Draft

Research

Subject Areas:
ecology, evolution, theoretical biology

Keywords:
host-pathogen interactions, plant resistance genes, frequency-dependent selection, trans-species polymorphism

Author for correspondence:
Joy Bergelson
e-mail: jb7684@nyu.edu

Evolutionary implications of plant host interactions with a generalist pathogen

Hanna Märkle[1,2] and Joy Bergelson[1,2]

[1]Center for Genomics and Systems Biology, New York University, 12 Waverly Place, New York, NY, 10003, USA
[2]Center for Genomics and Systems Biology (CGSB), New York University Abu Dhabi, PO Box 129188, Abu Dhabi, UAE

It is widely believed that eco-evolutionary feedbacks arising from host-pathogen interactions shape the number and frequency of resistance (R) genotypes and the allelic polymorphism that they harbor. A subset of R genes exhibit unusually strong signatures of balancing selection, sometimes even existing as trans-specific polymorphism. Here, we explore the role of alternative hosts on R gene evolution through a simple model of two closely related host species that share a single generalist pathogen. We ask (i) how shared interactions determine the R gene repertoires and polymorphism in each host and (ii) under which circumstances trans-specific polymorphism is maintained. Our results indicate that interactions with a generalist pathogen are more likely to sustain polymorphism at a shared ancestral R gene compared to the maintenance of polymorphism at R genes private to each host. The former often translates into trans-specific resistance gene polymorphism. Further, we observe that increasing the relative proportion of a single host species favors fixation of private resistance in the less common host, as the pathogen tracks the more common host. Thus, our model sheds light on how R gene dynamics are shaped by interactions with a shared generalist pathogen and the pathogen's response to unequal host frequencies.

© The Authors.

1. Introduction

Ecological interactions affect not only the short-term population dynamics of interacting species, but also the long-term evolution of genes at the interface of these interactions. This is especially true for species that interact strongly, such as hosts and their pathogens. Host-pathogen systems exhibit particularly interesting eco-evolutionary dynamics in that allele frequencies and population sizes cycle as epidemics rage and wane (1; 2; 3). As a result, host-pathogen systems offer some of the most compelling evidence for natural selection to date. This has led to broad interest in the ecological and population genetic dynamics of host-pathogen systems, beginning with the seminal work of Anderson and May 1982 (4; 5) and continuing to the current day (6).

Although there is growing evidence that community context can be an important driver of host-pathogen evolution (7), models have typically focused on pairs of host and pathogen interacting in isolation. However, species do not interact in isolation of the community in which they are embedded. For example, disease pressure imposed on members of grassland communities is determined not only by the host's relative abundance, but also by the abundance of closely related host species (8). The impact of other community members stems from their roles as alternative hosts for generalist pathogens. The impact of community context is likely widespread given that numerous pathogen species interact with multiple host species, especially those that are closely related (9; 10).

Another example of ecological context affecting host-pathogen interactions comes from the demonstration that strains of the generalist pathogen, *P. syringae*, are more likely to be recognized by local genotypes of its host, *A. thaliana* (11) than by allopatric hosts. The apparent maladaptation of *P. syringae* has been hypothesized to be the result of community-level dynamics driven by ecologically dominant alternative hosts. In particular, if generalist pathogens are most strongly selected to adapt to highly abundant and long-lived hosts, other community members, such as small and ephemeral hosts like *A. thaliana*, may have the opportunity to adapt to recognize a pathogen, without counter-selection for the pathogen to escape recognition. Taken together, these studies underscore the importance of incorporating multi-species interactions into host-pathogen coevolutionary models.

Here, we ask how the interaction of a generalist plant pathogen with two closely related hosts shapes their resistance (*R*) gene evolution. Plant *R* genes play a crucial role in effector-triggered immunity, a key component of the plant immune system (12), by recognizing small molecules, known as effectors, that bacterial and fungal pathogens secrete into the plant cell to promote their proliferation. Individual hosts have variably sized *R* gene repertoires (13) with widespread signatures of balancing selection at individual loci (1; 14; 15; 16; 17); that is, these loci can maintain multiple alleles for millions of generations, sometimes even predating the divergence of related host species. The existence of such trans-specific polymorphisms has been observed not only in plant *R* genes (18; 19), but also in immune genes of humans (20; 21) and *Daphnia* (22).

The long-term persistence of alternative *R* gene alleles has proven difficult to explain empirically. One intensively studied plant *R* gene, *Rps5*, exists as a stably maintained presence-absence polymorphism in *A. thaliana* (14) and has measurable costs and benefits (23; 24). Efforts to synthesize the fitness effects of *Rps5* and its associated effector with their underlying abundances found that this long-lived polymorphism was unlikely to be explained if *A. thaliana* and *P. syringae* were modeled in isolation (23). Instead, diffuse interactions involving several host species, as well as pathogens with multiple non-homologous effectors, could explain the observed polymorphism patterns. This highlights that other ecological members of the community must be considered to explain the ancient balanced polymorphism.

These observations led us to ask how interactions of a generalist pathogen with two closely related plant species sharing some, but not all, *R* genes determine the maintenance of *R* alleles and *R* gene polymorphism at private and shared loci. We were particularly interested in whether alternative hosts sharing a particular *R* gene promote the maintenance of trans-species polymorphism at plant resistance gene. Our intuition was that the combination of shared and private resistance gene polymorphisms could generate unique niches for generalist pathogen genotypes varying in effector repertoires. We further hypothesized that the generalist's interactions with two separate hosts could function as a way to decouple the pathogen genotype frequency cycles from cycles in the hosts. Finally, we consider the possibility that a higher frequency of one of the two hosts could result in the pathogen more closely tracking genotype frequencies in that host and thus diffuse the selective pressure on the less abundant host. In this case we expect the private resistance gene in the less abundant host to become more beneficial, as this resistance allele is not present in the more common host and thus, the pathogen less frequently encounters it. As a result we expect that pathogen genotypes tend to maintain the corresponding, but recognized, effector, as (i) it provides a benefit on genotypes of the more common host and (ii) it does not have any cost. To address this, we consider a model with three *R* gene loci - one that is ancestral to and shared between two hosts and two that are private. Though still a simple model, our simulations provide a first glimpse at the importance of an alternative host in shaping the evolution of *R* genes in *A. thaliana*.

2. Methods and Material

(a) A two host-generalist pathogen gene-for-gene (GFG) model

We incorporate two phylogenetically related host species H and M interacting with a single generalist pathogen P into an existing multi-locus gene-for-gene (GFG) model (25; 26). Hosts have relative proportions ϕ (for host H) and $(1-\phi)$ (for host M). We base our model on the two simple assumptions that (i) there is no competition between the two host species, as we are considering ephemeral hosts, and (ii) there is no preference of the pathogen for either host, beyond that certain pathogen genotypes are better at infecting certain host genotypes based on their *R* gene repertoires (see below).

Host genotypes Our two host species are characterized by a total number of three bi-allelic resistance loci (L_1, L_2, L_3), with the allelic states susceptible S (0) or resistant R (1). We use the notation $\boldsymbol{x} = \{x_1, x_2, x_3\}$ to denote a genotype in host species H and $\boldsymbol{u} = \{u_1, u_2, u_3\}$ to denote a genotype in host species M. The resistance locus L_3 is shared by the two hosts, predating their evolutionary split (henceforth called the shared locus), whereas locus L_1 is private to host M (hence $x_1 = 0$ for all host H genotypes) and locus L_2 is private to host H (hence $u_2 = 0$ for all host M genotypes). Our chosen locus structure is inspired by the notion that *R* genes are often localized in clusters (13; 27; 28). These clusters can be highly dynamic, with frequent occurrence of tandem gene duplications and gene conversion events (29; 30). In our specific case, we assume that independent tandem duplications in each host species result in the evolution of a new *R* gene in each of our two host species.

Pathogen genotypes The pathogen is characterized by three bi-allelic effector loci (E_1, E_2, E_3) with allelic states Avirulent A (0) or virulent V (1). We use the notation $\boldsymbol{y} = \{y_1, y_2, y_3\}$ to denote a pathogen genotype. Note that here we use the term Avirulent (A) to refer to an effector allele that can be recognized by the cognizant resistance allele, but promotes disease in the absence of the corresponding resistance allele (31). We are not, here, concerned with the amount of damage caused to the host.

Interaction between host and pathogen loci Rooted in the assumption of classic multi-locus gene-for-gene (GFG) interactions (25; 26; 32), we assume a strict one-to-one correspondence between each bi-allelic pathogen and host locus; that is L_1 interacts with E_1, and so forth. In other words, a pathogen effector allele at a given effector locus can only be recognized by a resistance allele at the matched, cognizant resistance locus.

There have been some discrepancies in how molecular interactions between *R* genes and effectors are represented in an interaction matrix, as required by the GFG model (33). Here, we will use the following notation for a GFG-interaction matrix:

$$\begin{array}{cc} & \begin{array}{cc} \mathrm{A} & \mathrm{V} \end{array} \\ \begin{array}{c} \mathrm{R} \\ \mathrm{S} \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \end{array},$$

where 1 designates the ability of a pathogen to escape recognition and 0 designates recognition of the pathogen allele (31), which results in decreased growth on the given host.

