

Infection genetics: gene-for-gene versus matching-alleles models and all points in between

Aneil Agrawal* and Curtis M. Lively

Department of Biology, Indiana University, Bloomington, IN 47405-3700, USA

ABSTRACT

Here we address the controversy between plant pathologists and invertebrate zoologists as to the genetic basis of infection. We show that the alternative models proposed by these groups represent two ends of a continuum. Specifically, the gene-for-gene (GFG) model of plant pathologists represents one end of a continuum, where a very broad host range is expected to occur in one pathogen genotype. The matching alleles (MA) model, favoured by invertebrate zoologists, represents the opposite end of the same continuum, where an exact genetic match is required for infection. Since it is known that matching-alleles models give highly dynamical behaviour and selection for recombination if parasites exert strongly negative effects on host fitness, we were especially interested in determining the nature of any such behaviour in the interior regions of the MA–GFG continuum. We found that the highly dynamical aspects of matching-alleles models were observed across most of the continuum, and that recombination would increase geometric mean fitness over about half of the continuum.

Keywords: evolution of sex, gene-for-gene model, host–parasite co-evolution, matching-alleles model, recombination, Red Queen hypothesis.

INTRODUCTION

Two major models have been presented for the genetics underlying infection in plants and animals. The gene-for-gene model is based on data from plant–pathogen interactions, especially crop plants (Flor, 1956). The key feature of the model is that one parasite genotype has ‘universal virulence’, which means that it can infect all host genotypes (Fig. 1C). Under theory, a cost of virulence is required to keep this generalized genotype from going to fixation. In contrast, matching-alleles models are based on self/non-self recognition systems in invertebrates (e.g. Grosberg and Hart, 2000). The key feature of these models is that infection (or resistance) requires a specific match between the host and parasite (Fig. 1A). ‘Universal virulence’ is not possible in these models and polymorphism is easily maintained by negative frequency-dependent selection. Matching-alleles models underlie most of the theory constructed to understand the effects of host–parasite co-evolution on sex and recombination (e.g. Hamilton *et al.*, 1990; Howard and Lively, 1994). This theory has come to be known as the Red Queen theory.

* Author to whom all correspondence should be addressed. e-mail: aagraval@bio.indiana.edu
Consult the copyright statement on the inside front cover for non-commercial copying policies.

In a seminal paper, Parker (1994) pointed out that the use of matching-alleles models for the study of sexual reproduction may hamper the generality of the Red Queen theory for sex. If the dynamical aspects of these models (Red Queen dynamics) were dependent on the assumptions of matching-alleles models, then the Red Queen theory was precluded from including plants where the gene-for-gene model was already well established. Using a series of simulation models, Parker (1994) then showed that the gene-for-gene model would not select for sexual reproduction or recombination. Hence, if the gene-for-gene model were truly representative of the plant kingdom, then parasites would be unlikely to explain the existence of outcrossing in natural populations of a large proportion of species.

During the same period, Frank (1994) noted that the gene-for-gene model was largely inductive. By this he meant that the model was constructed to explain data for resistance/susceptibility in the 1950s, and that new data were interpreted as consistent with the model, without considering alternative genetic models. He showed that one form of the matching-alleles model could generate results that would be interpreted as support for the gene-for-gene model and argued that general acceptance of the gene-for-gene model was premature on theoretical grounds.

These different points of view have led to an interesting exchange (Frank, 1996a,b; Parker, 1996), which focused attention on the need to understand better the genetic basis of infection in plants and, in particular, animals. Here we take the view that the gene-for-gene and matching-alleles models are two ends of a continuum (following Parker, 1994). Our construction of this continuum is similar to Parker's (1994) model, except that the relevant parameters are combined in a multiplicative rather than additive way. We show that small departures from strict gene-for-gene assumptions can lead to Red Queen dynamics.

THE MODELS

Figure 1C shows the fitness of each (haploid) parasite genotype on each (haploid) host genotype under the classic gene-for-gene model (Flor, 1956; Parker, 1994). For parasites, there are two possible alleles at each of the two loci, an 'avirulent' allele and a 'virulent' allele. (These are the names traditionally associated with these alleles, although it would be more appropriate to call them narrowly infectious and widely infectious alleles, respectively. For the sake of tradition, we use the standard nomenclature of avirulent and virulent alleles.) For each parasite locus there is a corresponding host locus for which there are two possible alleles, a 'susceptible' allele and a 'resistant' allele. Hosts carrying the susceptible allele can potentially be infected by parasites carrying either allele, but hosts carrying the resistant allele can only be infected by parasites carrying the virulent allele. When multiple loci are involved in the interaction between hosts and parasites (in the present model there are two loci), a host can resist a parasite if the host has a resistant allele at any locus for which the parasite has an avirulent allele at the corresponding locus.

