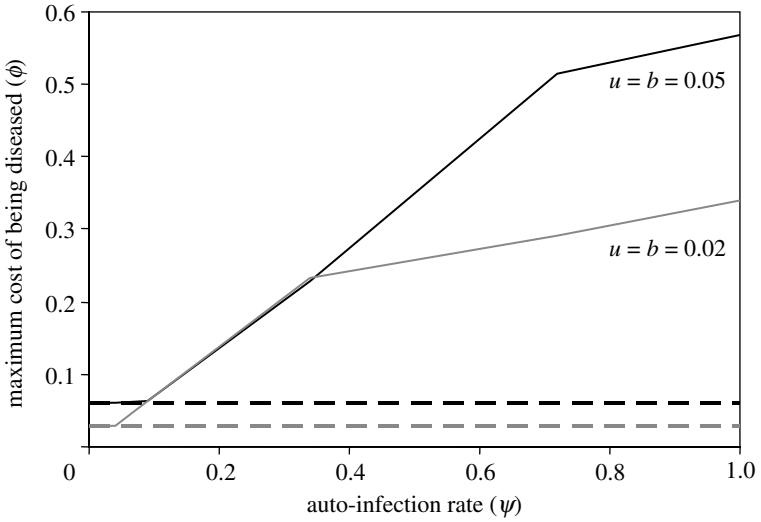

Figure 3. Outcomes of gene-for-gene coevolution for polycyclic parasites ($G=2$) in relation to the frequency of autoinfection (ψ) and the maximum cost of disease (ϕ) for $z=1.4$, with constitutive costs of resistance (u) and virulence (b) of $u=b=0.05$ (black lines) or 0.02 (grey lines).(i) When $\phi<u$ (dotted lines), there is a net cost of resistance so plant *res* alleles and parasite *AVR* alleles are fixed. (ii) At intermediate ϕ and intermediate to high ψ, polymorphism is stable (as in figure 1*b*). (iii) With high ϕ and sufficiently low ψ (above solid lines), *RES* and *avr* are fixed (as in figure 1*a*).

Table 3. Fitness of hosts and parasites in model C with two parasite generations ($G=2$) and autoinfection (ψ) (other parameters defined in §4) (* Not applicable: there is no auto-infection because *AVR* parasites fail to infect *RES* plants.).

	parasite genotypes (frequencies) within host generation g			fitness at beginning of host generation $g+1$	
	first generation	autoinfection (ψ) allo infection ($1-\psi$)	second generation	fitness of second parasite infection	host fitness
host genotype *RES* (R_g)	*AVR* (A_g)	n/a^*	*AVR* ($A_{g,1}$)	0	$1-u$
			avr ($a_{g,1}$)	$1-b$	$(1-u)(1-\varepsilon)$
	avr (a_g)	ψ	*avr* (a_g)	$1-b$	$(1-u)(1-\phi)$
		$1-\psi$	*avr* ($a_{g,1}$)	$1-b$	$(1-u)(1-\phi)$
		$1-\psi$	*AVR* ($A_{g,1}$)	0	$(1-u)(1-\varepsilon)$
host genotype *res* (r_g)	*AVR* (A_g)	ψ	*AVR* (A_g)	1	$1-\phi$
		$1-\psi$	*AVR* ($A_{g,1}$)	1	$1-\phi$
		$1-\psi$	*avr* ($a_{g,1}$)	$1-b$	$1-\phi$
	avr (a_g)	ψ	*avr* (a_g)	$1-b$	$1-\phi$
		$1-\psi$	*avr* ($a_{g,1}$)	$1-b$	$1-\phi$
		$1-\psi$	*AVR* ($A_{g,1}$)	1	$1-\phi$

(more negative), other parameters being equal, so the discriminant (Q_B) of $\boldsymbol{J}_B$ has a larger negative value. The range of parameter values at which polymorphism is stable is wider with higher constitutive costs (compare $u=b=0.05$ with $u=b=0.02$ in figure 2). This is consistent with predictions that costs of *RES* and *avr* help to maintain polymorphism (Leonard 1977; Bergelson *et al.* 2001*a*). For example, increasing u reduces selection for *RES* when f_A is high and enhances selection against *RES* when f_A is low. However, it does not imply that costs of *RES* and *avr* are sufficient to maintain polymorphism.