For instance, using the *Rps5* example, the resistance (R) allele would represent plants carrying the *Rps5*-resistance allele that recognizes strains harboring the effector *avrPphB* (the Avirulent A allele in our notation) and the (S) allele would correspond to plants lacking *Rps5* and thus unable to recognize strains harboring the *avrPphB* effector. Despite carrying the *Rps5* resistance allele, hosts will fail to detect pathogen strains that (i) have a sequence variant of the *avrPphB* effector or (ii) lack *avrPphB* (corresponding to a virulent V allele in our notations).

We use the following set of differential equations to describe changes in host and pathogen genotype frequencies over time:

$$\begin{aligned} \frac{df_H(\boldsymbol{x})}{dt} &= \left(\frac{w_H(\boldsymbol{x})}{\bar{w}_H} - 1\right) f_H(\boldsymbol{x}) \\ \frac{df_M(\boldsymbol{u})}{dt} &= \left(\frac{w_M(\boldsymbol{u})}{\bar{w}_M} - 1\right) f_M(\boldsymbol{u}) \\ \frac{df_P(\boldsymbol{y})}{dt} &= \left(\frac{w_P(\boldsymbol{y})}{\bar{w}_P} - 1\right) f_P(\boldsymbol{y}) \end{aligned} \tag{2.1}$$

where $w_S(\boldsymbol{z})$ is the fitness (life-time reproductive success) of genotype $\boldsymbol{z}$ in species S, $f_S(\boldsymbol{z})$ is the corresponding frequency of the genotype and $\bar{w}_S$ is the mean fitness of species S.

Genotypic costs It has been found that resistance and virulence can result in genotypic costs that reduce life-time reproductive success (fitness) as a result of a trade-off between the allocation of resources to reproduction versus resistance or virulence. Theoretical results have further shown that the cost of resistance is a necessary but not sufficient condition for maintaining *R* gene polymorphisms in simple models of one host interacting with one pathogen (34). Empirical studies measuring costs of resistance in plants have found mixed evidence for its extent (reviewed in 35). On the positive side, comparisons of isogenic *A. thaliana* lines with and without *Rps5* (23) or with and without *Rpm1* (36) found costs ranging between 5%-10% in the absence of the pathogen. However, no differential fitness was detected between pairs of isogenic lines with resistant or susceptible *Rps2* sequence variants (37). Results from *Rps2* led to the suggestion that pleiotropic costs and genomic architecture interact to determine the cost of resistance. *Rps2* seems to segregate as alternative, functional alleles whereas *Rpm1* and *Rps5* exist as indel polymorphisms. This observation furthermore suggested that indel polymorphisms might be a specific architecture that results from alleles harboring higher costs of resistance. In addition to a role of genetic architecture, the cost of resistance is prone to genotype-by-environment interactions (35; 38), as has also been found for non *R* gene forms of resistance. Genotype by environment interactions make costs more difficult to quantify(35; 39).

Effectors that escape recognition can similarly exhibit a cost that is manifest in terms of decreased aggressiveness/growth compared to strains carrying the effector allele that is recognized by a cognizant R gene. The extent of this cost varies between effectors, just as has been found for plant R genes. Bai et al (2000) used marker exchange mutants to study the genotypic costs of virulent alleles of the two effectors *avrXa7* and *avrXa10* in the *Xanthomonas oryzae pv. oryzae* strain PX086. While the virulent avrXa7 $^-$ allele reduced growth compared to avrXa7$^+$ (Avirulent) on a rice cultivar lacking the corresponding resistance gene, there was no such cost for the avrXa10$^-$ (virulent) compared to the avrXa10$^+$ (Avirulent) allele.

Fitness expressions In the following subsection, we derive fitness expressions for all genotypes. We present the equations for host H, but the same principles apply to host M. Following (26), we calculate the probability $Q_{\boldsymbol{xy}}$ that host H genotype $\boldsymbol{x}$ gets infected by pathogen genotype $\boldsymbol{y}$ as a function of the number of effective resistance alleles d_{xy}:

$$Q_{\boldsymbol{xy}} = \sigma^{d_{xy}} \text{ with } \sigma \in [0, 1]$$

A host resistance allele at locus k is effective if the interacting pathogen carries the corresponding Avirulence (A) allele and hence, $d_{xy} = \sum_{i=1}^{3} = x_i \cdot (1 - y_i)$.

Table 1. The number of effective resistance alleles d in all possible pairwise interactions between host genotypes (columns; host H left, host M right) and pathogen genotypes (rows). The probability of infection in the interactions is calculated as σ^d

	host H genotype				host M genotype			
pathogen genotype	000	010	001	011	000	100	001	101
000	0	1	1	2	0	1	1	2
100	0	1	1	2	0	0	1	1
010	0	0	1	1	0	1	1	2
001	0	1	0	1	0	1	0	1
110	0	0	1	1	0	0	1	1
101	0	1	0	1	0	0	0	0
011	0	0	0	0	0	1	0	1
111	0	0	0	0	0	0	0	0

We incorporate the negative fitness effect of infection for the host with the parameter β_H and model the positive effect of infection for the pathogen with the parameter β_P.

Costs of resistance and virulence are incorporated as a function of the proportion of loci with a resistance (or virulence) allele. The cost functions are characterized by two parameters $c_S^{(1)}$ and $c_S^{(2)}$ for each species S. $c_S^{(1)}$ depicts the maximum loss in fitness when host (or pathogen) genotype has all possible resistance (or virulence) alleles and is referred to as the maximum cost of resistance (or virulence). The parameter $c_S^{(2)}$ governs the shape of the function that links the number of resistance (virulence) alleles to the corresponding genotypic cost. If $c_S^{(2)} < 0$, these costs are decelerating, meaning that the first resistance (or virulence) allele is comparatively expensive, but the addition of subsequent alleles becomes progressively less expensive until the

maximum cost is reached. On the other hand, costs are accelerating for $c_S^{(2)} > 0$, where the first allele is comparatively inexpensive and each additional allele becomes more expensive. The cost of resistance (or virulence) reduces the fitness of a given genotype in the absence of the interacting partner and is modeled by Equation A 1.
Therefore, the fitness expressions can be written as follows:

$$
\begin{aligned}
w_H(\boldsymbol{x}) &= C_H(|\boldsymbol{x}|)exp\left[-\beta_H \sum_{\boldsymbol{y}} Q(\boldsymbol{x},\boldsymbol{y}) f_P(\boldsymbol{y})\right] \\
w_M(\boldsymbol{u}) &= C_H(|\boldsymbol{u}|)exp\left[-\beta_H \sum_{\boldsymbol{y}} Q(\boldsymbol{u},\boldsymbol{y}) f_P(\boldsymbol{y})\right] \\
w_P(\boldsymbol{y}) &= C_P(|\boldsymbol{y}|)exp\left[\beta_P \left(\phi_1 \sum_{\boldsymbol{x}} Q(\boldsymbol{x},\boldsymbol{y}) f_H(\boldsymbol{x}) + \phi_2 \sum_{\boldsymbol{u}} Q(\boldsymbol{u},\boldsymbol{y}) f_M(\boldsymbol{u})\right)\right]
\end{aligned}
\tag{2.2}
$$

A detailed overview of variables, parameters and functions can be found in Table 2.

Simulation of the model We extended the source code from (26) to simulate the dynamics of our system with two host species and the generalist pathogen. We first run our model assuming equal proportions of both hosts ($\phi = 0.5$) for 10,000 random combinations of maximum cost of resistance ($c_H^{(1)}$), maximum cost of virulence ($c_P^{(1)}$) and initial genotype frequencies, while fixing the shapes of the cost functions to $c_H^{(2)} = c_P^{(2)} = 3$ (accelerating costs). Maximum costs of resistance and virulence for individual simulations were independently drawn from a uniform distribution $U(0.01, 0.3)$. We randomly initialized the genotype frequencies within a single species, by first drawing a random number for each genotype from a uniform distribution $\mathcal{U}(0,1)$ and subsequently normalizing these numbers so that they sum to 1. Each simulation was run until $T_{max} = 100,000$ or until genotype frequencies reached a stable equilibrium.

For each simulation, we determined (i) the genotypes present at the end of the simulation, (ii) the maximum number of resistance alleles maintained in any host genotype that persisted, (iii) the number and identity of R-alleles maintained in each host population, (iv) the polymorphism status of each resistance gene and (v) the genotypes still present in the pathogen population. Next we repeated the same procedure for increasing proportions of host H; that is, for $\phi \in \{0.7, 0.9, 1\}$. Finally, we aggregated results across nine different pairwise combinations of the shapes of the cost functions for the host ($c_H^{(2)} \in \{-3, 0, 3\}$) and the pathogen ($c_P^{(2)} \in \{-3, 0, 3\}$), for four different proportions of host H. Therefore, we jointly analyzed a total of 9 (cost shape combinations) x 4 ($\phi \in \{0.5, 0.7, 0.9, 1\}$) x 10,000 (random combinations of $c_H^{(1)}$, $c_P^{(1)}$ and genotype frequencies) = 360,000 simulations.

3. Results

(a) Frequency-dependent selection maintains a subset of host and pathogen genotypes with fluctuating frequencies.

We start by presenting results for an equal proportion of the two hosts ($\phi = 0.5$) and for accelerating fitness costs ($c_H^{(2)} = c_P^{(2)} = 3$). In this scenario, each pathogen genotype responds equally to the genotype frequencies in both hosts and to their capacities to resist infection. Host-pathogen interactions drive dynamic reciprocal genotype frequency fluctuations that result in the loss of several genotypes (Figure 1).