Everything else being equal, there is an advantage to parasites that carry virulent alleles, as these parasites are able to infect a broader spectrum of hosts than parasites carrying avirulent alleles. Similarly, there is an advantage to hosts that carry resistant alleles, as these hosts are able to resist infection from a broader array of parasites than hosts carrying susceptible alleles. Intrinsic costs of virulent and resistant alleles have been suggested to maintain variation in populations (e.g. Parker, 1992).

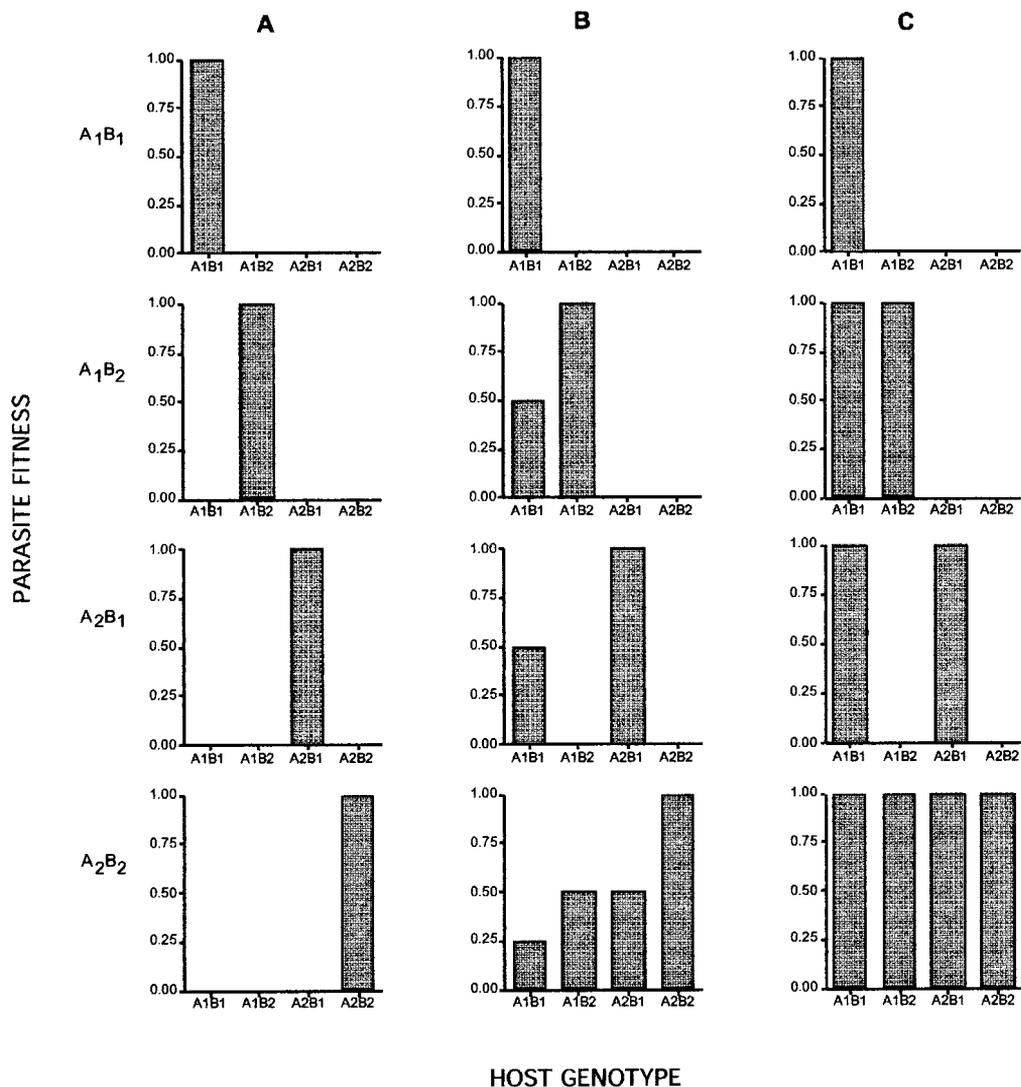


Fig. 1. Fitness of four pathogen genotypes on each host genotype under different models. (A) Pure matching alleles ($a = 0$). (B) Intermediate point on the MA–GFG continuum ($a = 0.5$). (C) Pure gene-for-gene ($a = 1$). Cost of virulence: $c = 0$ for A–C. Note that a pure gene-for-gene (GFG) system includes fitness bars that do not exist under the pure matching-alleles (MA) system. Intermediate points on the MA–GFG continuum ($0 < a < 1$) show partial fitness bars on these host genotypes.

In contrast, the rules of the matching-alleles model are relatively simple. A parasite must exactly match a host's genotype to elude the host's immune system and successfully infect the host. Thus, all parasite genotypes infect an equal number of host genotypes under the matching-alleles model.

