vol. 185, no. 5 the american naturalist may 2015

E-Article

# Host-Parasite Coevolutionary Dynamics with Generalized Success/Failure Infection Genetics

**Jan Engelstädter***

School of Biological Sciences, University of Queensland, Brisbane, Queensland 4072, Australia

*Submitted March 31, 2014; Accepted September 16, 2014; Electronically published February 25, 2015*

*Online enhancement:* appendix.

abstract: Host-parasite infection genetics can be more complex than envisioned by classic models such as the gene-for-gene or matching-allele models. By means of a mathematical model, I investigate the coevolutionary dynamics arising from a large set of generalized models of infection genetics in which hosts are either fully resistant or fully susceptible to a parasite, depending on the genotype of both individuals. With a single diploid interaction locus in the hosts, many of the infection genetic models produce stable or neutrally stable genotype polymorphisms. However, only a few models, which are all different versions of the matching-allele model, lead to sustained cycles of genotype frequency fluctuations in both interacting species ("Red Queen" dynamics). By contrast, with two diploid interaction loci in the hosts, many infection genetics models that cannot be classified as one of the standard infection genetics models produce Red Queen dynamics. Sexual versus asexual reproduction and, in the former case, the rate of recombination between the interaction loci have a large impact on whether Red Queen dynamics arise from a given infection genetics model. This may have interesting but as yet unexplored implications with respect to the Red Queen hypothesis for the evolution of sex.

*Keywords*: matching allele, gene for gene, diploidy, epistasis, coevolution, Red Queen hypothesis.

## Introduction

Host-parasite interactions have long been recognized as central drivers of evolution. One reason for this is that, due to the inherent antagonism of the interaction, natural selection can produce ongoing coevolutionary arms races of attack and counterattack. Of particular interest are so-called Red Queen (RQ) dynamics, which involve continuing oscillations of both host and parasite genetic variants. RQ dynamics can arise because parasites are under selection to infect common hosts, and hosts are under selection to resist common parasites, so that often both rare host and rare parasite variants will be favored by natural selection. (This has been termed "time-delayed," "indirect," or "virtual" negative frequency dependence but is distinct from normal frequency-dependent selection in a single species.) RQ dynamics are interesting for a number of reasons: they may lead to rapid evolutionary change that is observable in nature and in coevolution experiments (Decaestecker et al. 2007; Brockhurst and Koskella 2013), maintain genetic variation in both host and parasite populations, and produce selection for sex and recombination (the RQ hypothesis for the evolution of sex; Jaenike 1978; Hamilton 1980; Salathé et al. 2008).

The emergence of RQ dynamics is determined to a large extent by the infection genetics underlying a host-parasite system. Most previous models of host-parasite coevolution have considered one of a few "standard" models for infection genetics, in particular the gene-for-gene (GFG) model and the matching-allele (MA) model. The classic GFG model assumes that hosts carry either susceptibility or resistance alleles at one or several loci, whereas parasites carry either "avirulence" or "virulence" alleles (Flor 1955). A parasite can infect a host unless the host carries at least one resistance allele that matches a corresponding avirulence allele in the parasite. The GFG model has much empirical support in plant-pathogen systems and has been widely studied theoretically (Thompson and Burdon 1992; Brown and Tellier 2011). The main feature of the coevolutionary dynamics predicted to arise from this model is that, in the absence of other factors, such as fitness costs of the virulence allele, this allele will become fixed in the parasite population, and no persistent RQ dynamics will emerge (Jayakar 1970; Leonard 1977). The MA model (Frank 1993) assumes that a parasite can infect a host only if each of its alleles, carried at a number of loci, match corresponding alleles in the host. Thus, there is no universally successful parasite genotype in this model, and as a consequence, persistent oscillations of host and parasite allele frequencies can emerge.

Although these models are important to study theoretically as the simplest generic models of host-parasite inter-

* E-mail: j.engelstaedter@uq.edu.au.

Am. Nat. 2015. Vol. 185, pp. E000–E000. 
DOI: 10.1086/680476

actions, real host-parasite infection genetics are often more complex than either the GFG or the MA model. As cautioned already in the 1980s, GFG infection genetics may accurately describe only a subset of plant-plant pathogen systems (Barrett 1985), and more recent work has indeed uncovered a variety of factors other than resistance and avirulence genes that are involved in plant immunity and pathogen virulence (Bent and Mackey 2007). In animals, only a few systems have been studied extensively, but in a recent study Luijckx et al. (2013) demonstrated a complex pattern of dominance (and possibly epistasis) in the infection genetics of a *Daphnia magna–Pasteuria ramosa* system. Although these infection genetics exhibit the main MA property of reversibility of host susceptibility and resistance depending on parasite genotype, both the one-locus model and the two-locus model proposed by Luijckx et al. (2013) to explain their data are quite distinct from the canonical MA model.

The aim of this article is to explore the coevolutionary dynamics arising from a much more comprehensive set of infection genetics models than are given by the classic GFG and MA models. Specifically, I take a combinatorial approach in which I examine a large number of infection genetics models that are based on two major assumptions. First, the infection outcome of a host-parasite interaction is determined by only one or two host loci. This assumption is in line with the results obtained in the *D. magna–P. ramosa* system mentioned above (Luijckx et al. 2013) and also with a survey of quantitative trait locus studies that indicate that, in most animal and plant species, only a few loci are involved in host resistance (Wilfert and Schmid-Hempel 2008). Second, the infection outcome of a host-parasite interaction is assumed to be binary: a host with a given genotype is either fully susceptible or fully resistant to a parasite with a given genotype. Again, this assumption has empirical support from the *D. magna–P. ramosa* system, in which genotype-genotype interactions are highly specific (Luijckx et al. 2011).

In contrast to much previous theoretical work, I will also assume throughout that hosts are diploid and that the infection genetics can therefore be characterized by both dominance and (in the two-locus case) epistasis. Two main questions will be addressed in this investigation: (1) What types of infection genetic models produce RQ dynamics? (2) How do sex and recombination in the hosts influence the occurrence and properties of such dynamics?

## Methods

### *The Model*

Two antagonistically interacting populations are considered: hosts and parasites. Both of these populations are assumed to be of infinite size, so that random genetic drift effects can be ignored. Hosts are diploid, and their interaction with the parasites is determined by their allelic state at either one or two biallelic loci. Thus, in the one-locus scenario, there are three different genotypes (*aa, Aa, AA*), whereas in the two-locus scenario, there are 10 different genotypes (including the two phenotypically equivalent but genetically distinct double heterozygotes *AaBb* and *AabB*). In the basic model, the parasite population consists of two different genotypes, but model extensions with four and eight parasites will also be considered. Because the parasites reproduce clonally and without mutation, the exact genetic architecture (e.g., ploidy and number of loci) of the parasite genotypes is irrelevant. Frequencies of host and parasite genotypes in the two populations are denoted by the vectors $\mathbf{h}$ and $\mathbf{p}$.

For both hosts and parasites, I assume a life cycle of discrete, nonoverlapping generations that comprises two steps: selection and reproduction. The selection step is based on the infection matrix $\mathbf{L}$, in which rows represent host genotypes and columns represent parasite genotypes. I assume that a given parasite of genotype $j$ is either able to infect a host of genotype $i$, in which case $L_{ij} = 1$, or is unable to infect this particular host, in which case $L_{ij} = 0$. The infection matrices $\mathbf{L}$ will be represented in the remainder of this article as color diagrams (see fig. 1). Infection leads to a fitness reduction $s_{\mathrm{H}}$ in the host, whereas failure to infect entails a fitness cost $s_{\mathrm{P}}$ for the parasite. If we also assume that hosts and parasites encounter each other randomly according to the mass action principle, the recursion equations for the selection step are given by

$$h_i^+ = \frac{h_i[1 - s_{\mathrm{H}}(\mathbf{Lp})_i]}{\overline{W}_{\mathrm{H}}},$$
$$p_j^+ = \frac{p_j\left(1 - s_{\mathrm{P}}[\mathbf{h}^{\mathrm{T}}(\mathbf{1} - \mathbf{L})]_j\right)}{\overline{W}_{\mathrm{P}}}, \tag{1}$$

where $\overline{W}_{\mathrm{H}}$ and $\overline{W}_{\mathrm{P}}$ are the average fitnesses in hosts and parasites and $\mathbf{1}$ denotes a matrix of appropriate dimensions whose elements are all 1.