In the specific case ($c_H^{(1)} = c_P^{(1)} = 0.1$), all host genotypes with exactly one resistance allele

(shared or private) and all pathogen genotypes with exactly two virulence alleles are maintained (Figure 1ef). This is the result of trade-offs between the burden of infection and the costs of resistance in the hosts, and between costs of virulence and the ability to escape recognition in the pathogen. Pathogen genotypes with two or three Avirulence alleles ($P_{000}, P_{100}, P_{010}, P_{001}$) suffer from lower infection efficiency as they can be recognized by at least two host genotypes. This fitness loss outweighs the fitness advantage of having more Avirulence alleles. Therefore, these genotypes go extinct very early in the simulation (Figure 1ab). The opposite applies to the pathogen genotype P_{111}, which contains three virulence alleles. Although this genotype has the highest overall infection efficiency, it also bears the highest cost of virulence. Somewhat surprisingly, this pathogen genotype is not favored over pathogen genotypes with two virulence alleles for the simple reason that all of the host genotypes with two resistance alleles are lost early in the dynamics; and as a consequence, pathogen genotypes with two virulence alleles reach almost the same infection efficacy P_{111} as pathogen genotypes with three virulence alleles, but at a lower genotypic cost. Completely susceptible host genotypes (with no resistance alleles; H_{000} and M_{000}) rapidly go extinct due to the reduced fitness that they suffer from constant full strength infection.

Once only genotypes that maximize the fitness trade-off remain (around $t = 700$), negative-frequency dependent selection drives fluctuations in their frequencies. Whenever the frequency of the P_{110} pathogen increases, increases in the frequencies of H_{001} and M_{001} (Figure S2a,b) follow, as these are the only host genotypes able to recognize the P_{110} pathogen . Simultaneously, host genotypes H_{010} and M_{100} decrease in frequency (Figure S2c,d). As a result of the increases in frequency of H_{001} and M_{001}, the frequency of P_{110} slowly decreases and the frequencies of P_{011} and P_{101} correspondingly increase. This in turn triggers a decrease in H_{001} and M_{001} and a corresponding increase in H_{010} and M_{100} genotypes. Note that these changes are not initially synchronized between the two hosts. Over time, dynamics synchronize and fluctuations dampen (compare Figure 1a,b,c).

(b) Both costs of resistance and virulence determine the number of resistance alleles maintained.

Next, we tested for the sensitivity of these dynamics to (i) the maximum cost of resistance ($c_H^{(1)}$), (ii) the maximum cost of virulence ($c_P^{(1)}$) and (iii) the initial genotype frequencies for fixed shapes of the genotypic cost functions ($c_H^2 = c_P^{(2)} = 3$). We ran 10,000 simulations for random combinations of the three factors and found that the maximum number of resistance alleles maintained within a single host genotype decreases with an increase in the cost of resistance (Figure 2a). However, the maximum number of resistance alleles in a single genotype depends not only on the cost of resistance but also on the cost of virulence. That is, for a given maximum cost of resistance, we find that the maximum number of resistance alleles maintained within a single genotype generally increases with increasing costs of virulence (Figure 2b). These results highlight that host genotypic costs and pathogen genotypic costs jointly determine the number of resistance alleles that persist in the repertoire of a single host genotype.

(c) The less abundant host preferentially maintains the resistance allele at the private locus

We next assessed how the relative proportion (ϕ) of host species H interacts with the costs of resistance and virulence to determine resistance dynamics. Within this context there are two related but different aspects to be considered: (i) which resistance alleles are maintained and (ii) whether they become fixed or are maintained as a balanced polymorphism. As we consider a haploid model, the latter implies that there are at least two genotypes maintained in the population that differ in their allelic state at a given locus. We begin by presenting results on

the maintenance of resistance alleles and consider the maintenance of polymorphism in the next subsection.

When both hosts are equally abundant, private and shared resistance alleles are maintained for a wide range of cost combinations (Figure 3a and b) in both host species. This tolerance to diverse cost functions persists in the more abundant host H (increasing ϕ) as its relative frequency increases, although the proportion of parameter space in which both resistance alleles are lost expands. In contrast, the host species M tends to retain only the private resistance allele and tends to lose the shared resistance allele. These patterns are considerably robust across shape functions (Figure 4a). An additional interesting observation is that there are only a few cases where the shared resistance allele persists alone, in the absence of the private resistance alleles (light blue range in Figure 4a). In the special case $\phi = 1$ (the pathogen interacts with only host species H), we observe equal proportions of simulations in which only the private or only the shared, ancestral gene is maintained. This is a consequence of the two shared and private resistance loci becoming equivalent once the pathogen no longer interacts with host M.

The observed results hint at an effect of the initial genotype frequencies on the final outcome of the interaction. Indeed, we observe that for repeated simulations with fixed cost parameters and shapes, which genotypes fix or are maintained change with initial genotype frequencies (Figure S5).

(d) Fixation of the private resistance allele is favored in the less common host

From an evolutionary perspective, it is not only important which resistance loci persist, but also if they maintain a polymorphism rather than fixing in the population. When hosts are equally frequent, polymorphism is observed for both resistance genes (private and ancestral) for a wide range of cost combinations (see Figure 3c for relative proportions and Figure 3d for the specific dependence on maximum costs). The polymorphisms in the private loci are lost when the cost of resistance is low and the cost of virulence is intermediate to high (see Figure 3b and Figure 3d). The polymorphism at the shared gene, however, persists in both hosts under these same conditions.

As the proportions of the two hosts becomes more skewed, we find that there is a loss of polymorphism at both private and shared resistance genes in the less common host H_2 (Figure 3d bottom left to right) because the private locus becomes fixed and the shared resistance allele is lost (see Figure 3b bottom left to right). The frequency of this outcome in host M increases with increasing presence of host H (increasing ϕ) (see Figure 3b,d). In contrast to the minor host, the more common host species H maintains polymorphism for both resistance genes for a wide range of parameter combinations (Figure 3d).

(e) Polymorphism is more likely to be maintained at the shared relative to private loci, promoting trans-specific polymorphism.

Across a range of combinations of the maximum costs of resistance and virulence, and shapes of corresponding cost functions, we find frequent instances of polymorphism at the ancestral resistance gene, especially for the more common host species (Figure 4b), but often as a trans-specific polymorphism (Figure 6). In particular, we find that for the more abundant host species H, the most frequent outcome is the maintenance of polymorphism at both resistance genes (private and ancestral) (Figure 4b). Interestingly, polymorphism exclusive to the private resistance genes is much more rarely observed than polymorphism that is exclusive to the shared resistance

gene. Remarkably, we noticed the maintenance of trans-specific polymorphism, even with an increasing imbalance in the frequencies of the two host species (increasing ϕ).

(f) The pathogen adapts to the more common host species.

As the fitness of pathogen genotypes is affected by the relative proportion of both host species (through the parameter ϕ), we asked how co-evolutionary dynamics change as the relative abundance (ϕ) of host H increases. A frequency increase ($\phi > 0.5$) of host H can be interpreted in two ways: either (i) host H has a higher abundance compared to host M, or (ii) there is some direct or indirect (i.e. through a transmitting vector) preference of the pathogen for host M. We were specifically interested in whether we would observe a shift in the pathogen genotypes that are maintained as a result of more frequent interactions with host H. Averaging over different combinations of the maximum costs $c_H^{(1)}$ and $c_P^{(1)}$, we find a decrease in the probability that a pathogen genotype with full virulence (P_{111}) persists (blue colors) as interactions with host H become more frequent (Figure 5c; from left to the right). This reduction in the maintenance of the fully virulent pathogen is accompanied by an increase in the probability that genotypes with two virulence alleles are maintained (pink/red colors).

In the host, we detect some contrasting patterns in host H compared to host M. In host H the parameter space where genotypes with zero (H_{000}) and one (H_{010} and/or H_{001}) resistance allele co-exist expands as its frequency increases (Figure 5a), whereas in host M the parameter space where genotypes with exactly one resistance allele coexist increases as the frequency of host H increases. In many instances, retention of genotypes with only one resistance allele in host M corresponds to fixation of the genotype with the private resistance allele (M_{100}). Taken together, our results suggest that as the pathogen interacts more frequently with host H, it starts to track more closely the genotype frequencies in the common host and becomes concentrated in genotypes that maximize fitness on this host.

4. Discussion

Understanding the proximate causes that maintain high levels of diversity and polymorphism at plant resistance genes has puzzled researchers for a long time. Increasing evidence points to the importance of the community context in which these interactions occur. In line with these results, we show that interactions of closely related species with a shared generalist contribute to the maintenance of resistance genotypes and resistance gene polymorphism. Our results further highlight that evolutionary dynamics at shared and private resistance genes are contingent on the relative abundances of the host species, along with the genotypic costs of both resistance and virulence. Therefore, our model reveals a role of species-specific trade-offs in shaping reciprocal evolutionary dynamics. In addition, our results help explain how trans-species *R* gene polymorphisms persist as a consequence of interactions with a shared generalist pathogen.