From the perspective of the parasite, the difference between the two models is that, under the matching-alleles model, some parasite genotypes are not able to infect as many host

genotypes as they could under the gene-for-gene model (compare Figs 1A and 1C). An intermediate state between these two models exists when parasites can partially infect the remainder of the full set of susceptible host genotypes under the gene-for-gene model (Parker, 1994). We use the term ‘partial infection’ to indicate that the parasite is able to infect the host, but that the parasite is less successful than when it is able to cause a full infection. Hosts suffer less from parasites that can cause a partial infection than from those which are able to cause a full infection. Figure 1B shows an intermediate state between a matching-alleles and a gene-for-gene model.

Formally, we build this model by assigning a fitness for each parasite genotype on each host genotype as shown in Table 1. Variable a describes the extent to which the system works like a pure gene-for-gene system. By setting $a = 0$, the model becomes the matching-alleles model depicted in Fig. 1A. By setting $a = 1$, the model becomes the classic gene-for-gene model depicted in Fig. 1C. Intermediates between these two extremes occur when $0 < a < 1$. Figure 1B was produced for $a = 0.5$. Qualitative estimates of a could be gained by determining whether parasite strains infected a variable number of host genotypes and, where multiple hosts genotypes were infected, whether the severity of these infections varied among host genotypes. Intermediate values of a would be characterized as those cases for which the multiple host genotypes were both infected by the same parasite strain and showed variation in the severity of infection. For example, the results of Jarosz and Burdon (1996) suggest an intermediate value of a . They found that isolates of barley scald showed varying degrees of specialization. Some isolates were able to infect only a few families, but the average severity of infection was high; other isolates were able to infect many families, but the average severity of infection was low.

The variable k measures the intrinsic cost of a virulent allele. Because ‘virulent’ alleles only exist in the gene-for-gene model, this cost is always weighted by a . As a result, when the model is set to work as a pure matching-alleles model ($a = 0$), all terms involving cost disappear. When an individual carries virulent alleles at both loci, the cost of each virulent allele is multiplicative (Table 1).

The fitness of each host genotype when exposed to each parasite genotype is shown in Table 1. In general, the fitness of host i when infected by parasite j is

$$W_H(i,j) = (1 - ac)^z (1 - sW_{Pji})$$

where s is the maximum virulence of a parasite, c is the cost of carrying a resistance allele, z is the number of resistance alleles carried by the host (i.e. number of A_2 or B_2 alleles) and W_{Pji} is the fitness of parasite j on host i . When there is a cost for having a virulent allele ($k > 0$), W_{Pji} is less than unity and, as a result, the host experiences less than the maximum virulence (see Table 1). For hosts, there is a cost, c , associated with carrying ‘resistant’ alleles. This cost is always weighted by a so that costs only occur to the extent that the system works like a gene-for-gene system. Recall that costs are part of the gene-for-gene model but not part of the matching-alleles model. For hosts carrying resistant alleles at both loci, the cost of each resistant allele is multiplicative.

Although our model was inspired by that of Parker (1994), the parameterization is different. Only a single variable (our a) is required to move from a pure matching-alleles model to a pure gene-for-gene model. With this parameterization, we are able to investigate the effects of costs of virulence/resistance without being required to invoke such costs to move along the continuum.

Table 1. Fitness matrices: (A) the fitness of each of the j parasite genotypes when interacting with each of the i host genotypes; (B) the fitness of each of the i host genotypes when interacting with each of the j parasite genotypes

Parasite genotypes	Host genotypes			
	A ₁ B ₁ ($i = 1$)	A ₁ B ₂ ($i = 2$)	A ₂ B ₁ ($i = 3$)	A ₂ B ₂ ($i = 4$)
A. Parasite fitness: W_{pji}				
A ₁ B ₁ ($j = 1$)	1	0	0	0
A ₁ B ₂ ($j = 2$)	$a(1 - ak)$	$1 - ak$	0	0
A ₂ B ₁ ($j = 3$)	$a(1 - ak)$	0	$1 - ak$	0
A ₂ B ₂ ($j = 4$)	$a^2(1 - ak)^2$	$a(1 - ak)^2$	$a(1 - ak)^2$	$(1 - ak)^2$
B. Host fitness: $W_H(i, j) = (1 - ac)^z(1 - sW_{pji})$				
A ₁ B ₁ ($j = 1$)	$1 - s$	$(1 - ac)$	$(1 - ac)$	$(1 - ac)^2$
A ₁ B ₂ ($j = 2$)	$1 - sa(1 - ak)$	$(1 - ac)(1 - s(1 - ak))$	$(1 - ac)$	$(1 - ac)^2$
A ₂ B ₁ ($j = 3$)	$1 - sa(1 - ak)$	$(1 - ac)$	$(1 - ac)(1 - s(1 - ak))$	$(1 - ac)^2$
A ₂ B ₂ ($j = 4$)	$1 - sa^2(1 - ak)^2$	$(1 - ac)(1 - sa(1 - ak)^2)$	$(1 - ac)(1 - sa(1 - ak)^2)$	$(1 - ac)^2(1 - s(1 - ak)^2)$

Note: The variables are as follows: a = the extent to which the system works like a pure gene-for-gene system; c = the cost to a host of carrying a resistance allele (A₂ or B₂); k = the cost to a parasite of carrying a virulence allele (A₂ or B₂); z = the number of resistance alleles carried by a host ($z \in \{0, 1, 2\}$). Note that costs are multiplicative for both hosts and parasites.