(b) *Model with parasite dispersal*

The second epidemiological factor we consider is the dispersal of parasite propagules between hosts within a host generation, a widespread characteristic of polycyclic diseases of plants and animals. We consider individual parasites within a host as independent units of infection and disease transmission. As a plant grows, each new leaf may be infected by a spore produced either on the same plant or on another plant (autoinfection and alloinfection, respectively, Barrett 1980). We show that higher rates of autoinfection are more likely to lead to a stable host–parasite interaction.

On *res* plants, the probabilities of autoinfection and alloinfection are ψ and $1-\psi$ respectively (section 4 of electronic supplementary material). Host and parasite fitnesses are given in table 3. On *RES* plants, autoinfection is only possible by *avr* parasites. Exact analysis of this model (C) was not possible, so we applied approximate analysis and simulations to investigate the effects of ψ and ϕ (figure 3). An approximate value of the interior equilibrium and analysis of the Jacobian matrix ($\boldsymbol{J}_C$) are in section 4 of electronic supplementary material. Note that model B with $G=2$ is a special case of model C, because table 3 reduces to table 2 if $\psi=1$.

In model C, polymorphism is stable over a wider range of other parameters when ψ is higher (figure 3). As in model B, $\partial\Delta f_R/\partial f_R=0$ but $\partial\Delta f_A/\partial f_A<0$. This can be explained biologically as follows. In model C, increasing