The evolution of *R* gene repertoires depends on host-specific and pathogen specific fitness trade-offs. In line with previous findings (32; 40), we show that the evolutionary dynamics of *R* genes are determined not only by the cost of resistance, but also by the cost of virulence in the pathogen. Specifically, previous theoretical results have shown that the equilibrium frequency of the resistance allele in a simple one locus GFG model may depend solely on the cost of virulence (32; 40). Thus, our results underscore that an integrative understanding of host-specific and pathogen-specific trade-offs is required to understand the evolutionary dynamics of resistance genes and repertoires.

A better understanding of cost functions in both hosts and pathogens is required for a general understanding of pathogen-host dynamics. We observed variable evolutionary outcomes depending on the specific costs of resistance and virulence, and the shapes of the functions that link the number of R alleles/virulent effectors to these costs. This highlights the need for experiments that will generate a more refined understanding of these costs and how they accumulate as members of resistance/effector repertoires are added to a genome.

Measuring the gene-specific costs of resistance and virulence has proven to be challenging, especially as these costs are likely to depend on the ecological and/or genomic context. A meta-analysis by Bergelson and Purrington (39) found mixed evidence for the presence of costs associated with various types of resistance. Studies revealing that single presence/absence polymorphisms can incur costs as high as 10% (23; 36) have raised questions about how resistance gene repertoires, with up to hundreds of *R* gene, can be sustained. One possible explanation was offered by MacQueen et al. (37) who showed that genomic architecture and pleiotropy might alleviate costs of resistance. While these studies offered useful insights into the extent to which individual costs vary, it still remains an open question how the composite cost associated with a specific *R* gene repertoire is shaped by its constituent members.

In the theoretical literature, the topic of cost functions has been considered. Tellier and Brown (32; 34) argued that pathogen virulence might be best described by accelerating costs due to their important role in establishing and promoting disease, and the known redundancy in effector function. In the presence of redundancy, the loss of one effector may be compensated for by an alternative, functionally redundant effector. However, as more effectors are lost, the chance that another effector can compensate by performing the necessary function is reduced. This suggests that the marginal cost associated with loss of virulence alleles should increase as the number of effectors decreases. On the other hand, Tellier and Brown (32; 34) reasoned that the cost of resistance should be decelerating because R proteins often converge to the same downstream processes and hubs. Nevertheless, accelerating costs in the host remain a possibility given that an increase in the number of resistance alleles should increase the chance of incompatible resistance alleles that trigger hybrid necrosis(41; 42; 43). Indeed, results from a recent theoretical study (44) suggest that accelerating cost-benefit trade-offs between resistance and fecundity may be a valid first assumption for various forms of first order and second order epistatic interactions.

To our knowledge there are only a few empirical studies that characterize the specific shape of the cost functions. Huang et al. (2017) (45) found evidence for a decelerating growth-defense trade-off in a bacterial prey when interacting with a co-evolved ciliate predator but an accelerating trade-off when interacting with the ancestral ciliate predator.

In a study of *P. infestans* isolates grown on a potato cultivar lacking all corresponding resistance genes, Montarry et al (2010) (46) found a negative correlation between the number of virulence factors and both aggressiveness components and overall fitness; this relationship was best characterized as linear or accelerating. Finally, Wichmann and Bergelson (47) showed that sequentially mutating the functional effectors *avrBs2*, *avrBs3*, *avrBs1* and *avrBs4* (virulent) in the *Xanthomonas axonopodis* pv. *vesicatoria* strain Xav19 decreased the emergence of the corresponding mutant strains in field conditions, with often additive effects but a dominant, and epistatic, role of *avrBs2* on the other effectors.

Pathogen maladaptation to ephemeral hosts is a result of the pathogen adapting to the more common host species. One commonly held prediction in the local adaptation literature (48; 49) is that pathogens tend to be ahead of the host in co-evolutionary interactions due to their shorter generation times and higher mutation rates. By investigating patterns of resistance of *A. thaliana* accessions to resident and non-resident strains of the generalist pathogen *P. syringae*, Kniskern et al. (11) found significantly higher levels of resistance to local versus non-resident pathogen strains. They suggested that this may a consequence of pathogens adapting to more commonly encountered hosts, such crops grown in the vicinity, thereby resulting in pathogen maladaptation

to its ephemeral host. Our results provide some evidence for this hypothesis. First, we observe a tendency in the less common host to fix for the private resistance allele. Second, and at the same time, we see an overall shift in the pathogen towards maintaining genotypes with a lower number of virulence alleles. In many cases, this corresponds to a loss of the virulence allele that escapes recognition by the private resistance allele in the less common (here, ephemeral) host. Thus, interactions of an ephemeral, selfing host such as *A. thaliana*, with a generalist pathogen, provide an additional mechanism to promote pathogen maladaptation besides (for alternative explanations, see 49; 50; 51).

Current limitations of the model While our model revealed an interesting mechanism that can promote the maintenance of polymorphism, we employed several simplified assumptions that warrant future investigation. Currently, our model is based on a very short repertoire that includes a maximum of two functional resistance genes per species and three effectors in the pathogen. In reality, species specific *R* gene repertoires range from few to thousands of *R* genes (52). In addition to accounting for the complexity of repertoire sizes, future research should address the known network structure of *R* gene and effector interactions that are far more complex than a set of strict one-to-one interactions (53; 54; 55; 56). For instance, effector recognition by R proteins can be mediated through direct interactions with the effector or indirect interactions mediated through effector-induced modifications of other host proteins. Studies have shown that several effectors can converge to the same target and these modifications may be monitored by different *R* genes (55; 57). The *Pseudomonas syringae* effectors AvrRpm1 and AvrB, for example, induce phosphorylation of the RIN4 protein which is sensed by the Rpm1 in *A. thaliana*. The effector AvrRpt2, on the other hand, cleaves RIN4, a modification that is recognized by Rps2. This example most likely only highlights a small subset of the plethora of interactions that may play out in nature.

Furthermore, we consider a very specific architecture of the resistance and effector genes, and do not account for reintroduction of previously extinct genotypes as a result of recombination. Our model best reflects a resistance gene cluster and a pathogenicity island in the pathogen (58; 59), where each host species has separately gained a new resistance gene through a process such as tandem gene duplication. However, *R* gene clusters in plants are known to be highly dynamic (30) and subject to both recombination and gene conversion (29; 60). The importance of recombination has been shown in a model (61) demonstrating that the evolution of broad-spectrum and specific-resistance genes is contingent on the amount of recombination. Recombination will additionally play a role in the generation of novel specificities (62), a topic not addressed by our current model. Investigating the effect of recombination is thus an important extension for future work.

Our model is based purely on genotype frequencies and neither incorporates the effect of genetic drift nor eco-evolutionary feedback (host-pathogen, host-host). Hence, we implicitly assume that the population sizes of the hosts and the pathogen are time-invariant and large. Yet, eco-evolutionary feedback can create direct frequency dependent selection, which generally promotes stable maintenance of polymorphism (5; 34; 63; 64; 65**?**). The incorporation of eco-evolutionary feedback can qualitatively alter the outcomes of host-pathogen co-evolution and expand the range of parameters supporting the maintenance of polymorphism (65). For this reason, the amount of polymorphism maintained in our model is likely conservative. On the other hand, genetic drift, (that is, random changes in allele frequencies), can alter allele frequency trajectories, sometimes even causing the loss of a deterministically stable polymorphism (66; 67).

The community context matters for the evolution of resistance genes Analyses of *R* gene polymorphism patterns have revealed diverse evolutionary histories (13; 17). In many cases, these histories are more consistent with trench-warfare dynamics (1), which are characterized by the rise and fall of allele frequencies, than arms-race dynamics (68) in which alleles are recurrently

fixed. Theoretical studies of simple host-pathogen interaction models have shown that (i) under a wide range of scenarios, polymorphism tends to be transient (reviewed in 34) and is promoted by (ii) direct frequency-dependent selection (dNFDS), whereby the fitness of an allele/genotype depends on its own frequency. The mechanisms generating dNFDS, including epidemiological dynamics and asynchronous host and pathogen life cycles, help maintain polymorphisms (34; 65). Diffuse host-pathogen interactions also help in maintaining the polymorphism by increasing the range of cost conditions in which *R* gene polymorphisms persist (23).

Here we further develop this idea by studying interactions of two closely related hosts with a generalist pathogen. Our motivation is to understand the evolution of *R* genes in *A. thaliana*, for which there are strong signatures of balancing selection at many *R* genes (1; 14; 15; 23; 69). *A. thaliana* frequently co-occurs with other hosts from the Brassicaceae family. These other plant species not only share similar life-history traits (70) but also serve as host to the same generalist pathogen, *Pseudomonas syringae*. The results of our model add a new explanation for how these ancient balanced polymorphisms could have arisen as a result of the context in which interactions between *A. thaliana* and its pathogens take place. The same idea could potentially help explain the maintenance of trans-specific polymorphism at immune genes underlying interactions in other systems, such as *Daphnia* and its pathogen *Pasteuria ramosa* (22).

Our results further demonstrate that *R* genes with new recognition specificities may be selectively favored if they protect against a generalist pathogen that also interacts with a more abundant host sharing a portion of its resistance gene repertoire. We also find that interactions with a generalist pathogen can dynamically drive species-specific extinctions of orthologous *R* gene specificities in one species while maintaining them in a related species. Last, our model suggests a potential mechanism that can maintain trans-species polymorphism in hosts with sympatric distributions and similar life-history traits. Thus, trans-specific polymorphisms may not require specific functions of the resistance gene but instead follow from the impact of ecological interactions with a shared pathogen on a set of orthologous and private *R* genes.