Since our focus is on the genetic interface between hosts and parasites, we make several simplifying assumptions. We use a deterministic, mass-action model with no epidemiology or stochasticity. This process is analogous to assuming that each host encounters a single, randomly chosen parasite. We also assume random mating for both hosts and parasites.

The relative fitness of the i th host genotype is:

$$w_H(i) = \left(\frac{1}{\bar{W}_H} \right) \sum_{j=1}^{n_p} W_H(i, j)p(j)$$

where $W_H(i, j)$ is the fitness of the i th host genotype when it encounters the j th parasite genotype (as shown in Table 1), n_p is the number of parasite genotypes, and $p(j)$ is the frequency of the j th parasite genotype. The average fitness in the host population is

$$\bar{W}_H = \sum_{i=1}^{n_H} h(i) \sum_{j=1}^{n_p} W_H(i, j)p(j)$$

where $h(i)$ is the frequency of the i th host genotype and n_H is the number of host genotypes.

Similarly, the relative fitness of the j th parasite genotype is:

$$w_P(j) = \left(\frac{1}{\bar{W}_P} \right) \sum_{i=1}^{n_H} W_P(j, i)h(i)$$

where $W_P(j, i)$ is the fitness of the j th parasite genotype when it encounters the i th host genotype (as shown in Table 1). The average fitness in the parasite population is

$$\bar{W}_p = \sum_{j=1}^{n_p} p(j) \sum_{i=1}^{n_H} W_p(j, i) h(i)$$

All simulations were initiated with intermediate allele frequencies for both host and parasite. Initial genotype frequencies were generated from the allele frequencies assuming no linkage disequilibrium. For parasites, we included a low (bi-directional) mutation rate of $\mu = 10^{-6}$ for both loci.

Simulations were run for 4000 generations to allow the system to stabilize before taking measurements. To assess allele-frequency dynamics, the average variance in allele frequency across the two host loci was measured between generations 4000 and 5000. The geometric mean fitness of the hosts was also measured over this period. For each set of parameter values, the simulation was run with no recombination between host loci ($r = 0$) and with a low level of recombination between host loci ($r = 0.1$). Parasites were always assumed to have free recombination ($r = 0.5$).

RESULTS AND DISCUSSION

Our primary interest was in the allele-frequency dynamics along the matching-alleles/gene-for-gene (MA-GFG) continuum. Dynamics showing high amplitudes and short periods were of particular interest, because they are more likely to generate selection for sexual reproduction and high rates of recombination (Barton, 1995; Peters and Lively, 1999). Because the allele-frequency dynamics can be complex, it was not possible to summarize them with measures of amplitudes and periods. Instead, we evaluated the dynamics by plotting the variance in host allele frequency (averaged over the two loci) across the continuum for different costs of resistance and virulence (Fig. 2). The continuum was modelled as a linear function of the variable a , which allows the system to change continuously from one model to the other.

Consistent with Sasaki (2000), we found that pure gene-for-gene systems ($a = 1$) will cycle if costs exist for both host resistance (c) and parasite virulence (k). We also found that the variance in allele frequency can be nearly as large as for pure matching-alleles models depending on the magnitude of these costs (e.g. $c = 0.1$ and $k = 0.5$; see Fig. 2C). The cycle works as follows: When the universally virulent genotype is common, there is selection against host resistance alleles because they do not prevent infection, yet they have a cost. Thus the non-resistant host alleles increase in frequency. This change in the host population is followed by selection against the virulent alleles in the parasite, which have become unnecessary for successful infection, yet have a cost. This change is then followed by selection for the host resistance alleles, which are now effective in preventing infection by the avirulent parasite alleles. Finally, there is selection for the virulent alleles in the parasite population, thus completing the cycle.

The continuum – without costs

In a pure gene-for-gene model, the resistant alleles go to fixation and the system does not cycle when there is no cost to resistance (Fig. 2A). However, the system does cycle if assumption of pure gene-for-gene genetics is relaxed ($a < 1$). In fact, even slight move-