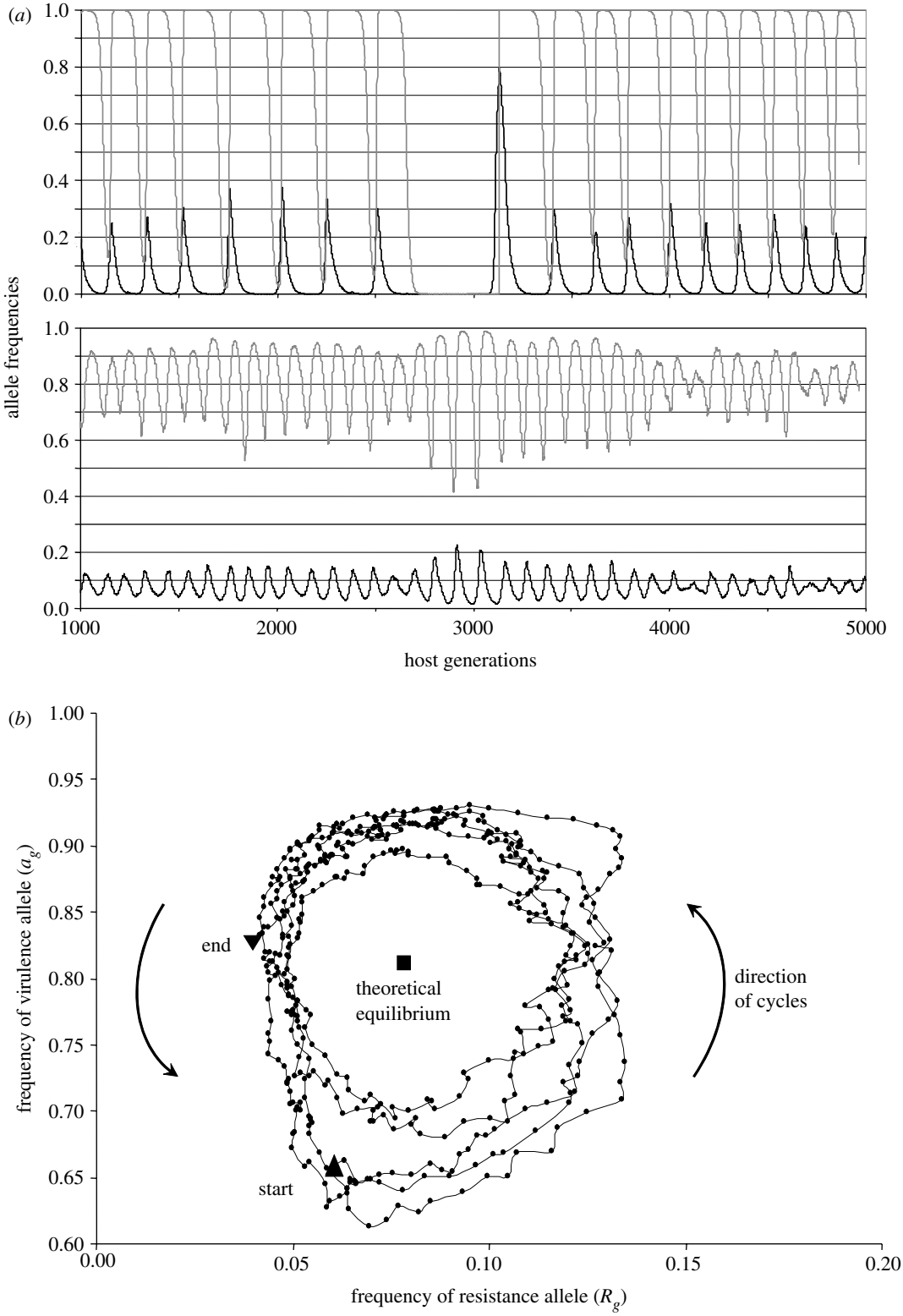

Figure 4. Dynamics of virulence (grey line) and resistance (black line) alleles for polycyclic parasites ($G=2$) in a finite population with genetic drift. $u=b=0.05$, $\phi=0.3$, $z=1.4$, $m=0.05$, population size $=1000$. (*a*) Top: unstable cycling ($\psi=0.1$) and bottom: stable cycling ($\psi=0.95$). (*b*) Trajectory of allele frequencies for five cycles showing irregular, anticlockwise cycling around the stable equilibrium ($\psi=0.95$).

alloinfection (decreasing ψ) tends to make successive parasite generations on the same plant independent of one another, causing selection against *avr* to become independent of its frequency ($\partial\Delta f_A/\partial f_A=0$ when $\psi=0$, section 4 of electronic supplementary material). When ψ is small and A is high, most *RES* plants infected by *avr* parasites in generation g, x encounter an *AVR* parasite in g, $x+1$. Increasing ψ, however, increases the probability of these plants remaining *avr*-infected in g, $x+1$. Hence for higher A and increasing ψ, selection for *avr* and against *AVR* becomes stronger ($\partial\Delta f_A/\partial f_A$ is more negative).

In polycyclic diseases, therefore, the stability of GFG systems depends on the outcome of infection in g, 1 influencing the course of g, 2. Autoinfection of *RES* plants infected by *avr* parasites is required for direct FDS. This is realistic in the GFG situation because the probability of

successful infection of a *RES* plant by an *avr* parasite is very much greater than that for an *AVR* parasite. The absence of autoinfection is implicit in polycyclic models which exhibit unstable behaviour, similar to that of model C with $\psi=0$ (Leonard & Czochor 1980; Kesinger *et al.* 2001).

(c) *Coevolution and genetic drift*

Model C2 is a stochastic version of model C, developed to investigate a realistic situation with finite, variable host and parasite population sizes (here, both $N_{\max}=1000$). Parasites have two generations per host generation; there is autoinfection and alloinfection (based on table 3 and eqns S12–S14 in section 4 of electronic supplementary material) and plant fitness loss function as in equation (4.1). Mutations occur between *RES* and *res* (and between *avr* and *AVR*) alleles with probability 10^{-5} per genome per generation. In host generation g, a random number of individuals of each genotype is added to or removed from the populations, the maximum change being a fraction $\pm$ m (here, $m=0.05$). For instance, the number of *RES* plants added in generation $g+1$ is

$$\Delta R_g = N_{\max} R_g m\theta,$$

where θ is a random number from a uniform distribution between -1 and 1.