5. Conclusion

The evolutionary histories of plant *R* genes are impressively diverse. Understanding the processes underlying these evolutionary dynamics is an ongoing work in progress. Motivated by the accumulating evidence of the importance of community context, we presented a multi-locus gene-for-gene model of two related host species interacting with a generalist pathogen. Our results offer a new perspective on how shared resistance among closely related hosts can couple their resistance gene evolution and, as a result, shape their resistance gene repertoire evolution. We additionally found evidence that ephemeral hosts may have the co-evolutionary advantage of being rare in diffuse co-evolutionary interactions, and profit from pathogens tracking more abundant hosts. This provides an opportunity for new resistance specificities to become fixed in their population. Finally, we showed that trans-specific polymorphism may be promoted by interactions with a shared generalist pathogen and potentially be a more common outcome than previously thought. While our model is simplified in many aspects, we hope it will stimulate further theoretical developments that embrace the roles of community context and genomic context in driving the evolution of *R* gene repertoires.

Acknowledgements. The authors thank Mercedes Pascual and Enrique Rojas for helpful discussions. This work was supported by Tamkeen under the NYU Abu Dhabi Research Institute Award to the NYUAD Center for Genomics and Systems Biology (ADHPG-CGSB) to JB, and by support from the Simons Foundation to JB.

Data availability: All codes for simulating the model and data underlying the figures are available at `https://github.com/PopRGen/TSPolyR2.git`.

Draft 0000000

6. Figures & Tables

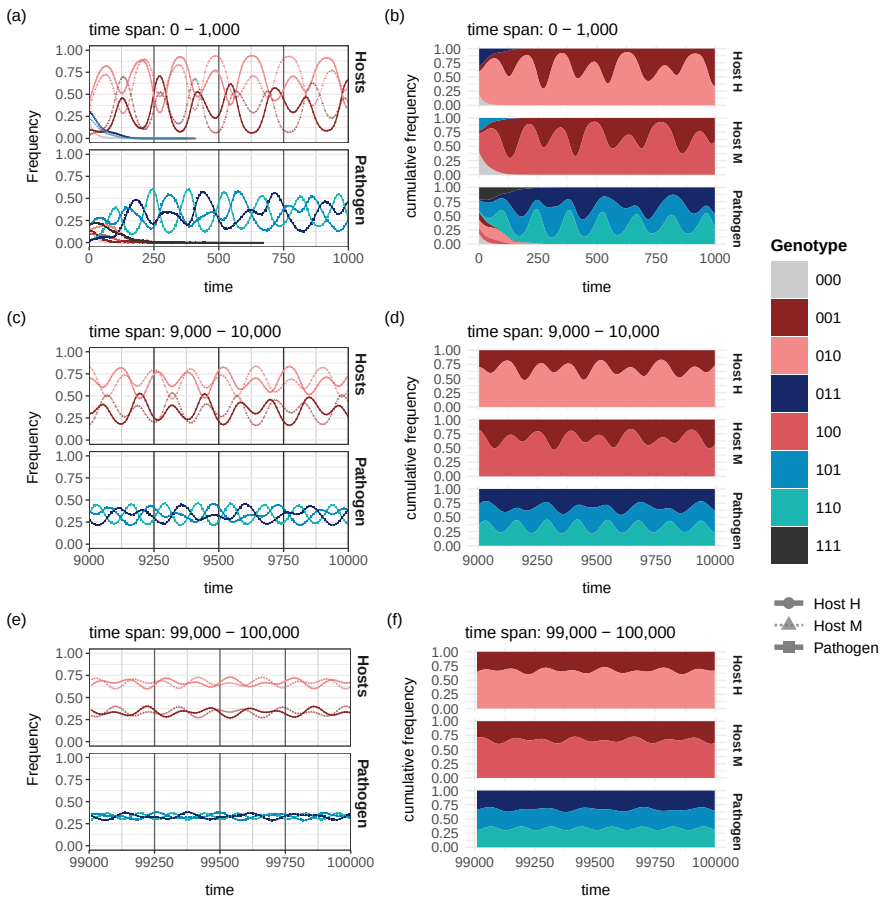

Figure 1. Frequency dynamics in different time intervals for a single simulation for $c_H^{(2)} = c_P^{(2)} = 3$, $c_H^{(1)} = c_P^{(1)} = 0.1$, $\phi = 0.5$, $\beta_H = \beta_P = 1$, seed $= 1,600$, $\sigma = 0.85$. On the top left the dynamics of all present genotypes are shown for time intervals: $0 - 1,000$, $9,000 - 10,000$ and $99,000 - 100,000$. The same dynamics are shown as a cumulative plot for each species on the left. Note that we use the same color codes for host and pathogen genotypes. For hosts 0 corresponds to the susceptible allele at the given locus and 1 reflects the corresponding resistance allele. For pathogen 0 represents the avirulence allele that can be recognized by a host with a corresponding resistance allele at the locus and 1 depicts a virulence allele at the locus which can escape recognition by both the susceptible and resistant allele at the given locus.

Draft 0000000

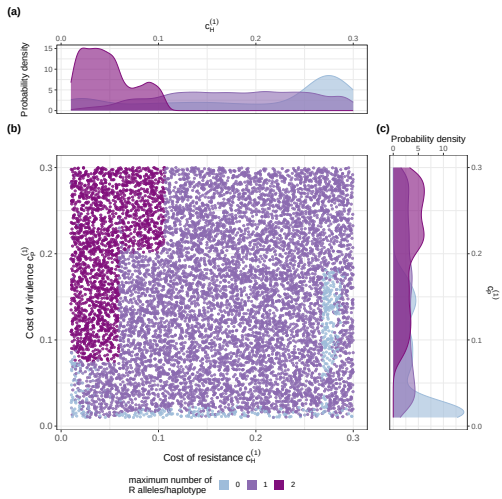

Figure 2. Maximum number of resistance alleles maintained in a single genotype for different combinations of the maximum costs of resistance ($c_H^{(1)}$) and virulence ($c_P^{(1)}$) based on 10,000 randomly drawn combinations of $c_H^{(1)} \in \mathcal{U}(0.01, 0.3)$ and $c_P^{(1)} \in \mathcal{U}(0.01, 0.3)$ and initial genotype frequencies. Note we here show the genotype with the highest number of resistance alleles maintained in host H (that is $\arg\max_{\boldsymbol{x}} \sum_{i=1}^{3} x_i$). This does not exclude the possibility that additional genotypes with fewer resistance alleles have been maintained. Subfigure A shows the probability density for different maximum numbers of resistance alleles (that is $\arg\max_{\boldsymbol{x}} \sum_{i=1}^{3} L_i$) per genotype for increasing costs of resistance $c_H^{(1)}$. Subfigure B show the the corresponding marginal probability densities for the cost of virulence $c_P^{(1)}$. Each point in subfigure C shows the correspoding information for a single random combination of $c_H^{(1)}$ (x-axis) and $c_P^{(1)}$ (y-axis). Points are colored according to the maximum number of resistance alleles found in any maintained host genotype (i.e if genotypes H_{010} and H_{011} are maintained the point would be colored in dark purple).

Draft 0000000

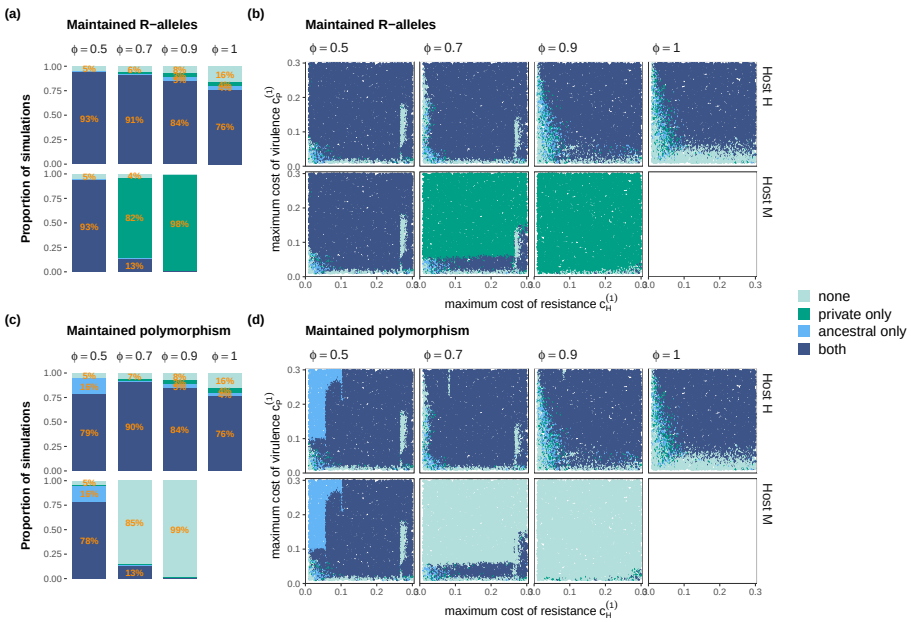

Figure 3. Maximum number of resistance alleles maintained in a single genotype for different combinations of the maximum costs of resistance ($c_H^{(1)}$) and virulence ($c_P^{(1)}$). Note we here show the genotype with the highest number of resistance alleles maintained in host H. This does not exclude the possibility that additional genotype with fewer resistance alleles have been maintained in a single simulation. Subfigure A shows the probability density for different maximum numbers of resistance alleles (that is argmax$\sum_{i=1}^{3} L_i$) per genotype for increasing costs of resistance based on 10,000 randomly drawn combinations of $c_H^{(1)} \in \mathcal{U}(0.01, 0.3)$ and $c_P^{(1)} \in \mathcal{U}(0.01, 0.3)$ and initial genotype frequencies.