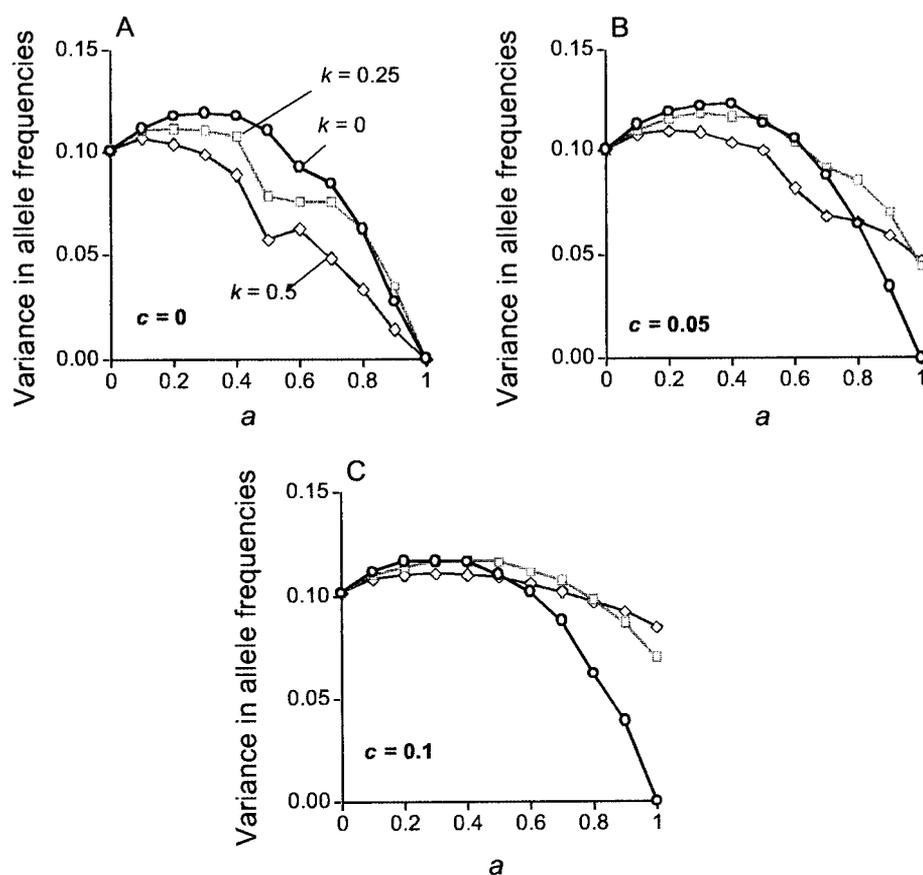


Fig. 2. The variance in host allele frequency as a function of the parameter a , the extent to which loci act as gene-for-gene loci, and the cost of virulence, k , in the parasite (\circ , $k=0$; \diamond , $k=0.25$; \square , $k=0.5$). (A) Cost of resistance in host, $c=0$. (B) Cost of resistance in host, $c=0.05$. (C) Cost of resistance in host, $c=0.10$. The parameter s (maximum virulence) was set to 0.5 for all runs. Note that the system cycles for pure gene-for-gene genetics ($a=1$) when the cost of resistance and virulence are both non-zero (see also Sasaki, 2000). Also note that the system cycles without costs for values of $a < 1$.

ment along the MA–GFG continuum results in cycling and very steep increases in the variance in allele frequencies (Fig. 2). Surprisingly, the system oscillated strongly in the absence of costs to either resistance or virulence when the genetic system was 50% gene-for-gene; in fact, the variance in allele frequencies at this point was slightly greater than that observed for the pure matching-alleles model ($a=0$; see Fig. 2A). More generally, moving along the continuum from pure gene-for-gene ($a=1$) to pure matching-alleles ($a=0$), we found that the amplitude of oscillations increases quickly, the period of oscillations decreases and the mean allele frequency becomes closer to 0.5. Hence, while experimental studies may suggest gene-for-gene genetics, the system could still have important matching-alleles properties (strong oscillatory dynamics) at the population level.

The continuum – with costs

Costs of virulence/resistance remain as an important empirical question. A few recent studies have demonstrated costs (Bergelson, 1994; Fineblum and Rausher, 1995; Berenbaum *et al.*, 1996; Kraaijeveld and Godfray, 1997; Mauricio, 1998), but many others have failed to detect such costs (Parker, 1992; Bergelson and Purrington, 1996), perhaps because they are difficult to detect in complex genetic systems (Frank, 2000). We have found that costs for both virulence and resistance reduce the difference in dynamical properties between pure gene-for-gene systems and pure matching-alleles systems. From a conceptual perspective, costs make a gene-for-gene system more similar to a matching-alleles system, because they prevent any one host (parasite) genotype from being the best against every parasite (host) genotype. However, the mechanisms are different. In matching-alleles systems, all the parasite genotypes are specialists, but they specialize on different host genotypes. In