Reproduction in the parasites is assumed to be always asexual and hence does not affect the genotype frequencies. Denoting the parasite frequencies in the next generation by $p_j'$, this means that $p_j' = p_j^+$ for all parasite genotypes $j$. For the hosts, either asexual (clonal) or sexual reproduction is assumed. With asexual reproduction, $h_j' = h_j^+$. Sexual reproduction involves Mendelian segregation of alleles and, when two loci are considered, may also involve recombination. With sexual reproduction and assuming random mating, we have

$$h_i' = (\mathbf{h}^+)^{\mathrm{T}}\mathbf{M}_i\mathbf{h}^+ \tag{2}$$

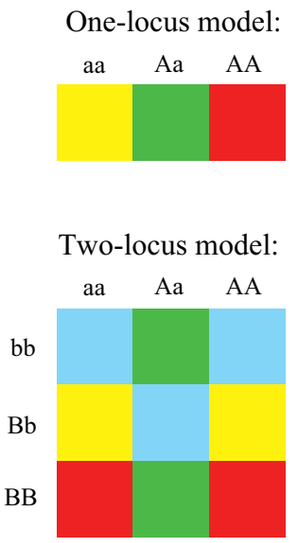


**Figure 1:** Illustration of the color diagrams used to denote individual infection matrices. Each square in the diagram represents one host genotype, and its color indicates whether this host can be infected by both parasite genotypes (red), by parasite 1 only (yellow), by parasite 2 only (blue), or by neither of the parasites (green). For example, the diagram for the one-locus model corresponds to the infection matrix ((1,0), (0,0), (1,1)), where rows indicate the infection outcome for the three host genotypes *aa, Aa*, and *AA* and columns for the two parasite genotypes. Note that, in the two-locus model, the two double heterozygotes (*AaBb* and *AabB*) are assumed to be phenotypically identical (middle square), although they are genetically distinct and have their own genotype frequencies in the model.

for all host genotypes $i$. Here, the tensor $M$ accounts for Mendelian inheritance and may, in the two-locus model, depend on the recombination rate $r$. Specifically, $M_{i,kl}$ is the proportion of offspring with genotype $i$ that are produced by parents with genotypes $k$ and $l$.

### *Simulation Methods*

To investigate the impact of the infection genetics on the coevolutionary dynamics, a combinatorial approach was taken in which all possible infection matrices **L** were screened (exploiting symmetries in these matrices, as detailed in the "Results" section below). The one-locus model was treated analytically through fixed-point and stability analyses. All analytical results were confirmed through simulations in which the recursion equations were iterated numerically. The two-locus model was found to be not amenable to analytical treatment, so extensive simulations were performed for all infection matrices and different parameter combinations. Initially, 12 replicate simulations were run for each infection matrix, with a burn-in phase of 5,000 generations and a 1,000-generation measurement phase. These simulations were initialized with random genotype frequencies in both hosts and parasites. This screening was then followed by a more intensive screening of the subset of matrices that exhibited consistent allele frequency fluctuations (defined as exhibiting an amplitude of allele frequencies $>0.01$ during the measurement phase in all 12 replicates). In this second screening, the simulations were run with a burn-in phase of 8,000 generations, a measurement phase of 2,000 generations, and all combinations of the initial allele frequencies 0.01, 0.1, 0.9, and 0.99 at both host loci and in the parasite population (i.e., 64 simulations per infection matrix). Here, the host population was initialized in Hardy-Weinberg and linkage equilibrium.

## Results

### *Models with Asexual Reproduction*

With clonal host reproduction, selection operates on host genotypes without the intermingling forces of segregation and recombination. As a result, both the one-locus and two-locus diploid models essentially reduce to a one-locus multiallele haploid model. Host genotypes conferring the same resistance patterns to the two parasites can be grouped, because they behave identically. This means that there will be at most four classes of host genotypes that correspond to a maximum of four alleles in a single-locus haploid model: those resistant to parasite 1 only, those resistant to parasite 2 only, those not resistant to any parasite, and those resistant to both parasites.

The ensuing evolutionary dynamics in these models are relatively straightforward to predict (analysis not shown). When there is one host genotype that provides protection against both parasite genotypes, this genotype will inevitably go to fixation. When there are more than one of those "superhost" genotypes, they will jointly spread, drive all other genotypes to extinction, and remain as a neutral polymorphism in the population. Conversely, one of the parasites may be able to infect all host genotypes; this "superparasite" will also go to fixation and produce a neutral polymorphism in the host population. RQ dynamics (i.e., cyclic coevolutionary dynamics of genotype frequencies) will emerge if and only if two conditions are met: (1) some host genotypes can be infected only by parasite 1, whereas others can only be infected by parasite 2, and (2) no host genotype is resistant to both parasites. Condition (1) can be summarized as "reciprocal specificity" in the genetic interaction or, alternatively, as the presence of host genotype versus parasite genotype (GxG) interactions inherent in the infection genetics. Note that, although RQ dynamics emerge under these conditions and may continue for many generations, those dynamics do not continue indefinitely: an outward spiral of genotype frequency

dynamics ensues, with the coevolutionary cycles increasing slowly in amplitude but decreasing in speed. As shown by Seger (1988), however, even a very low mutation rate can stabilize these cycles and lead to sustained RQ dynamics.

### *One-Locus Model with Sexual Reproduction*

When reproduction is sexual, we can exploit the fact that segregation of alleles leads to Hardy-Weinberg equilibrium in every generation. Denoting the frequency of allele $A$ within the host population by $x$ and the frequency of parasite genotype 2 by $y$, the model can then be reduced to only two recursion equations:

$$x' = x\frac{1 - s_H\begin{pmatrix}0\\1-x\\x\end{pmatrix}^T \mathbf{L}\begin{pmatrix}1-y\\y\end{pmatrix}}{1 - s_H\begin{pmatrix}(1-x)^2\\2x(1-x)\\x^2\end{pmatrix}^T \mathbf{L}\begin{pmatrix}1-y\\y\end{pmatrix}},$$
$$y' = y\frac{1 - s_P\begin{pmatrix}(1-x)^2\\2x(1-x)\\x^2\end{pmatrix}^T (\mathbf{1}-\mathbf{L})\begin{pmatrix}0\\1\end{pmatrix}}{1 - s_P\begin{pmatrix}(1-x)^2\\2x(1-x)\\x^2\end{pmatrix}^T (\mathbf{1}-\mathbf{L})\begin{pmatrix}1-y\\y\end{pmatrix}}. \quad (3)$$

(Here, **1** denotes again a matrix where all elements are 1s.) For each possible infection matrix **L**, equilibria of this simplified system of recursion equations can now be derived analytically and their stability determined through standard eigenvalue analyses of the corresponding Jacobian matrices (analysis not shown). This endeavor is facilitated by exploiting symmetries inherent in many of the infection matrices: although there are $2^{3\times2} = 64$ possible matrices **L**, the labeling of host alleles, as well as the labeling of the two parasites, is arbitrary, so there are classes of (at most four) equivalent matrices producing essentially the same dynamics. In terms of the color diagrams (fig. 1*A*), equivalent matrices are obtained by mirroring the diagram horizontally and/or by switching the colors yellow and blue. Eliminating such equivalent matrices leaves 24 "unique" infection matrices (representatives of their equivalency classes) that are sufficient to consider.