Draft 0000000

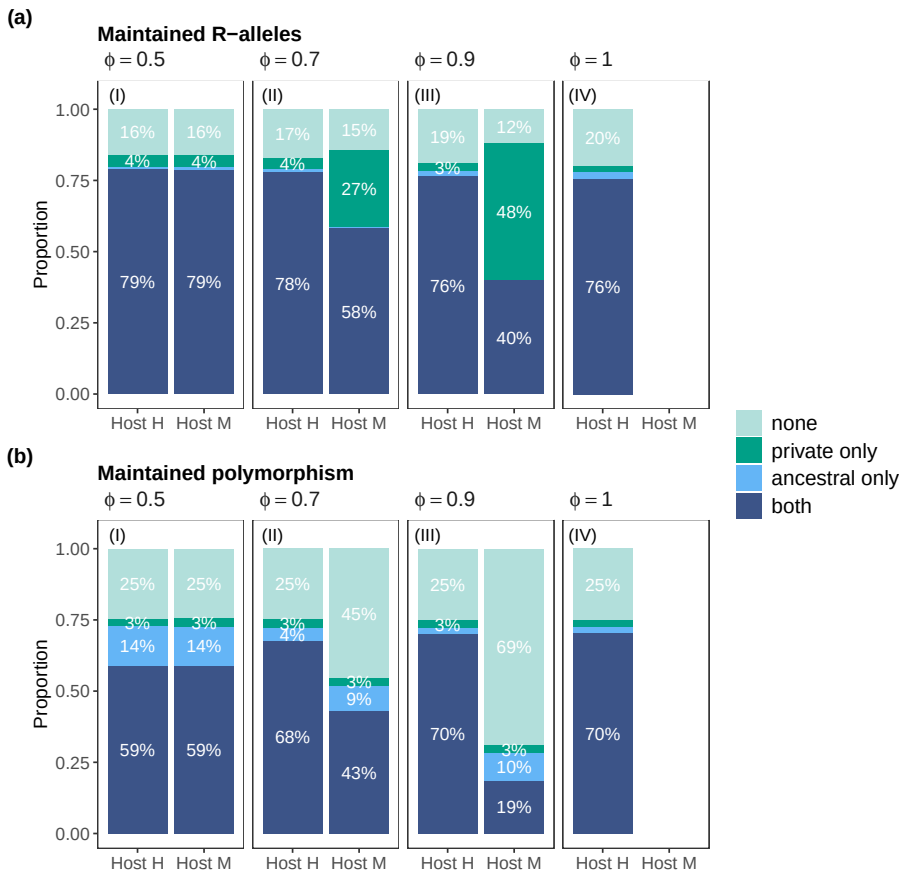

Figure 4. Patterns of (a) maintained resistance alleles and (b) polymorphism in hosts H and M for a wide range of cost combinations and shapes depending on the relative proportion ϕ of host H (from left to right). Each pair of stacked bars corresponds to a total of 90,000 simulations with 10,000 simulations each for all pairwise combinations of $c_H^{(2)} \in \{-3, 0, 3\}$ and $c_P^{(2)} \in \{-3, 0, 3\}$. The 10,000 simulations for each combination of $c_H^{(2)}$, $c_P^{(2)}$ and ϕ, have been run by randomly drawing the initial frequencies of all genotypes, $c_H^{(1)} \sim \mathcal{U}(0.01, 0.3)$ and $c_P^{(1)} \sim \mathcal{U}(0.01, 0.3)$. All simulations where summarized at $t = 100,000$.

Draft 0000000

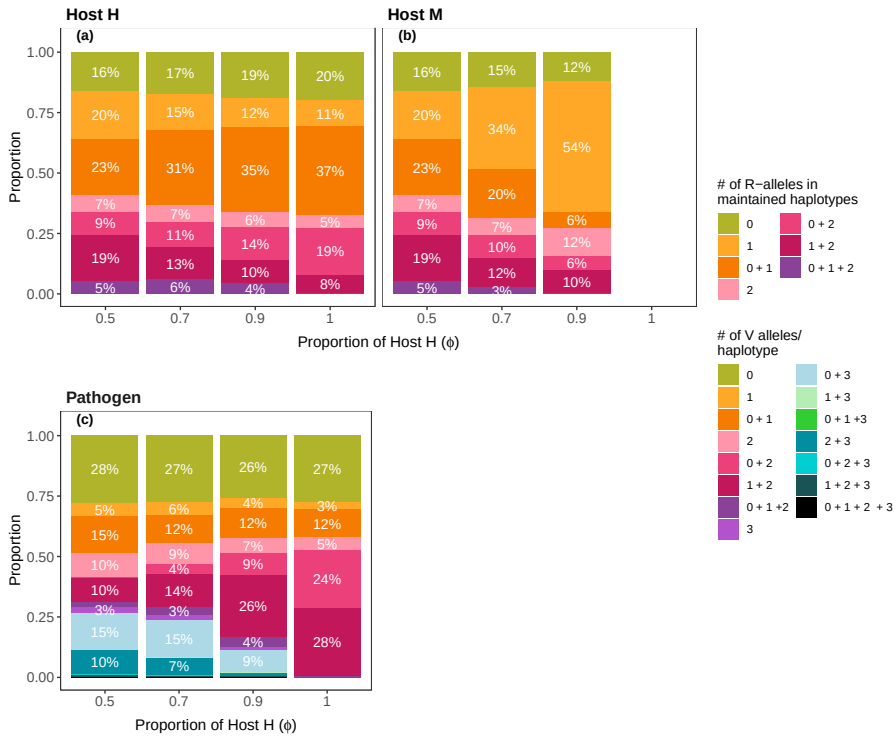

Figure 5. Maintained genotype combinations in host H, host M and the pathogen for a wide range of cost combinations and shapes depending on the relative proportion ϕ (x-axis) of host H. The legends are to be read as follows: If there is a single number it means that only genotypes with the corresponding number of resistance **R**/virulence **V**) alleles have been maintained in the given species; if there are several numbers combined with a '+', then genotypes with different numbers of resistance R/virulence V alleles are still present at the end of the simulation. Note that the color codes in the hosts and the pathogen are the same for combinations which can occur in hosts and pathogens. Also note that there might be several genotypes with a given number of R-alleles/V-alleles maintained.

Each stacked bars corresponds to a total of 90,000 simulations with 10,000 simulations each for all pairwise combinations of $c_H^{(2)} \in \{-3, 0, 3\}$ and $c_P^{(2)} \in \{-3, 0, 3\}$. The 10,000 simulations for each combination of $c_H^{(2)}$, $c_P^{(2)}$ and ϕ, have been run by randomly drawing the initial frequencies of all genotypes, $c_H^{(1)} \sim \mathcal{U}(0.01, 0.3)$ and $c_P^{(1)} \sim \mathcal{U}(0.01, 0.3)$. All simulations where summarized at $t = 100,000$.

Draft 0000000

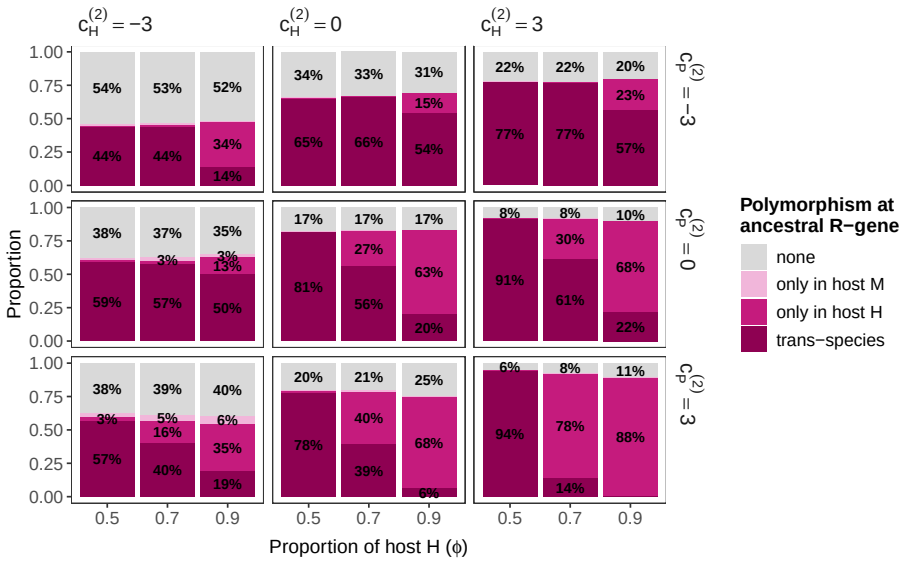

Figure 6. Proportion of simulations where polymorphism at the shared resistance gene is maintained for different shapes of the cost function in the host (columns) and pathogen (rows). Each subplot shows the results for increasing proportions ϕ of host species H. Each bar presents the results of 10,000 simulations with $c_H^{(1)} \sim \mathcal{U}(0.01, 0.3)$, $c_P^{(1)} \sim \mathcal{U}(0.01, 0.3)$ and random initialization of genotype frequencies.