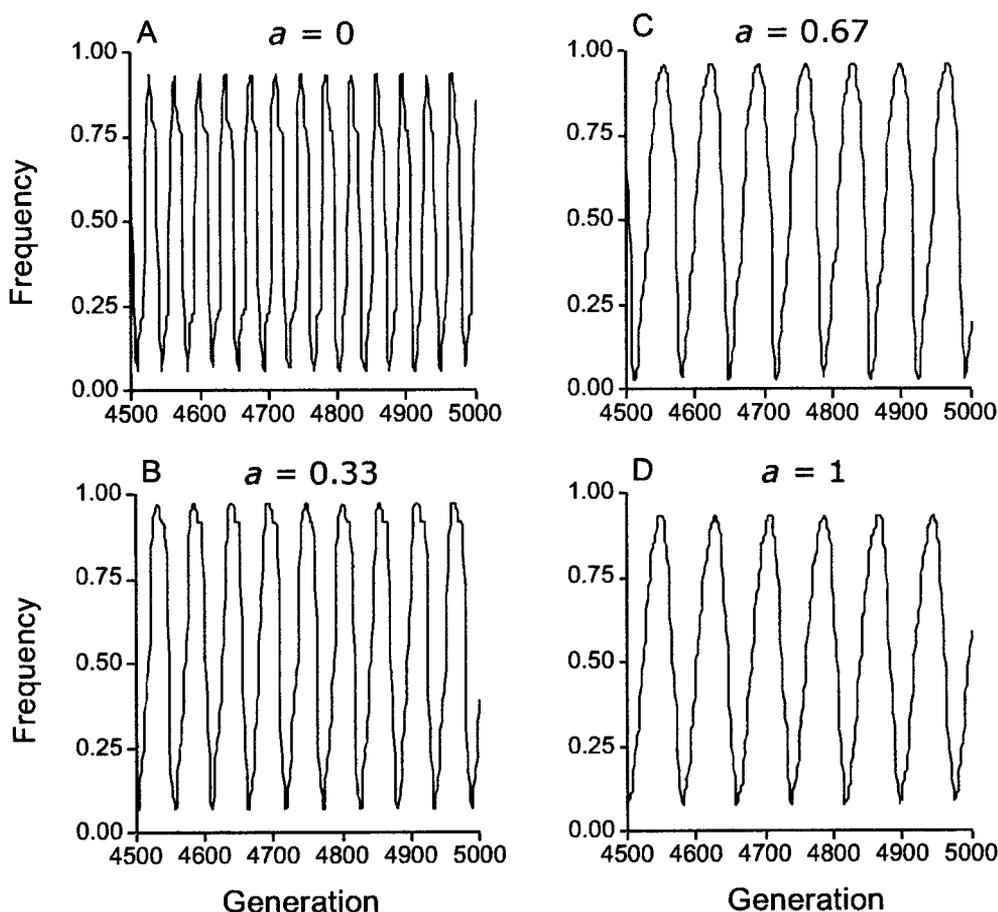


Fig. 3. Host allele-frequency dynamics (A, allele) for a host population with recombination ($r = 0.1$). Cost of resistance in host, $c = 0.10$; cost of virulence in parasite, $k = 0.5$. Four points along the MA–GFG continuum are shown: (A) $a = 0$; (B) $a = 0.33$; (C) $a = 0.67$; (D) $a = 1$. Note that the period of oscillations increases with larger values of a .

gene-for-gene systems, increasing the cost increases the trade-off between generalist (A_2B_2) and specialist (e.g. A_1B_1) parasite genotypes. Moving along the MA-GFG continuum gradually changes the system from being composed of pure specialists to one in which pure specialists co-occur with pure generalists. Our results show that the allele-frequency dynamics are highly oscillatory across most of this continuum (Fig. 2B,C).

Representative allele-frequency dynamics are shown in Fig. 3 (low recombination) and Fig. 4 (no recombination). As a increases, oscillations decline only slightly in amplitude, but increase in period more dramatically (approximately two-fold). Comparing Figs 3 and 4, we found that allele-frequency oscillations were more regular in recombining hosts