Figure 2 summarizes the dynamical properties emerging from these 24 infection matrices **L**. The four corners of the state space ($(x,y) \in \{(0,0), (0,1), (1,0), (1,1)\}$), representing extinction of one host allele and one parasite genotype, are always equilibrium points. In addition, there can also be internal equilibria. In general, neither the position nor the stability of the equilibria depend on the strength of selection, $s_H$ and $s_P$, with the only exception being the infection matrix shown in figure 2*L* (see below). In most cases, either an asymptotically stable or a neutrally stable equilibrium will be reached. In some cases, this may lead to a polymorphism of host alleles that is stably preserved in the population due to overdominance (fig. 2*K*, 2*M*, 2*V*).

There are five representative infection matrices that exhibit fluctuating dynamics around an internal equilibrium (fig. 2*F*, 2*L*, 2*O*, 2*R*, 2*T*). The common property of these matrices is that they involve GxG interactions at one or more pairs of host genotypes (indicated by both a yellow and a blue square in the diagrams). This can give rise to complex eigenvalues associated with the internal equilibria and thus oscillatory dynamics around these equilibria. However, in three out of these five infection matrices (fig. 2*F*, 2*R*, 2*T*), these oscillations are always only transient, and the system quickly reaches an equilibrium state. With the matrix in figure 2*F*, in which one of the homozygous genotypes represents a "superhost," the system always converges to a state in which host allele $A$ has become extinct and the two parasite genotypes are no longer under selection. With the matrix in figure 2*R*, the internal equilibrium itself, $(\hat{x}, \hat{y}) = (1/3, 2/3)$, is asymptotically stable (see also fig. A1*A* for example dynamics; figs. A1–A3 available online). The matrix in figure 2*T* gives rise to dynamics that very quickly spiral outward (i.e., oscillations that increase in amplitude and become slower), so that the system eventually ends up in one of the four corner equilibria (see also fig. A1*B*). This is because of the underdominance inherent in this matrix: the heterozygote hosts can be infected by both parasites, inhibiting invasion of alleles that are beneficial in the homozygous state and thereby bringing the oscillations to a halt. The matrix in figure 2*L*, representing the opposite case of overdominance, is interesting because it is the only matrix in which the stability of the internal equilibrium depends on the strength of selection. Specifically, the equilibrium $(\hat{x}, \hat{y}) = (1/2, 1/2)$ is asymptotically stable for $s_P < 4/5$ (see fig. A1*C* for the resulting dynamics in this case). Only with very strong selection against unsuccessful parasites, $s_P > 4/5$, does the internal equilibrium become unstable and persistent oscillatory dynamics emerge. Finally, the matrix shown in figure 2*O* always produces prolonged oscillations of both host allele and parasite genotype frequencies, independent of the strength of selection. Technically, these dynamics also exhibit outward spiraling (dynamics shown in fig. A1*D*), but at a much slower rate than in the underdominant case shown in figures 2*T* and A1*B*. Moreover, even very low mutation rates can lead to truly persistent RQ dynamics with the matrix in figure 2*O* (results not shown).

Comparing the results of the above analysis in sexual populations with the situation of asexual reproduction, it can be seen that sex can both produce persistent RQ dynamics that would not occur with asexual reproduction (matrix in fig. 2*L*) and annihilate oscillations that would arise with asexual reproduction (matrices in fig. 2*Q*, 2*R*,

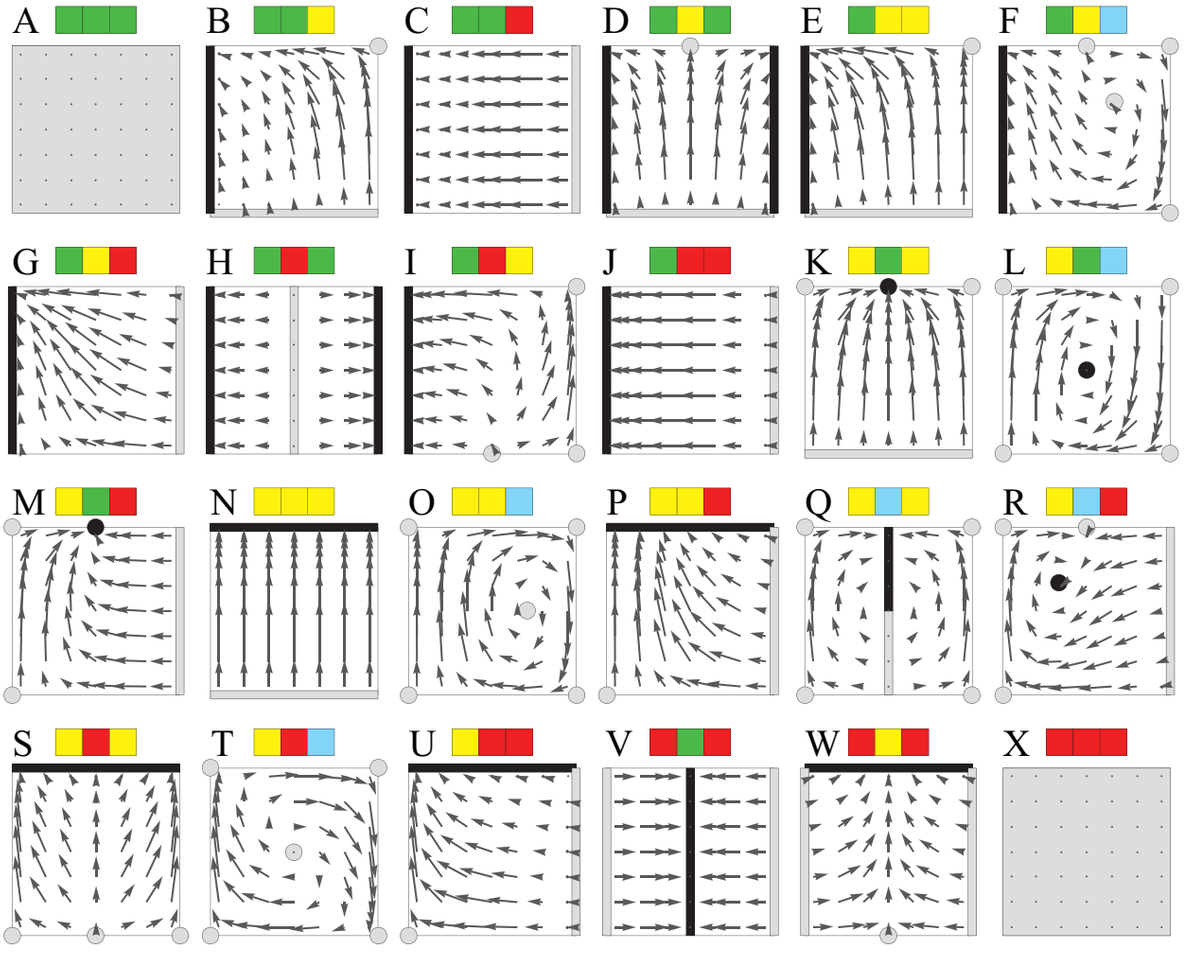


**Figure 2:** Evolutionary dynamics in the one-locus model for each of the 24 representative infection matrices **L**. In each panel, the infection matrix is shown as a color diagram on top on the plots (see fig. 1). In each plot, the horizontal axis gives the frequency $x$ of host allele $A$, whereas the vertical axis gives the frequency $y$ of parasite genotype 2. Black circles indicate stable equilibria of the system, whereas gray circles indicate unstable equilibria. Black and gray thick lines indicate a continuum of equilibria that are externally stable or unstable, respectively. Note that both the position and the stability of equilibria are independent of $s_H$ and $s_P$, with the exception of $L$, where the internal equilibrium can be either stable (as shown) or unstable. In $A$ and $X$, the entire state space consists of neutral equilibria. Arrows indicate the change in frequencies from one generation to the next; here $s_H = s_P = 0.8$ was assumed throughout.

$2T$). In both cases, this is because meiosis and syngamy ensure that, as long as both alleles are maintained in the host population, the two homozygous genotypes will always produce individuals carrying the heterozygous genotype and vice versa. It should also be noted that, with sexual reproduction, the presence of a superhost genotype no longer precludes RQ dynamics, as long as this is a heterozygous genotype (fig. 2$L$).