References

1 Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. 1999 Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis.. *Nature* **400**, 667–71. Place: England.

2 Retel C, Kowallik V, Huang W, Werner B, Künzel S, Becks L, Feulner PGD. 2019 The feedback between selection and demography shapes genomic diversity during coevolution. *Science Advances* **5**, eaax0530. (10.1126/sciadv.aax0530)

3 Hall AR, Scanlan PD, Morgan AD, Buckling A. 2011 Host-parasite coevolutionary arms races give way to fluctuating selection: Bacteria-phage coevolutionary dynamics. *Ecology Letters* **14**, 635–642. (10.1111/j.1461-0248.2011.01624.x)

4 Anderson RM, May RM. 1982 Coevolution of hosts and parasites. *Parasitology* **85**, 411–426. (10.1017/S0031182000055360)

5 May RM, Anderson RM. 1983 Epidemiology and genetics in the coevolution of parasites and hosts.. *Proceedings of the Royal Society of London. Series B, Biological sciences* **219**, 281–313. Place: England.

6 Buckingham LJ, Ashby B. 2022 Coevolutionary theory of hosts and parasites. *Journal of Evolutionary Biology* **35**, 205–224. (10.1111/jeb.13981)

7 Evensen C, White A, Boots M. 2024 Multispecies interactions and the community context of the evolution of virulence. *Proceedings of the Royal Society B: Biological Sciences* **291**, 20240991. (10.1098/rspb.2024.0991)

8 Parker IM, Saunders M, Bontrager M, Weitz AP, Hendricks R, Magarey R, Suiter K, Gilbert GS. 2015 Phylogenetic structure and host abundance drive disease pressure in communities. *Nature* **520**, 542–544. (10.1038/nature14372)

9 Woolhouse MEJ, Taylor LH, Haydon DT. 2001 Population Biology of Multihost Pathogens. *Science* **292**, 1109–1112. (10.1126/science.1059026)

10 Barrett LG, Kniskern JM, Bodenhausen N, Zhang W, Bergelson J. 2009 Continua of specificity and virulence in plant host–pathogen interactions: causes and consequences. *New Phytologist* **183**, 513–529. (10.1111/j.1469-8137.2009.02927.x)

11 Kniskern JM, Barrett JM, Bergelson J. 2011 Maladaptation in wild populations of the generalist plant pathogen Pseudomonas syringae. *Evolution* **65**, 818–30. (10.1111/j.1558-5646.2010.01157.x)

12 Jones JDG, Dangl JL. 2006 The plant immune system. *Nature* **444**, 323–329. (10.1038/nature05286)

13 Van de Weyer AL, Monteiro F, Furzer OJ, Nishimura MT, Cevik V, Witek K, Jones JD, Dangl JL, Weigel D, Bemm F. 2019 A Species-Wide Inventory of NLR Genes and Alleles in Arabidopsis thaliana. *Cell* **178**, 1260–1272.e14. (10.1016/j.cell.2019.07.038)

14 Tian D, Araki H, Stahl E, Bergelson J, Kreitman M. 2002 Signature of balancing selection in Arabidopsis. *Proceedings of the National Academy of Sciences* **99**, 11525–11530. (10.1073/pnas.172203599)

15 Rose LE, Bittner-Eddy PD, Langley CH, Holub EB, Michelmore RW, Beynon JL. 2004 The Maintenance of Extreme Amino Acid Diversity at the Disease Resistance Gene, *RPP13* , in *Arabidopsis thaliana*. *Genetics* **166**, 1517–1527. (10.1534/genetics.166.3.1517)

16 Bergelson J, Kreitman M, Stahl EA, Tian D. 2001 Evolutionary dynamics of plant R-genes.. *Science (New York, N.Y.)* **292**, 2281–5. Place: United States.

17 Bakker EG, Toomajian C, Kreitman M, Bergelson J. 2006 A Genome-Wide Survey of *R* Gene Polymorphisms in *Arabidopsis*. *The Plant Cell* **18**, 1803–1818. (10.1105/tpc.106.042614)

18 Wu Q, Han TS, Chen X, Chen JF, Zou YP, Li ZW, Xu YC, Guo YL. 2017 Long-term balancing selection contributes to adaptation in Arabidopsis and its relatives. *Genome Biology* **18**, 217. (10.1186/s13059-017-1342-8)

19 Koenig D, Hagmann J, Li R, Bemm F, Slotte T, Neuffer B, Wright SI, Weigel D. 2019 Long-term balancing selection drives evolution of immunity genes in Capsella. *eLife* **8**, e43606. (10.7554/eLife.43606)

20 Azevedo L, Serrano C, Amorim A, Cooper DN. 2015 Trans-species polymorphism in humans and the great apes is generally maintained by balancing selection that modulates the host immune response. *Human Genomics* **9**, 21. (10.1186/s40246-015-0043-1)

21 Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, Venn O, Bowden R, Bontrop R, Wall JD, Sella G, Donnelly P, McVean G, Przeworski M. 2013 Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees. *Science* **339**, 1578–1582. (10.1126/science.1234070)

22 Cornetti L, Fields PD, Du Pasquier L, Ebert D. 2024 Long-term balancing selection for pathogen resistance maintains trans-species polymorphisms in a planktonic crustacean. *Nature Communications* **15**, 5333. (10.1038/s41467-024-49726-8)

23 Karasov TL, Kniskern JM, Gao L, DeYoung BJ, Ding J, Dubiella U, Lastra RO, Nallu S, Roux F, Innes RW, Barrett LG, Hudson RR, Bergelson J. 2014 The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature* **512**, 436–440. (10.1038/nature13439)

24 Gao L, Roux F, Bergelson J. 2009 Quantitative fitness effects of infection in a gene-for-gene system. *New Phytologist* **184**, 485–494. (10.1111/j.1469-8137.2009.02959.x)

25 Sasaki A. 2000 Host-parasite coevolution in a multilocus gene-for-gene system. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **267**, 2183–2188. (10.1098/rspb.2000.1267)

26 Ashby B, Boots M. 2017 Multi-mode fluctuating selection in host–parasite coevolution. *Ecology Letters* **20**, 357–365. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.12734 (https://doi.org/10.1111/ele.12734)

27 Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003 Genome-Wide Analysis of NBS-LRR–Encoding Genes in Arabidopsis[W]. *The Plant Cell* **15**, 809–834. (10.1105/tpc.009308)

28 van Wersch S, Li X. 2019 Stronger When Together: Clustering of Plant NLR Disease resistance Genes. *Trends in Plant Science* **24**, 688–699. (10.1016/j.tplants.2019.05.005)

29 Leister D. 2004 Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance gene. *Trends in genetics: TIG* **20**, 116–122. (10.1016/j.tig.2004.01.007)

30 Michelmore RW, Meyers BC. 1998 Clusters of Resistance Genes in Plants Evolve by Divergent Selection and a Birth-and-Death Process. *Genome Research* **8**, 1113–1130. (10.1101/gr.8.11.1113)

31 Thompson JN, Burdon JJ. 1992 Gene-for-gene coevolution between plants and parasites. *Nature* **360**, 121–125. (10.1038/360121a0)

32 Tellier A, Brown JKM. 2007 Polymorphism in Multilocus Host–Parasite Coevolutionary Interactions. *Genetics* **177**, 1777–1790. (10.1534/genetics.107.074393)

33 Thrall PH, Barrett LG, Dodds PN, Burdon JJ. 2016 Epidemiological and Evolutionary Outcomes in Gene-for-Gene and Matching Allele Models. *Frontiers in Plant Science* **6**. (10.3389/fpls.2015.01084)

34 Brown JKM, Tellier A. 2011 Plant-Parasite Coevolution: Bridging the Gap between Genetics and Ecology. *Annual Review of Phytopathology* **49**, 345–367. (10.1146/annurev-phyto-072910-095301)

35 Bergelson J. 1994 The Effects of Genotype and the Environment on Costs of Resistance in Lettuce. *The American Naturalist* **143**, 349–359.