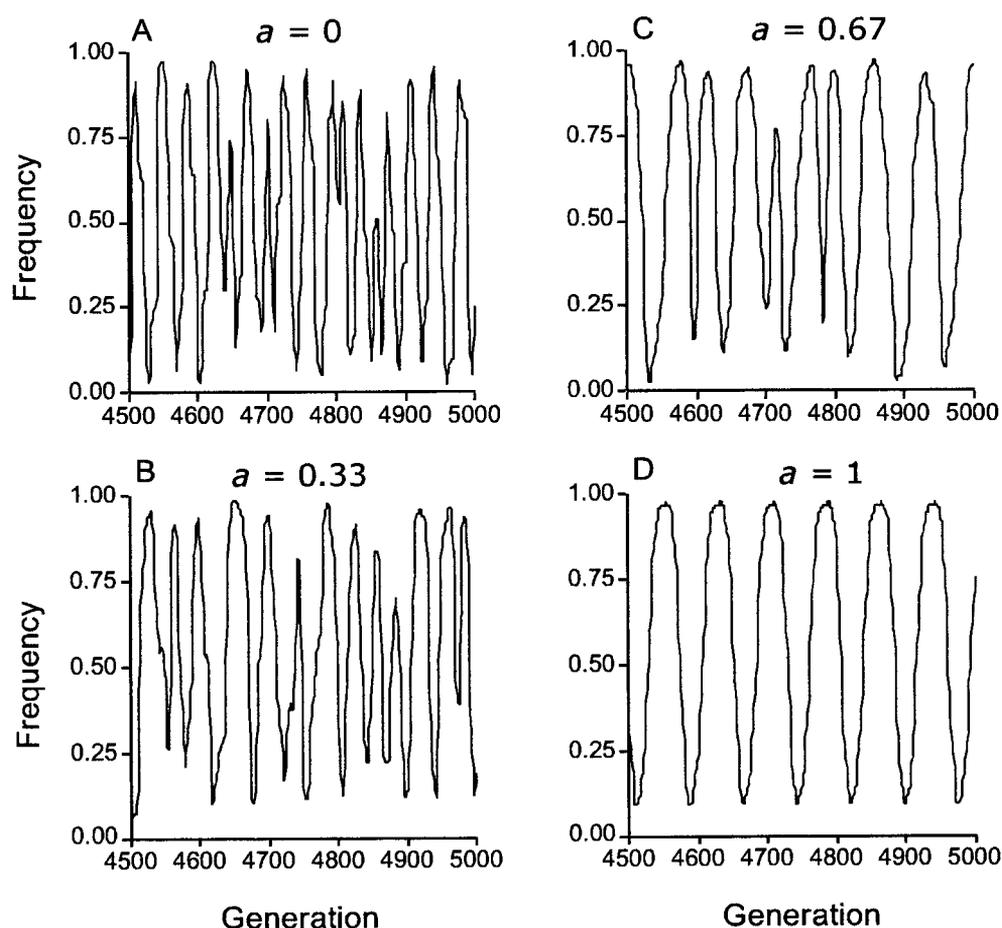


Fig. 4. Host allele-frequency dynamics (A_1 allele) for a host population without recombination ($r = 0$). Cost of resistance in host, $c = 0.10$; cost of virulence in parasite, $k = 0.5$. Four points along the MA-GFG continuum are shown: (A) $a = 0$; (B) $a = 0.33$; (C) $a = 0.67$; (D) $a = 1$. Note that the period of oscillations and the evenness of the oscillations increase with larger values of a . Also note that the oscillations are less even, in general, than those observed for the same parameter values when recombination is in the model (see Fig. 3).

than in non-recombining hosts, except for large values of a when the interaction was more like a gene-for-gene system.

Recombination and mean fitness

To determine the importance of recombination, we compared the geometric mean fitness of a host population with a low level of recombination ($r = 0.1$) to the geometric mean fitness of a host population with no recombination ($r = 0$). In Fig. 5, the difference in geometric mean fitness is plotted as a percentage of the recombining population's geometric mean fitness. Positive values indicate that the recombining host population had a higher geometric mean fitness than the non-recombining populations when co-evolution was simulated under the same parameters. In general, positive values are found in the matching-alleles half of