### *Two-Locus Model with Sexual Reproduction*

With two host loci, the dynamical system could not be treated analytically, so this model was investigated through extensive simulations, screening again all possible infection matrices **L**. In the two-locus model, there are $2^{9\times2} = 262{,}144$ such matrices, but as in the one-locus model, these matrices form equivalency classes, because the labeling of alleles at both host loci, labeling of the two loci themselves, and labeling of the two parasites are arbitrary, so that one infection matrix **L** can have up to 15 equivalent matrices. Considering only a single representative member of each of these equivalency classes reduces the total number of matrices to be screened to 17,676.

Not surprisingly, the majority of infection matrices resulted in dynamics in which alleles at one or both host loci became fixed or reached stable polymorphic equilib-

ria. Nevertheless, there are many infection matrices that did produce persistent oscillatory dynamics of parasite genotype frequencies and host allele frequencies at both host loci. In what follows, I will focus on those latter infection matrices, called RQ matrices (RQMs).

Figure 3 shows that the presence or absence of sexual reproduction and recombination has a large impact on sets of RQMs, with only very little overlap between these sets across the different modes of reproduction (fig. 3*A*). Given that the absence of superhosts and superparasites inevitably leads to RQ dynamics with asexual but not with sexual reproduction (see above), it is a rather intuitive result that many matrices are RQMs with asexual but not with sexual reproduction. However, the opposite is also true, indicating that many matrices that are RQMs with sexual reproduction are characterized by the presence of superhosts. The rate of recombination between the two host loci also has a strong impact on whether a matrix is an RQM (fig. 3*B*). By contrast, the strength of selection has a much weaker effect and tends to produce a nested pattern, in that matrices that are RQMs with weak selection also tend to be RQMs with stronger selection but not vice versa (fig. 3*C*, 3*D*).

It can also be seen in figure 3 that, although the number of infection matrices that are classified as RQMs increases monotonically with the strength of selection on both hosts

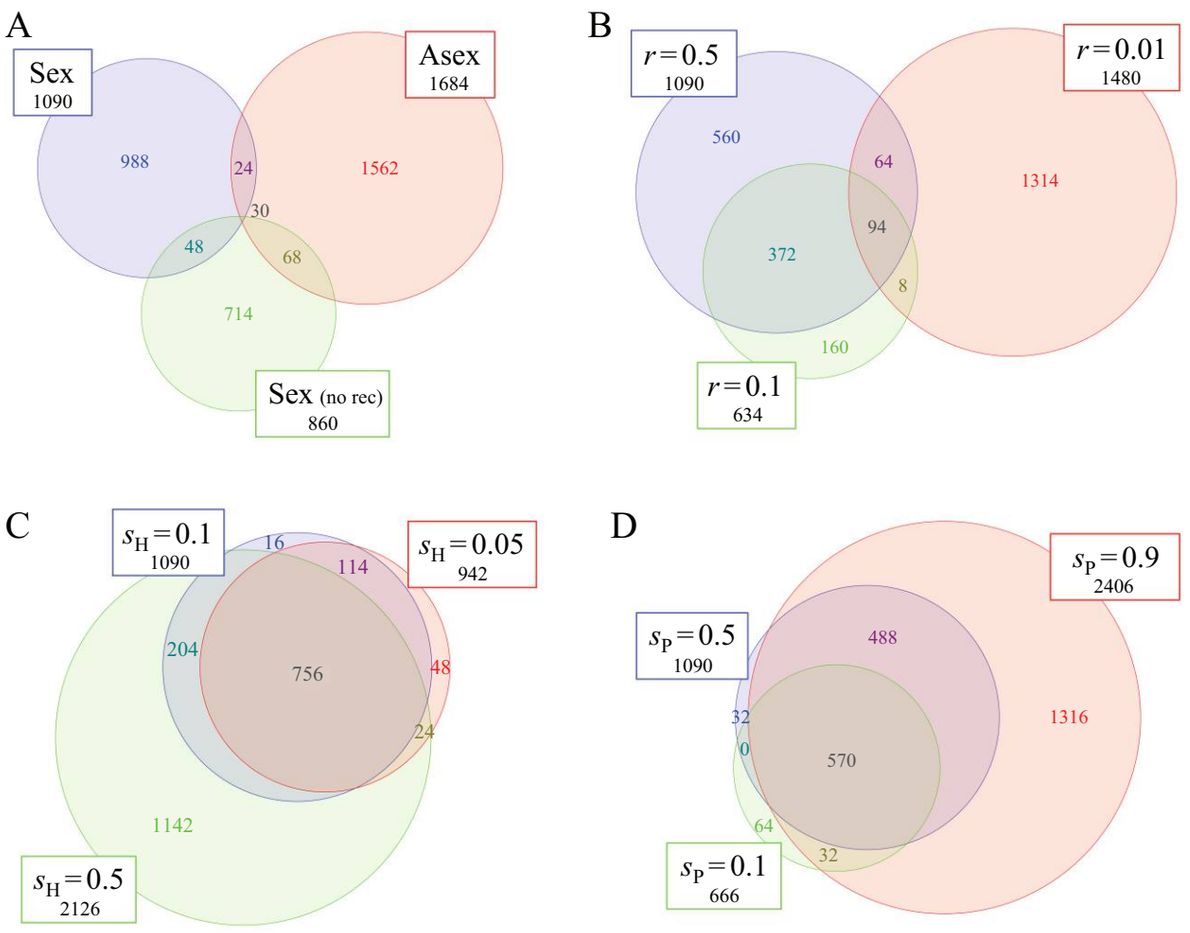


**Figure 3:** Venn diagrams showing sets of infection matrices **L** that consistently lead to Red Queen dynamics under different conditions in the two-locus model. These conditions are (*A*) sexual reproduction with free recombination ($r = 0.5$; Sex), sexual reproduction with complete linkage of the two loci ($r = 0$; Sex (no rec)), and asexual (clonal) reproduction (Asex); (*B*) different recombination rates between the loci; (*C*) different strengths of selection against infected hosts; and (*D*) different strengths of selection against unsuccessful parasites. Numbers within circles indicate the number of matrices **L** in the respective subset, whereas the total numbers of matrices within each set is given underneath its label. Unless otherwise indicated, reproduction is always assumed to be sexual with free recombination, and selection coefficients take the values $s_H = 0.1$ and $s_P = 0.5$.

and parasite, the effect of sex and recombination rates is more subtle. Asexual reproduction entails more RQMs than sexual reproduction with any of the recombination rates tested, but the numbers are similar to those with sexual reproduction, and $r = 0.01$. With sexual reproduction, the number of RQMs is nonmonotonic with respect to recombination rates, with the highest number found for low (but positive) recombination rates and the lowest number found for an intermediate recombination rate. The strong and nonmonotonic impact of recombination rates is also illustrated in figure A3. It should be noted, however, that the absolute numbers of RQMs under different parameter values are arguably of limited biological relevance. This is because the infection matrices that describe actual host-parasite systems in nature probably constitute a very specific subset of all possible infection matrices, and within this as-yet-unknown subset, the numbers of RQMs may well follow different trends than when all infection matrices are considered.

Figure 4*A* shows some example RQMs for the case of sexual reproduction with free recombination. Some of these matrices do not entail superhost or superparasite genotypes and are thus also RQMs with asexual reproduction. However, many RQMs do involve superhosts (indicated by green squares); these superhosts are invariably single or double heterozygotes and therefore do not select for fixation of host alleles when reproduction is sexual. It can also be seen from these examples that the number of host genotypes that can be infected by both parasites (red squares in diagrams) varies considerably among RQMs. Figure 4*B* gives infection matrices that are very similar to those in Figure 4*A* but are not RQMs with sexual reproduction. This demonstrates that subtle differences in the infection genetics can lead to qualitatively different coevolutionary dynamics. Indeed, aside from the requirements that, with sexual reproduction, RQMs must not involve any double-homozygous superhosts and that there are some GxG interactions, no further necessary or sufficient conditions for an infection matrix being an RQM could be conjectured. Finally, figure 4*C* shows the infection matrix that has recently been proposed to explain data in a *Daphnia magna–Pasteuria ramosa* system (in addition to a one-locus model; Luijckx et al. 2013). Rather intuitively, under this model, the B allele becomes fixed in the population, and oscillation of allele frequencies occurs only at locus A (not shown).