The results of model C2 were largely identical to those of the deterministic model C, as stable polymorphism in both species was only achieved if disease was polycyclic with a high proportion of autoinfection. Unstable behaviour (top figure 4*a*), which occurs when ψ is low (here, 0.1), typically shows recurrent fixation of *avr* and *RES* alleles as their trajectories spiral towards the boundaries. When there is mainly autoinfection ($\psi=0.95$), the system is quasi-stable (bottom figure 4*a* and figure 4*b*). While allele frequencies spiral towards equilibrium, stochastic events nudge them away. This results in cycling around the theoretical equilibrium, particularly when genetic drift is limited ($m<0.1$; figure 4*b*). Higher values of genetic drift ($m>0.1$) lead to increased stochasticity and higher probability of allele fixation (not shown).

5. OTHER GENE-FOR-GENE MODELS

Many factors have been proposed to contribute to maintain polymorphism in victim–exploiter systems such as host–parasite interactions and GFG systems in particular (Leonard 1977; Frank 1993; Jeger 1997; Damgaard 1999; Sasaki 2000; Bergelson *et al.* 2001*a*; Sasaki *et al.* 2002; Thrall & Burdon 2002; Brown 2003*b*; Salathe *et al.* 2005; Segarra 2005) to the extent that it has been proposed that a complex of many interacting factors is essential to maintain GFG polymorphism (Thrall & Burdon 2002; Brown 2003*a*,*b*; De Meaux & Mitchell-Olds 2003). On the contrary, the analysis presented here unifies different selection pressures in a simple but comprehensive theoretical framework. Host–parasite interactions imply the existence of indirect FDS, such that host gene frequencies depend on parasite gene frequencies and *vice versa*. A condition for stable polymorphism, however, is that there must also be direct FDS, such that the fitness effects of *RES* or *avr* alleles or both are negatively dependent on their own frequencies.

We argue that most of the many factors which have been proposed to account for GFG polymorphism can be viewed as special cases of this general theory. Detailed analysis of these factors, however, is beyond the scope of this paper. These factors fall into two general classes. One class generates FDS and therefore may maintain stable polymorphism, following the 'trench warfare' concept of (Stahl *et al.* 1999). Several models incorporate differences between the life cycles of a host and a parasite (Leonard 1977), including perennial plant growth habit (Jeger 1997); seed banks (Damgaard 1999); parasites dispersing in a spatially structured host population (Thrall & Burdon 2002); and the parasite completing several generations during a single host generation (§4). If this causes parasite numbers to increase, with corresponding damage to the host, we predict that $\partial\Delta f_A/\partial f_A<0$, by analogy with models B and C. In the metapopulation model of (Damgaard 1999), increasing pressure of a GFG disease increases susceptibility to a second, non-specific disease. This causes stronger pressure from the two diseases combined where R is high, so once again, $\partial\Delta f_R/\partial f_R<0$. Note that several models include more than one factor which we predict to lead to stability (Damgaard 1999; Thrall & Burdon 2002).

A second class of factor may increase the expectation or variance of the lifetime of a *RES* or *AVR* allele, leading to transient polymorphism and successive replacement of *RES–AVR* allele pairs in an 'arms race' (Bergelson *et al.* 2001*b*; Holub 2001). Genetic drift and high mutation rates generate irregular fluctuations of R and a in time (Kirby & Burdon 1997) or space (Sasaki *et al.* 2002) when these stochastic processes are important compared with selection. When they are less significant, however, host and parasite alleles become fixed, with an 'arms race' of successive alleles. This transient polymorphism differs from irregular cycling around an interior equilibrium (§4*c* and Stahl *et al.* 1999; Tian *et al.* 2002) because the latter process leads to stable polymorphism in large populations but the former does not.

6. CONCLUSION

The model presented here provides a general explanation for stable or quasi-stable polymorphism in a coevolving host and parasite with realistically small constitutive costs of resistance and virulence (§2). It generates testable hypotheses about situations in which polymorphism is most likely to be stable and therefore detectable. Polymorphism is predicted to be the most prevalent in strongly polycyclic diseases (model B, figure 2) with high autoinfection (model C, figure 3). These are distinctive features of the powdery mildew, downy mildew and rust diseases of plants, in which GFG interactions are especially well known. 'Trench warfare' processes, however, do not preclude 'arms race' processes involving several loci (Holub 2001), which may contribute to the establishment of new *RES* and *avr* alleles.