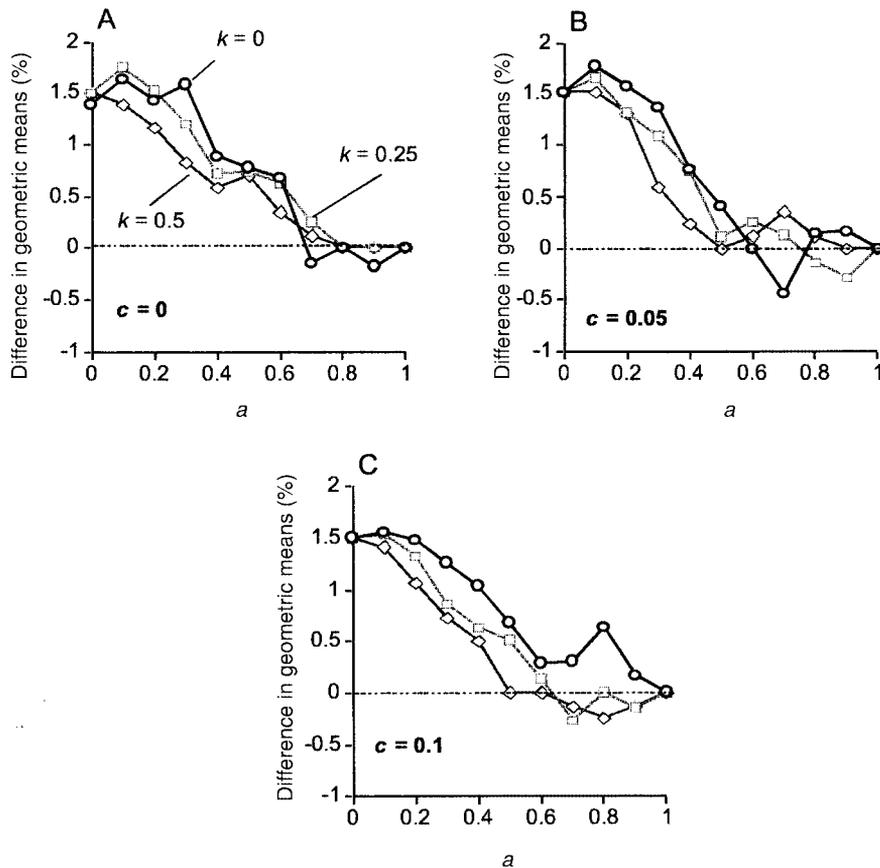


Fig. 5. Difference in geometric mean fitness between recombining ($r = 0.1$) and non-recombining hosts ($r = 0$) for different costs of virulence (\circ , $k = 0$; \diamond , $k = 0.25$; \square , $k = 0.5$). The difference is plotted as a percentage of the mean fitness of the recombining host. Positive values indicate that a recombining host has a higher geometric mean fitness than a non-recombining host when co-evolution is simulated under the same parameters. Positive values are generally found in the region of $a < 0.5$. (A) Cost of resistance in host, $c = 0$. (B) Cost of resistance in host, $c = 0.05$. (C) Cost of resistance in host, $c = 0.10$.

the continuum ($a < 0.5$; Fig. 5). In the remainder of the continuum, the differences in fitness tended to be smaller and both positive and negative values were observed. These results suggest that recombination will be selected across much of the continuum, but not in systems that behave much like pure gene-for-gene systems. The longer period of gene-for-gene-like systems may prevent parasites from favouring sex and recombination, as suggested by Parker (1994).

ACKNOWLEDGEMENTS

We thank S. Frank, M. Parker and P. Thrall for helpful comments on earlier versions of the manuscript. This work was supported by a US National Science Foundation grant to C.M.L. (DEB 9904840) and by a fellowship from the Natural Sciences and Engineering Research Council of Canada to A.F.A.

REFERENCES

- Barton, N.H. 1995. A general model for the evolution of recombination. *Genet. Res.*, **65**: 123–144.
- Berenbaum, M.R., Zangerl, A.R. and Nitao, J.K. 1996. Constraints on chemical coevolution: wild parsnips and the parsnip webworm. *Evolution*, **40**: 1215–1228.
- Bergelson, J. 1994. The effect of genotype and the environment on costs of resistance in lettuce. *Am. Nat.*, **143**: 349–359.
- Bergelson, J. and Purrington, C.B. 1996. Surveying patterns in the cost of resistance in plants. *Am. Nat.*, **148**: 536–558.
- Fineblum, W.L. and Rausher, M.D. 1995. Tradeoff between resistance and tolerance to herbivore damage in a morning glory. *Nature*, **377**: 517–520.
- Flor, H.H. 1956. The complementary genetic systems in flax and flax rust. *Adv. Genet.*, **8**: 29–54.
- Frank, S.A. 1994. Recognition polymorphism in host–parasite genetics. *Phil. Trans. R. Soc. Lond. B*, **346**: 283–293.
- Frank, S.A. 1996a. Problems inferring the specificity of plant–pathogen genetics. *Evol. Ecol.*, **10**: 323–325.
- Frank, S.A. 1996b. Statistical properties of polymorphism in host–parasite genetics. *Evol. Ecol.*, **10**: 307–317.
- Frank, S.A. 2000. Specific and non-specific defense against parasitic attack. *J. Theor. Biol.*, **202**: 283–304.
- Grosberg, R.K. and Hart, M.W. 2000. Mate selection and the evolution of highly polymorphic self/nonself recognition genes. *Science*, **289**: 2111–2114.
- Hamilton, W.D., Axelrod, R. and Tanese, R. 1990. Sexual reproduction as an adaptation to resist parasites: a review. *Proc. Natl. Acad. Sci. USA*, **87**: 3566–3573.
- Howard, R.S. and Lively, C.M. 1994. Parasitism, mutation accumulation and the maintenance of sex. *Nature*, **367**: 554–557.
- Jarosz, A.M. and Burdon, J.J. 1996. Resistance to barley scald (*Rhynchosporium secalis*) in wild barley grass (*Hordeum glaucum* and *Hordeum leporinum*) populations in south-eastern Australia. *Aust. J. Agric. Res.*, **47**: 413–425.
- Kraaijeveld, A.R. and Godfray, H.C.J. 1997. Trade-off between parasitoid resistance and larval competitive ability in *Drosophila melanogaster*. *Nature*, **389**: 278–280.
- Mauricio, R. 1998. Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *Am. Nat.*, **151**: 20–28.
- Parker, M.A. 1992. Disease and plant population genetic structure. In *Plant Resistance to Herbivores and Pathogens: Ecology, Evolution, and Genetics* (R.S. Fritz and E.L. Simms, eds), pp. 345–362. Chicago, IL: University of Chicago Press.

- Parker, M.A. 1994. Pathogens and sex in plants. *Evol. Ecol.*, **8**: 560–584.
- Parker, M.A. 1996. The nature of plant–parasite specificity. *Evol. Ecol.*, **10**: 319–322.
- Peters, A.D. and Lively, C.M. 1999. The Red Queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. *Am. Nat.*, **154**: 393–405.
- Sasaki, A. 2000. Host–parasite coevolution in a multilocus gene-for-gene system. *Proc. R. Soc. Lond. B*, **267**: 2183–2188.