Example coevolutionary dynamics for some of the RQMs in figure 4*A* are shown in figure A2, illustrating that these dynamics can differ considerably with respect to the speed and amplitude of allele frequency oscillations, even though the selection coefficients are kept constant in these simulations. A more complete picture of key statistics of coevolutionary dynamics resulting from RQMs with varying mode of reproduction and recombination rate is provided in figure 5. Sexual versus asexual (clonal) reproduction is clearly a key factor for all of these statistics. With asexual reproduction, the distribution of the variance in allele frequencies, as well as the amplitude in allele frequency oscillations, is quite distinct from (i.e., more widespread and uniform than) the corresponding distributions with sexual reproduction. Asexual reproduction appears to lead to little variance in cycle length of allele frequency oscillations or mean host fitness, in direct contrast to the RQMs with sexual reproduction, and on average, cycles are slower and hosts harmed more with asexual than with sexual reproduction. On the other hand, the distributions of the statistics describing the coevolutionary dynamics are much more similar across different recombination rates, although marked differences are also apparent here. Figure A3 shows some further example dynamics that underline both the quantitative and qualitative impact

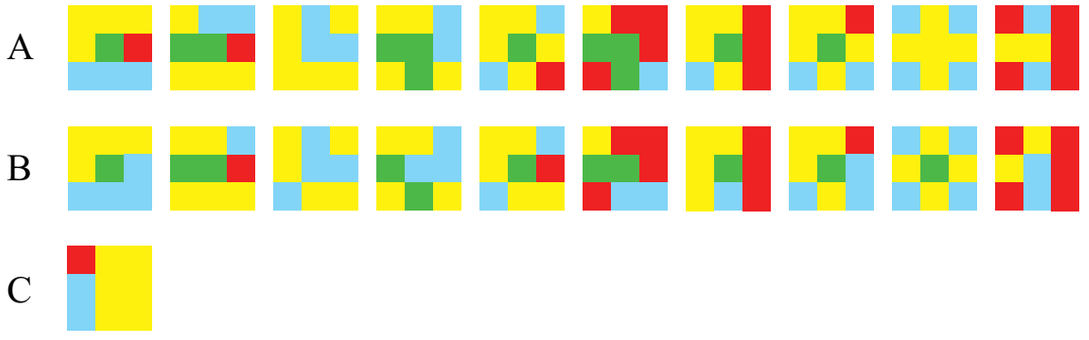


**Figure 4:** Examples for infection matrices in the two-locus model (shown as color diagrams; see fig. 1) that (*A*) produce sustained Red Queen (RQ) dynamics or (*B*) do not produce RQ dynamics, in both cases with sexual reproduction under free recombination and with selection coefficients $s_H = 0.1$ and $s_P = 0.5$. *C* shows the infection matrix that represents one of the models proposed for infection genetics in the *Daphnia magna–Pasteuria ramosa* system (Luijckx et al. 2013).

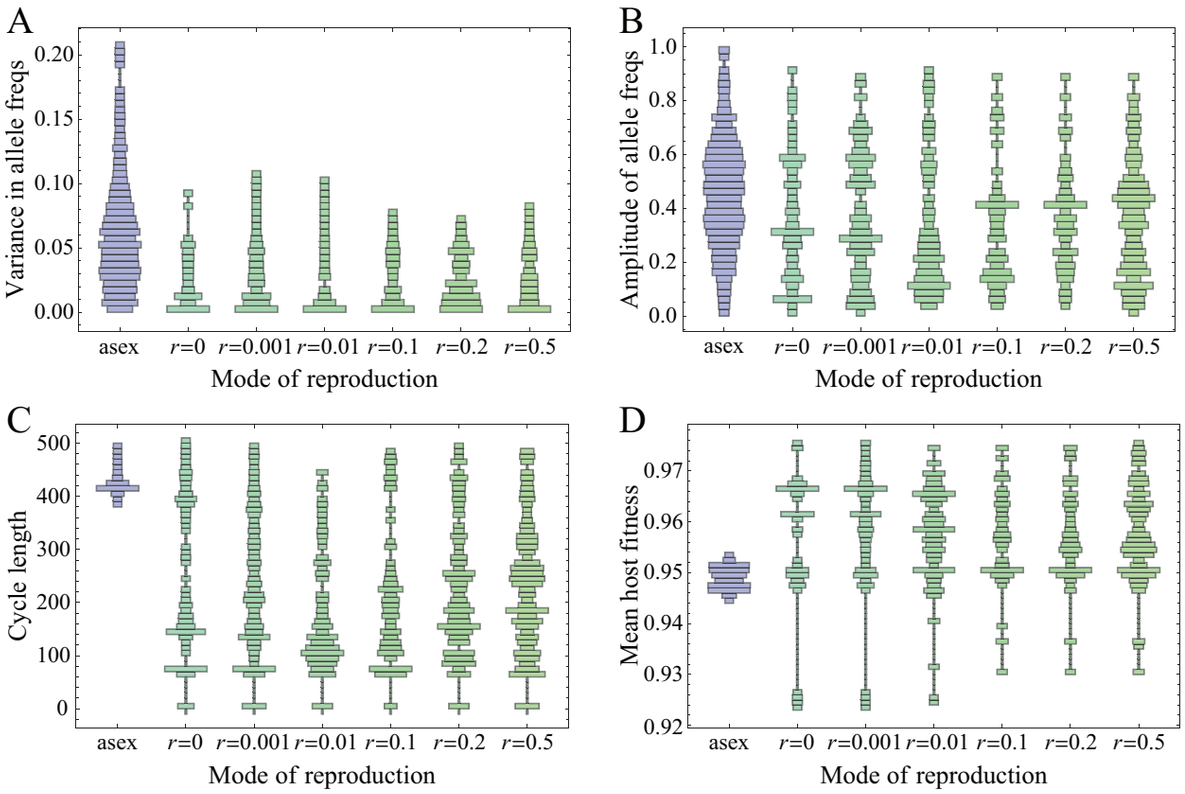


**Figure 5:** Distribution of key statistics describing Red Queen (RQ) dynamics in the two-locus model under asexual or sexual reproduction with various recombination rates. Distributions are taken over all infection matrices **L** that lead to persistent RQ dynamics at both host loci. *A*, Variance in host allele frequencies over time, averaged over the two loci. *B*, Amplitude of host allele frequencies, averaged over the two loci. *C*, Length of coevolutionary cycles in host allele frequencies, averaged over the two loci. *D*, Geometric mean over time of mean host fitness. Selection coefficients take the values $s_H = 0.1$ and $s_P = 0.5$.

of the recombination rate on the coevolutionary dynamics; this figure also indicates that the effect of recombination is not necessarily monotonic.

### *Model with More than Two Parasites*

So far, only infection matrices with two parasite genotypes have been considered. The model presented here can readily be extended to include an arbitrary number of (asexually reproducing) parasite genotypes. Unfortunately, the number of possible infection matrices grows very quickly with increasing parasite number, such that exhaustive screening of all infection matrices becomes infeasible. Therefore, additional simulations with four and eight parasites were run with a nonexhaustive set of randomly chosen infection matrices. To achieve a broad range of infection matrices, matrices were sampled with an increasing number $m$ of genotype-genotype interactions leading to infection (i.e., number of 1s in the infection matrices **L**). Specifically, with four parasites, 100 matrices for each $m \in \{8, 9, 10, ..., 28\}$, and with eight parasites, 100 matrices for each $m \in \{16, 18, 20, ..., 56\}$ were randomly sampled. For each infection matrix, 12 replicate simulations with random initial frequencies were run under sexual reproduction with free recombination. The results were compared with those obtained in the two-parasite model using 200 random infection matrices for each $m \in \{4, 5, 6, ..., 14\}$.