The role of multiple GFG loci (Frank 1993; Sasaki 2000; Thrall & Burdon 2002; Salathe *et al.* 2005; Segarra 2005) requires further investigation in the light of the theory presented here. For example, does the existence of multiple loci further stabilize a system in which

epidemiological and ecological factors generate FDS, or does it further increase the variance of the lifetime of transiently polymorphic alleles?

The conditions for stable polymorphism in a coevolving host and parasite system (equation (2.5)) are generally applicable to long-term stability of any victim–exploiter system (equation (2.2)). These include prey–predator (Roughgarden 1996; Kot 2001) and host–parasite systems (Apanius *et al*. 1997; Hill 2001), as well as genetic interactions between hosts and parasites other than the GFG relationship (Frank 1992).

We thank anonymous reviewers for valuable recommendations for improving the manuscript. This research was supported in part by the Biotechnology and Biological Sciences Research Council.

REFERENCES

Apanius, V., Penn, D., Slev, P. R., Ruff, L. R. & Potts, W. K. 1997 The nature of selection on the major histocompatibility complex. *Crit. Rev. Immunol.* **17**, 179–224.

Barrett, J. A. 1980 Pathogen evolution in multilines and variety mixtures. *Z. Pflanzenk. Pflanzens* **87**, 383–396.

Bergelson, J. & Purrington, C. B. 1996 Surveying patterns in the cost of resistance in plants. *Am. Nat.* **148**, 536–558. (doi:10.1086/285938)

Bergelson, J., Dwyer, G. & Emerson, J. J. 2001*a* Models and data on plant–enemy coevolution. *Annu. Rev. Genet.* **35**, 469–499. (doi:10.1146/annurev.genet.35.102401.090954)

Bergelson, J., Kreitman, M., Stahl, E. A. & Tian, D. C. 2001*b* Evolutionary dynamics of plant R-genes. *Science* **292**, 2281–2285. (doi:10.1126/science.1061337)

Brown, J. K. M. 2003*a* A cost of disease resistance: paradigm or peculiarity? *Trends Genet.* **19**, 667–671. (doi:10.1016/j.tig.2003.10.008)

Brown, J. K. M. 2003*b* Little else but parasites. *Science* **299**, 1680–1681. (doi:10.1126/science.1083033)

Campbell, C. L. & Madden, L. V. 1990 *Introduction to plant disease epidemiology*. New York, NY: Wiley.

Damgaard, C. 1999 Coevolution of a plant host–pathogen gene-for-gene system in a metapopulation model without cost of resistance or cost of virulence. *J. Theor. Biol.* **201**, 1–12. (doi:10.1006/jtbi.1999.1007)

Dangl, J. L. & Jones, J. D. G. 2001 Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833. (doi:10.1038/35081161)

De Meaux, J. & Mitchell-Olds, T. 2003 Evolution of plant resistance at the molecular level: ecological context of species interactions. *Heredity* **91**, 345–352. (doi:10.1038/sj.hdy.6800342)

Flor, H. H. 1971 Current status of gene-for-gene concept. *Annu. Rev. Phytopathol.* **9**, 275–275. (doi:10.1146/annurev.py.09.090171.001423)

Frank, S. A. 1992 Models of plant pathogen coevolution. *Trends Genet.* **8**, 213–219.

Frank, S. A. 1993 Coevolutionary genetics of plants and pathogens. *Evol. Ecol.* **7**, 45–75. (doi:10.1007/BF01237734)

Hill, A. V. S. 2001 The genomics and genetics of human infectious disease susceptibility. *Annu. Rev. Genomics Hum. Genet.* **2**, 373–400. (doi:10.1146/annurev.genom.2.1.373)

Holub, E. B. 2001 The arms race is ancient history in *Arabidopsis*, the wildflower. *Nat. Rev. Genet.* **2**, 516–527. (doi:10.1038/35080508)

Hughes, A. L. & Nei, M. 1992 Maintenance of MHC polymorphism. *Nature* **355**, 402–403. (doi:10.1038/355402b0)

Jeger, M. J. 1997 An epidemiological approach to modelling the dynamics of gene-for-gene interactions. In *The gene-for-gene relationship* (eds I. R. Crute, E. B. Holub & E. J. J. Burdon), pp. 191–209. Wallingford, UK: CAB International.