With a total number of eight parasites, the median number of parasite genotypes maintained in the population is only marginally higher than that with four parasites, and only very few infection matrices in the eight-parasite scenario were found that support more than four parasites (fig. 6*A*). The number of parasite genotypes exhibiting fluctuating dynamics increases with the total number of parasites considered but usually also remains low. On the other hand, increasing the number of parasite genotypes considerably increases the number of host loci at which polymorphism is maintained and also increases the

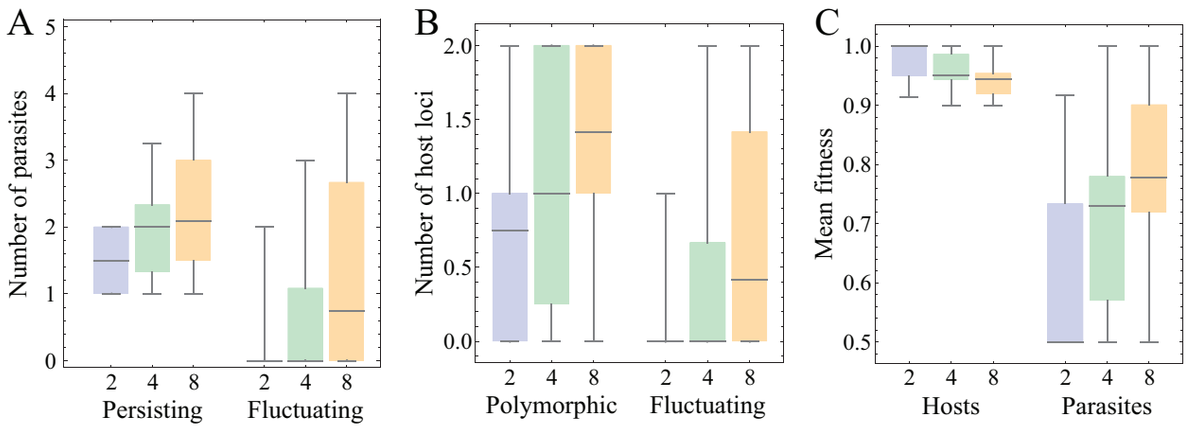


**Figure 6:** Box plots comparing coevolutionary dynamics with different numbers of parasite genotypes (blue: two genotypes; green: four genotypes; orange: eight genotypes). Each box summarizes the distribution of a key statistic derived from simulating the coevolutionary dynamics arising from 2,200 (two parasites) or 2,100 (four and eight parasites) infection matrices, averaged over 12 replicate simulations for each matrix. Horizontal lines show medians, boxes show interquartile ranges, and whiskers show the 95% interquantile range. In *A*, the number of parasite genotypes that persist in the population, as well as the number of parasite genotypes that exhibit fluctuating dynamics, is shown. *B* shows how many host loci remain polymorphic and at how many loci fluctuating allele frequencies are observed. In *C*, the distribution of the geometric mean over time of mean population fitness in both hosts and parasites is shown.

number of host loci at which fluctuating allele frequencies are observed (fig. 6*B*). This is also reflected in the number of RQMs, as defined in the previous subsection, that are present in the matrix samples (only 6 with two parasites, 37 with four parasites, and 144 with eight parasites). Finally, mean host fitness decreases and mean parasite fitness increases with increasing number of parasite genotypes (fig. 6*C*). This is expected, because if each parasite can infect, on average, the same number of hosts (i.e., a roughly constant ratio of $m$ and parasite number due to the sampling strategy described above), adding more parasites will generally harm the host and benefit the parasite population.

## Discussion

### *Relationship to Standard Infection Genetics Models*

The models considered in this article represent a generalization of standard infection genetics models, such as the GFG and MA models. Accordingly, different versions of both of these models are embedded in the present model. This is best seen in the one-locus case. On the one hand, two of the representative infection matrices (eight in total) can be considered diploid versions of the GFG model (see fig. 2*P*, 2*U*). These matrices differ in the phenotype of the heterozygous individuals, where the hosts' resistance gene $a$ is either dominant (fig. 2*P*) or recessive (fig. 2*U*). As in the haploid GFG model, without additional fitness costs, one of the parasite genotypes—carrying what is usually called the "virulence gene" in the literature on plant infection genetics—always becomes fixed with these two infection matrices. Recently, Fenton et al. (2009) proposed and investigated an "inverse gene-for-gene" (IGFG) model in which the roles of hosts and parasites are essentially reversed in the infection genetics; two simple versions of this model are also embedded in the set of infection matrices considered here, with the expected dynamical properties (fig. 2*B*, 2*E*).

On the other hand, three of the representative infection matrices (eight in total) are diploid versions of the MA model (fig. 2*L*, 2*O*, 2*T*). These three matrices are characterized by GxG interaction at the two homozygous host genotypes, where *aa* hosts can be infected only by parasite 1, and *AA* hosts can be infected only by parasite 2. The matrices differ in whether the heterozygous host genotype is resistant to both parasites (resulting in overdominance; fig. 2*L*), resistant to only one parasite (resulting in dominance of one resistance allele and recessiveness of the other; fig. 2*O*), or susceptible to both parasites (resulting in underdominance; fig. 2*T*). Of these three models, the underdominance model mimics a matching infection system in which the host is susceptible as soon as the parasite carries an allele that matches a corresponding host allele. The overdominance model is perhaps more in line with the original description of the MA model (Frank 1993), in which a host is resistant if it carries an allele that matches a corresponding allele in the parasite; today, this is usually referred to as the "inverse matching-allele" model. Only in the dominance case or, when selection on the parasites is

strong, with overdominance, do persistent allele frequency oscillations arise. An MA-like infection matrix is thus a necessary but not sufficient condition for RQ dynamics in the one-locus version of the present model.

In the two-locus model, many of the infection matrices could also be regarded as GFG models or MA models, with a variety of epistatic interactions between the two loci. Moreover, the complete set of matrices also contains matrices that combine both of these models (e.g., GFG at one and MA at the other locus, or one locus determining whether the interaction at the other locus is GFG or MA). The presence of GxG interactions (i.e., the existence of hosts that can be infected only by parasite 1 and of hosts that can be infected only by parasite 2) is a prerequisite for RQ dynamics to occur, and this is the fundamental property of the MA model. Nevertheless, in light of the hundreds of different infection matrices that fulfill this condition and the striking differences between these models (see fig. 4), referring to all of these models as MA models arguably means to stretch this term beyond recognition.

### *Relationship to Previous Models with Nonstandard Infection Genetics*

Several previous authors have put forward models that are extensions of the classic GFG and MA models in directions other than the one considered in this article. Parker (1994) and Agrawal and Lively (2002) examined haploid two-locus models that allowed for a continuum of infection genetics between the pure GFG and MA models, showing that small deviations from the GFG model are sufficient to produce RQ dynamics even in the absence of fitness costs of virulence and resistance alleles. Similarly, Agrawal and Lively (2003) and Fenton et al. (2012) have considered mixtures of different models (MA, GFG, and IGFG) in which these models account for different stages in the infection process (e.g., recognition and defense). Switkes and Moody (2001) analyzed a model for coevolutionary interactions between a diploid and a haploid species with one biallelic locus in each species. This model allows for arbitrary fitness effects resulting for all possible genotype-genotype interaction and therefore includes as a special case the one-locus version of the present model. However, due to the large number of parameters, only general results concerning the number and position of equilibria could be obtained by these authors. Engelstädter and Bonhoeffer (2009) also considered a model with arbitrary fitness effects, but their model considered two haploid species with two loci each, and their approach was computational rather than analytical. These authors showed that infection matrices with a high degree of "antagonicity"—defined as the amount of negative correlation between host and parasite fitnesses across all genotype-genotype combinations—are more likely to exhibit RQ dynamics. (Note that, in the present model as in most other models of host-parasite coevolution, antagonicity is always maximal by default.) A similar approach using generalized fitness interaction matrices was taken by Kwiatkowski et al. (2012), who investigated the tripartite coevolutionary dynamics between parasites, hosts, and their protective symbionts.

Overall, it can be seen from these and other examples (e.g., Otto and Nuismer 2004) that a distinctive feature of the present model is that infection is interpreted not as a probabilistic but as a deterministic process: depending on its own genotype and the genotype of the host, a parasite can either infect a host (in which case, this host has its fitness reduced by a certain fixed amount) or fail to infect a host (in which case, it is the parasite's fitness that is reduced). This is, of course, a gross simplification, but one that has some empirical support (Luijckx et al. 2011). Moreover, because fitness is always relative in a population genetic framework, and because only overall fitness and not individual fitness components are considered, the model can also be interpreted much more broadly. In this interpretation, there are two types of genotype-genotype combinations of hosts and parasites, of which the first is, on average, more favorable for the host and disadvantageous for the parasite (e.g., low success rate of infection), whereas the other type is favorable for the parasite and disadvantageous for the host (e.g., high infection success).

### *The Impact of Sex and Recombination on Coevolutionary Dynamics*

For the same infection genetics, marked differences were observed between the coevolutionary dynamics in sexual versus asexual populations and those in populations with different recombination rates. In particular, only a minority of infection matrices were found to produce RQ dynamics with both sexual and asexual reproduction. This could have a number of interesting consequences that are yet to be explored. For example, it is not clear what coevolutionary dynamics to expect in host populations with a life cycle of cyclical parthenogenesis (e.g., as observed in water fleas, aphids, monogonont rotifers, and many unicellular eukaryotes): Can a single round of sexual reproduction per life cycle annihilate RQ dynamics that would occur in purely asexual species, or vice versa? Similarly but more subtly, variation in recombination rates between host individuals (including between males and females) may influence the coevolutionary dynamics in a complex manner.

It is also interesting to ponder the results of this study in the context of the RQ hypothesis for the evolution of sex (Jaenike 1978; Bell 1982; Hamilton et al. 1990; Salathé et al. 2008; reviewed in Lively 2010*b*). This hypothesis states that

sex and recombination are favored during rapid antagonistic coevolution, because recombination allows the production of genotypes that are currently underrepresented in the population but have disproportionately high fitness. Many of the infection matrices that produce RQ dynamics with sexual production are characterized by the existence of heterozygous genotypes that are resistant to both parasites. This means that the RQ hypothesis would not work in those situations, because asexual clones carrying these genotypes would inevitably spread to fixation. However, a situation is also conceivable in which RQ dynamics do occur in an asexual population and produce selection for sex (this, of course, would depend on other costs and benefits of sex), but where, as a result of the spread of sexual individuals, the RQ dynamics come to a halt. If this can indeed happen, the absence of oscillatory genotype frequency dynamics in a sexual population cannot be taken as evidence against the RQ hypothesis (see also Lively 2010*a*).

### *Model Limitations*

To enable the screening of a wide range of infection genetics models for their impact on the resulting coevolutionary dynamics, evolutionary forces other than recombination and natural selection acting on the different host and parasite genotypes were ignored in this model. Directional selection (independent of genotype frequencies of the other species) is known to be an important driver of coevolutionary dynamics, in particular in the form of costs of resistance and infectivity in GFG models, in which this type of selection determines the presence or absence of RQ dynamics (Sasaki 2000). Moreover, the sexual version of the model assumes random mating between individuals; inbreeding or other forms of assortative mating may have a strong impact on the coevolutionary dynamics. Mutation is another important evolutionary force that is absent in the model proposed here. For example, Seger (1988) showed that low mutation rates are necessary to produce persistent RQ dynamics in a discrete-time model with MA infection genetics, but that there exists a threshold mutation rate above which a stable interior equilibrium is reached. This result is fully applicable to asexual versions of the present model, and preliminary simulations show that, by preventing outward spiraling of allele frequencies, mutation will stabilize coevolutionary oscillations also in the sexual model (results not shown). By contrast, random genetic drift is expected to annihilate RQ dynamics, especially in cases in which the amplitude of allele frequency oscillations is high, so that an allele can easily become fixed or become extinct through drift. However, allele frequency oscillations were often found to occur at intermediate allele frequencies (see figs. 4*B*, A2 for examples), so that even in finite populations, RQ dynamics may be expected to occur with many infection matrices. Finally, spatial structure and migration are well-recognized factors promoting diversity in host-parasite coevolutionary dynamics, in particular within the geographical mosaic framework of coevolution (e.g., Thompson 1994; Gomulkiewicz et al. 2000; Thompson and Cunningham 2002).

The model proposed here is also very simplistic, in that it is based on a standard population genetic framework that tracks genotype frequencies over time but ignores ecological considerations. Many previous models have been much more explicit in that respect. The classic work by May and Anderson (1983), for example, demonstrates that when host-parasite coevolution is modeled explicitly as an epidemiological process, chaotic dynamical behavior can emerge, especially when not only the frequency of host genotypes but also overall host abundance is influenced by the parasites. Another example consists of abiotic environmental factors, such as temperature or availability of nutrients, that may modify the infection genetics (Wolinska and King 2009), such that fluctuations in these factors can alter the coevolutionary dynamics both quantitatively and qualitatively (Mostowy and Engelstädter 2011).

Finally, it should be noted that, because the focus of this work was on exploring the role of infection matrices and host genetic architecture (including mode of reproduction and recombination rates) on RQ dynamics, only a limited number of selection coefficients were investigated (usually $s_H = 0.1$ and $s_P = 0.5$). In reality, a wide range of both parameters is expected. The parameter $s_H$ will depend on host virulence but may also incorporate overall prevalence of the parasites (i.e., low parasite numbers can be modeled as low $s_H$). Selection on parasites may be expected to be very strong ($s_P \approx 1$) in situations where failure to cause an infection manifests at a stage when the parasite is no longer able to infect other hosts (e.g., when the parasite has already entered the host but is eliminated by the immune system). However, in cases in which failure to infect is caused by failure to enter the host, selection on the parasite population may be weaker, because unsuccessful parasites may be able to subsequently infect a different host.

### *Toward an Empirical Understanding of Infection Matrices*

The sets of infection matrices considered here may contain some matrices that are found in natural systems, but they will most likely also contain many that will never be observed. A major empirical task is thus to identify sets of infection matrices that are biologically plausible. The most direct approach to this problem is to determine the genetic basis of host defense and parasite counterdefense in individual host-parasite systems through cross-infection studies. Unfortunately, because of the need to genetically dissect different host strains and, at the very least, to isolate and

propagate individual parasite genotypes, this is a very difficult task and has thus far been feasible in only very few systems (e.g., Luijckx et al. 2013). Moreover, as first pointed out by Frank (1996), an inherent problem with this approach is that it relies on the sampling of genotypes from polymorphism currently available in the host and parasite populations under study. This means that genotypes that occur at a low frequency in the population will usually be missed, and wrong conclusions about the overall infection matrix and the expected coevolutionary dynamics may be drawn (see also Dybdahl et al. 2014 for a discussion of this problem). One way to reduce this problem would be to extensively sample many genotypes through time and space to achieve convergence to the "real" infection matrix involving all relevant genotypes. From a theoretical side, it would be interesting to extend Frank's (1993) analyses and ascertain to what extent we can infer generalized infection matrices such as those studied here with cross-infection studies of genotypes sampled from naturally existing polymorphism.

An alternative approach that has recently been advocated by Dybdahl et al. (2014) is to derive plausible infection matrices from experimentally established molecular principles about host immunology and parasite infection. This approach might uncover different types of infection matrices with different underlying host and parasite genes that govern the various stages of the infection process, such as host recognition by a parasite, parasite entry into the host, and detection and eradication of the parasite by the immune system. These matrices could then be combined (mathematically, in the simplest case, as a Hadamard product) to construct infection matrices for the entire infection process that are in accord with the molecular mechanisms. Such composite infection matrices have been studied for two-stage infection processes involving either an MA and a GFG infection matrix (Agrawal and Lively 2003) or an IGFG and a GFG infection matrix (Fenton et al. 2012). However, a more systematic exploration involving other combinations of infection matrices, such as more than two stages of infection and diploid host genetics, remains to be conducted.