Kesinger, J. C., Allen, L. J. S. & Strauss, R. E. 2001 Discrete-time models for gene frequencies and population densities in plant pathosystems. *Nonlinear Anal. Theory Methods Appl.* **47**, 1489–1500. (doi:10.1016/S0362-546X(01)00284-X)

Kirby, G. C. & Burdon, J. J. 1997 Effects of mutation and random drift on Leonard's gene-for-gene coevolution model. *Phytopathology* **87**, 488–493.

Kot, M. 2001 *Elements of mathematical ecology*. Cambridge, UK: Cambridge University Press.

Leonard, K. J. 1977 Selection pressures and plant pathogens. *Ann. NY Acad. Sci. USA* **287**, 207–222.

Leonard, K. J. & Czochor, R. J. 1980 Theory of genetic interactions among populations of plants and their pathogens. *Annu. Rev. Phytopathol.* **18**, 237–258. (doi:10.1146/annurev.py.18.090180.001321)

Nuismer, S. L. & Otto, S. P. 2005 Host–parasite interactions and the evolution of gene expression. *PLoS Biol.* **3**, 1283–1288. (doi:10.1371/journal.pbio.0030203)

Roughgarden, J. 1996 *Theory of population genetics and evolutionary ecology: an introduction*. New York, NY: Prentice Hall.

Salathe, M., Scherer, A. & Bonhoeffer, S. 2005 Neutral drift and polymorphism in gene-for-gene systems. *Ecol. Lett.* **8**, 925–932. (doi:10.1111/j.1461-0248.2005.00794.x)

Sasaki, A. 2000 Host–parasite coevolution in a multilocus gene-for-gene system. *Proc. R. Soc. B* **267**, 2183–2188. (doi:10.1098/rspb.2000.1267)

Sasaki, A., Hamilton, W. D. & Ubeda, F. 2002 Clone mixtures and a pacemaker: new facets of Red-Queen theory and ecology. *Proc. R. Soc. B* **269**, 761–772. (doi:10.1098/rspb.2001.1837)

Segarra, J. 2005 Stable polymorphisms in a two-locus gene-for-gene system. *Phytopathology* **95**, 728–736.

Stahl, E. A., Dwyer, G., Mauricio, R., Kreitman, M. & Bergelson, J. 1999 Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**, 667–671. (doi:10.1038/23260)

Thompson, J. N. & Burdon, J. J. 1992 Gene-for-gene coevolution between plants and parasites. *Nature* **360**, 121–125. (doi:10.1038/360121a0)

Thrall, P. H. & Burdon, J. J. 2002 Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal. *Plant Pathol.* **51**, 169–184. (doi:10.1046/j.1365-3059.2002.00683.x)

Thrall, P. H. & Burdon, J. J. 2003 Evolution of virulence in a plant host–pathogen metapopulation. *Science* **299**, 1735–1737. (doi:10.1126/science.1080070)

Thrall, P. H., Burdon, J. J. & Young, A. 2001 Variation in resistance and virulence among demes of a plant host–pathogen metapopulation. *J. Ecol.* **89**, 736–748.

Tian, D. C., Araki, H., Stahl, E., Bergelson, J. & Kreitman, M. 2002 Signature of balancing selection in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **99**, 11 525–11 530. (doi:10.1073/pnas.172203599)

Tian, D., Traw, M. B., Chen, J. Q., Kreitman, M. & Bergelson, J. 2003 Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74–77. (doi:10.1038/nature01588)

Vera Cruz, C. M., Bai, J. F., Ona, I., Leung, H., Nelson, R. J., Mew, T. W. & Leach, J. E. 2000 Predicting durability of a disease resistance gene based on an assessment of the fitness loss and epidemiological consequences of avirulence gene mutation. *Proc. Natl Acad. Sci. USA* **97**, 13 500–13 505. (doi:10.1073/pnas.250271997)