### *Conclusions*

The complexity of empirically observed infection genetics underlines the importance of going beyond the simple, established models if we want to understand how hosts and their parasites coevolve. Considering generalized sets of matrices defining infection success for all combinations of host and parasite genotypes provides a natural way forward in that respect (e.g., Agrawal and Lively 2002; Otto and Nuismer 2004; Engelstädter and Bonhoeffer 2009; Kwiatkowski et al. 2012). The set of matrices analyzed here, assuming a binary outcome of infection, has three desirable properties. First, because it is agnostic about the mechanistic basis of host-pathogen interactions, it is highly flexible and allows for complex interactions between alleles at the same locus and at different loci (dominance and epistasis). Second, despite this flexibility, this matrix set is still manageable, at least for a low number of host alleles and parasite genotypes. By contrast, it is very difficult to meaningfully screen other matrix sets that are even more general (e.g., Switkes and Moody 2001; Engelstädter and Bonhoeffer 2009). Finally, the matrix set proposed here contains not only standard infection genetic models, such as the MA and GFG models as special cases, but also a recently proposed model explaining experimental results in the *Daphnia magna–Pasteuria ramosa* system (Luijckx et al. 2013). Provided that further empirical support for complex patterns of success/failure infection genetics accumulates, it would be interesting to further explore the coevolutionary dynamics resulting from infection matrices with binary outcomes in models incorporating more ecological and epidemiological realism.

## Acknowledgments

I would like to thank Y. Michalakis and two anonymous reviewers for helpful comments on the manuscript. This work was supported by the Swiss National Science Foundation (grant PZ00P3_132934).

## Literature Cited

Agrawal, A., and C. M. Lively. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evolutionary Ecology Research 4:79–90.

Agrawal, A. F., and C. M. Lively. 2003. Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. Proceedings of the Royal Society B: Biological Sciences 270:323–334.

Barrett, J. 1985. The gene-for-gene hypothesis: parable or paradigm. Pages 215–225 *in* D. Rollinson and R. M. Anderson, eds. Ecology and genetics of host-parasite interactions. Academic Press, London.

Bell, G. 1982. The masterpiece of nature: the evolution and genetics of sexuality. Croom Helm, London.

Bent, A. F., and D. Mackey. 2007. Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. Annual Review of Phytopathology 45:399–436.

Brockhurst, M. A., and B. Koskella. 2013. Experimental coevolution of species interactions. Trends in Ecology and Evolution 28:367–375.

Brown, J. K. M., and A. Tellier. 2011. Plant-parasite coevolution: bridging the gap between genetics and ecology. Annual Review of Phytopathology 49:345–367.

Decaestecker, E., S. Gaba, J. A. M. Raeymaekers, R. Stoks, L. Van Kerckhoven, D. Ebert, and L. De Meester. 2007. Host-parasite "Red Queen" dynamics archived in pond sediment. Nature 450:870–873.

Dybdahl, M. F., C. E. Jenkins, and S. L. Nuismer. 2014. Identifying the molecular basis of host-parasite coevolution: merging models and mechanisms. American Naturalist 184:1–13.

Engelstädter, J., and S. Bonhoeffer. 2009. Red Queen dynamics with non-standard fitness interactions. PLoS Computational Biology 5: e1000469.

Fenton, A., J. Antonovics, and M. A. Brockhurst. 2009. Inverse-gene-for-gene infection genetics and coevolutionary dynamics. American Naturalist 174:E230–E242.

———. 2012. Two-step infection processes can lead to coevolution between functionally independent infection and resistance pathways. Evolution 66:2030–2041.

Flor, H. H. 1955. Host-parasite interaction in flax rust—its genetics and other implications. Phytopathology 45:680–685.

Frank, S. A. 1993. Specificity versus detectable polymorphism in host-parasite genetics. Proceedings of the Royal Society B: Biological Sciences 254:191–197.

———. 1996. Statistical properties of polymorphism in host-parasite genetics. Evolutionary Ecology 10:307–317.

Gomulkiewicz, R., J. N. Thompson, R. D. Holt, S. L. Nuismer, and M. E. Hochberg. 2000. Hot spots, cold spots, and the geographic mosaic theory of coevolution. American Naturalist 156:156–174.

Hamilton, W. D. 1980. Sex versus non-sex versus parasite. Oikos 35:282–290.

Hamilton, W. D., R. Axelrod, and R. Tanese. 1990. Sexual reproduction as an adaptation to resist parasites (a review). Proceedings of the National Academy of Sciences of the USA 87:3566–3573.

Jaenike, J. 1978. An hypothesis to account for the maintenance of sex within populations. Evolutionary Theory 3:191–194.

Jayakar, S. D. 1970. A mathematical model for interaction of gene frequencies in a parasite and its host. Theoretical Population Biology 1:140–164.

Kwiatkowski, M., J. Engelstädter, and C. Vorburger. 2012. On genetic specificity in symbiont-mediated host-parasite coevolution. PLoS Computational Biology 8:e1002633.

Leonard, K. J. 1977. Selection pressures and plant pathogens. Annals of the New York Academy of Sciences 287:207–222.

Lively, C. M. 2010*a*. An epidemiological model of host-parasite coevolution and sex. Journal of Evolutionary Biology 23:1490–1497.

———. 2010*b*. A review of Red Queen models for the persistence of obligate sexual reproduction. Journal of Heredity 101:S13–S20.

Luijckx, P., F. Ben-Ami, L. Mouton, L. Du Pasquier, and D. Ebert. 2011. Cloning of the unculturable parasite *Pasteuria ramosa* and its *Daphnia* host reveals extreme genotype-genotype interactions. Ecology Letters 14:125–131.

Luijckx, P., H. Fienberg, D. Duneau, and D. Ebert. 2013. A matching-allele model explains host resistance to parasites. Current Biology 23:1085–1088.

May, R. M., and R. M. Anderson. 1983. Epidemiology and genetics in the coevolution of parasites and hosts. Proceedings of the Royal Society B: Biological Sciences 219:281–313.

Mostowy, R., and J. Engelstädter. 2011. The impact of environmental change on host-parasite coevolutionary dynamics. Proceedings of the Royal Society B: Biological Sciences 278:2283–2292.

Otto, S. P., and S. L. Nuismer. 2004. Species interactions and the evolution of sex. Science 304:1018–1020.

Parker, M. A. 1994. Pathogens and sex in plants. Evolutionary Ecology 8:560–584.

Salathé, M., R. D. Kouyos, and S. Bonhoeffer. 2008. The state of affairs in the kingdom of the Red Queen. Trends in Ecology and Evolution 23:439–445.

Sasaki, A. 2000. Host-parasite coevolution in a multilocus gene-for-gene system. Proceedings of the Royal Society B: Biological Sciences 267:2183–2188.

Seger, J. 1988. Dynamics of some simple host-parasite models with more than 2 genotypes in each species. Philosophical Transactions of the Royal Society B: Biological Sciences 319:541–555.

Switkes, J. M., and M. E. Moody. 2001. Coevolutionary interactions between a haploid species and a diploid species. Journal of Mathematical Biology 42:175–194.

Thompson, J. N. 1994. The coevolutionary process. University of Chicago Press, Chicago.

Thompson, J. N., and J. J. Burdon. 1992. Gene-for-gene coevolution between plants and parasites. Nature 360:121–125.

Thompson, J. N., and B. M. Cunningham. 2002. Geographic structure and dynamics of coevolutionary selection. Nature 417:735–738.

Wilfert, L., and P. Schmid-Hempel. 2008. The genetic architecture of susceptibility to parasites. BMC Evolutionary Biology 8:187.

Wolinska, J., and K. C. King. 2009. Environment can alter selection in host-parasite interactions. Trends in Parasitology 25:236–244.

Associate Editor: Yannis Michalakis
Editor: Susan Kalisz