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

37 MacQueen A, Sun X, Bergelson J. 2016 Genetic architecture and pleiotropy shape costs of Rps2-mediated resistance in Arabidopsis thaliana. *Nature Plants* **2**, 16110. (10.1038/nplants.2016.110)

38 Brown JKM, Rant JC. 2013 Fitness costs and trade-offs of disease resistance and their consequences for breeding arable crops. *Plant Pathology* **62**, 83–95. (10.1111/ppa.12163)

39 Bergelson J, Purrington CB. 1996 Surveying Patterns in the Cost of Resistance in Plants. *The American Naturalist* **148**, 536–558. (10.1086/285938)

40 Leonard KJ. 1994 Stability of Equilibria in a Gene-for-Gene Coevolution Model of Host-Parasite Interactions. *Phytopathology* **84**, 70. (10.1094/Phyto-84-70)

41 Chae E, Bomblies K, Kim ST, Karelina D, Zaidem M, Ossowski S, Martín-Pizarro C, Laitinen RA, Rowan BA, Tenenboim H, Lechner S, Demar M, Habring-Müller A, Lanz C, Rätsch G, Weigel D. 2014 Species-wide Genetic Incompatibility Analysis Identifies Immune Genes as Hot Spots of Deleterious Epistasis. *Cell* **159**, 1341–1351. (10.1016/j.cell.2014.10.049)

42 Wan W, Kim S, Castel B, Charoennit N, Chae E. 2021 Genetics of autoimmunity in plants: an evolutionary genetics perspective. *New Phytologist* **229**, 1215–1233. (10.1111/nph.16947)

43 Calvo-Baltanás V, Wang J, Chae E. 2020 Hybrid Incompatibility of the Plant Immune System: An Opposite Force to Heterosis Equilibrating Hybrid Performances. *Frontiers in Plant Science* **11**, 576796. (10.3389/fpls.2020.576796)

44 Hulse SV, Bruns EL. 2024 The emergence of nonlinear evolutionary trade-offs and the maintenance of genetic polymorphisms. *Biology Letters* **20**, 20240296. (10.1098/rsbl.2024.0296)

45 Huang W, Traulsen A, Werner B, Hiltunen T, Becks L. 2017 Dynamical trade-offs arise from antagonistic coevolution and decrease intraspecific diversity. *Nature Communications* **8**, 2059. (10.1038/s41467-017-01957-8)

46 Montarry J, Hamelin FM, Glais I, Corbi R, Andrivon D. 2010 Fitness costs associated with unnecessary virulence factors and life history traits: evolutionary insights from the potato late blight pathogen Phytophthora infestans. *BMC Evolutionary Biology* **10**, 283. (10.1186/1471-2148-10-283)

47 Wichmann G, Bergelson J. 2004 Effector Genes of *Xanthamonas axonopodis* pv. *vesicatoria* Promote Transmission and Enhance Other Fitness Traits in the Field. *Genetics* **166**, 693–706. (10.1534/genetics.166.2.693)

48 Kaltz O, Shykoff JA. 1998 Local adaptation in host–parasite systems. *Heredity* **81**, 361–370. (10.1046/j.1365-2540.1998.00435.x)

49 Gandon S, Michalakis Y. 2002 Local adaptation, evolutionary potential and host–parasite coevolution: interactions between migration, mutation, population size and generation time. *Journal of Evolutionary Biology* **15**, 451–462. (10.1046/j.1420-9101.2002.00402.x)

50 Gandon S, Capowiez Y, Dubois Y, Michalakis Y, Olivieri I. 1996 Local adaptation and gene-for-gene coevolution in a metapopulation model. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **263**, 1003–1009. (10.1098/rspb.1996.0148)

51 Kaltz O, Gandon S, Michalakis Y, Shykoff JA. 1999 LOCAL MALADAPTATION IN THE ANTHER-SMUT FUNGUS *MICROBOTRYUM VIOLACEUM* TO ITS HOST PLANT *SILENE LATIFOLIA* : EVIDENCE FROM A CROSS-INOCULATION EXPERIMENT. *Evolution* **53**, 395–407. (10.1111/j.1558-5646.1999.tb03775.x)

52 Baggs E, Dagdas G, Krasileva K. 2017 NLR diversity, helpers and integrated domains: making sense of the NLR IDentity. *Current Opinion in Plant Biology* **38**, 59–67. (10.1016/j.pbi.2017.04.012)

53 Wu CH, Abd-El-Haliem A, Bozkurt TO, Belhaj K, Terauchi R, Vossen JH, Kamoun S. 2017 NLR network mediates immunity to diverse plant pathogens. *Proceedings of the National Academy of Sciences* **114**, 8113–8118. (10.1073/pnas.1702041114)

54 Karasov TL, Chae E, Herman JJ, Bergelson J. 2017 Mechanisms to Mitigate the Trade-Off between Growth and Defense. *The Plant Cell* **29**, 666–680. (10.1105/tpc.16.00931)

55 Weßling R, Epple P, Altmann S, He Y, Yang L, Henz SR, McDonald N, Wiley K, Bader KC, Gläßer C, Mukhtar MS, Haigis S, Ghamsari L, Stephens AE, Ecker JR, Vidal M, Jones JD, Mayer KF, Ver Loren van Themaat E, Weigel D, Schulze-Lefert P, Dangl JL, Panstruga R, Braun P. 2014 Convergent Targeting of a Common Host Protein-Network by Pathogen Effectors from Three Kingdoms of Life. *Cell Host & Microbe* **16**, 364–375. (10.1016/j.chom.2014.08.004)

56 Ngou BPM, Jones JD, Ding P. 2021 Plant immune networks. *Trends in Plant Science* p. S1360138521002430. (10.1016/j.tplants.2021.08.012)

57 Mukhtar MS, Carvunis AR, Dreze M, Epple P, Steinbrenner J, Moore J, Tasan M, Galli M, Hao T, Nishimura MT, Pevzner SJ, Donovan SE, Ghamsari L, Santhanam B, Romero V, Poulin MM, Gebreab F, Gutierrez BJ, Tam S, Monachello D, Boxem M, Harbort CJ, McDonald N, Gai L, Chen H, He Y, European Union Effectoromics Consortium, Vandenhaute J, Roth FP, Hill DE, Ecker JR, Vidal M, Beynon J, Braun P, Dangl JL. 2011 Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. *Science* **333**, 596–601. (10.1126/science.1203659)

58 Araki H, Tian D, Goss EM, Jakob K, Halldorsdottir SS, Kreitman M, Bergelson J. 2006 Presence/absence polymorphism for alternative pathogenicity islands in Pseudomonas viridiflava, a pathogen of Arabidopsis. *Proceedings of the National Academy of Sciences* **103**, 5887–5892. (10.1073/pnas.0601431103)

59 McCann HC, Guttman DS. 2008 Evolution of the type III secretion system and its effectors in plant–microbe interactions. *New Phytologist* **177**, 33–47. (10.1111/j.1469-8137.2007.02293.x)

60 MacQueen A, Tian D, Chang W, Holub E, Kreitman M, Bergelson J. 2019 Population Genetics of the Highly Polymorphic RPP8 Gene Family. *Genes* **10**, 691. (10.3390/genes10090691)

61 Hulse SV, Antonovics J, Hood ME, Bruns EL. 2023 Host–pathogen coevolution promotes the evolution of general, broad-spectrum resistance and reduces foreign pathogen spillover risk. *Evolution Letters* **7**, 467–477. (10.1093/evlett/qrad051)

62 Märkle H, Saur IML, Stam R. 2022 Evolution of resistance (R) gene specificity. *Essays in Biochemistry* **66**, 551–560. (10.1042/EBC20210077)

63 Frank SA. 1991 Ecological and genetic models of host-pathogen coevolution. *Heredity* **67**, 73–83. (10.1038/hdy.1991.66)

64 Frank SA. 1993 Coevolutionary genetics of plants and pathogens. *Evolutionary Ecology* **7**, 45–75. (10.1007/BF01237734)

65 Ashby B, Iritani R, Best A, White A, Boots M. 2019 Understanding the role of eco-evolutionary feedbacks in host-parasite coevolution. *Journal of Theoretical Biology* **464**, 115–125. (10.1016/j.jtbi.2018.12.031)

66 Tellier A, Moreno-Gámez S, Stephan W. 2014 Speed of adaptation and genomic footprints of host-parasite coevolution under arms race and trench warfare dynamics.. *Evolution; international journal of organic evolution* **68**, 2211–24. Place: United States.

67 Märkle H, Tellier A. 2020 Inference of coevolutionary dynamics and parameters from host and parasite polymorphism data of repeated experiments.. *PLoS computational biology* **16**, e1007668.

68 Dawkins R, Krebs JR. 1979 Arms races between and within species. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **205**, 489–511. (10.1098/rspb.1979.0081)

69 Karasov TL, Almario J, Friedemann C, Ding W, Giolai M, Heavens D, Kersten S, Lundberg DS, Neumann M, Regalado J, Neher RA, Kemen E, Weigel D. 2018 Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host & Microbe* **24**, 168–179.e4. (10.1016/j.chom.2018.06.011)

70 Frachon L, Mayjonade B, Bartoli C, Hautekèete NC, Roux F. 2019 Adaptation to Plant Communities across the Genome of Arabidopsis thaliana. *Molecular Biology and Evolution* **36**, 1442–1456. (10.1093/molbev/msz078)

A. Supplementary information

Supplementary Equations

The following (cost) function is used to calculate the relative fitness of genotype $\boldsymbol{z}$ with range ($|\boldsymbol{z}|$) in the absence of an interaction:

$$C_s(|\boldsymbol{z}|) = \begin{cases} 1 - c_S^{(1)} \dfrac{\left(1 - exp(c_s^{(2)} \cdot |\boldsymbol{z}|)\right)}{\left(1 - exp(c_s^{(2)})\right)} & \text{if } c_s^{(2)} \neq 0 \\ 1 - c_S^{(1)} \cdot |\boldsymbol{z}| & \text{if } c_s^{(2)} = 0 \end{cases} \tag{A 1}$$

where the index S indicates the species ($S \in \{H_1, H_2, P\}$) of genotype $\boldsymbol{z}$ and the parameters $c_S^{(1)}$ and $c_S^{(2)}$ govern the shape of the function.

B. Supplementary Figures

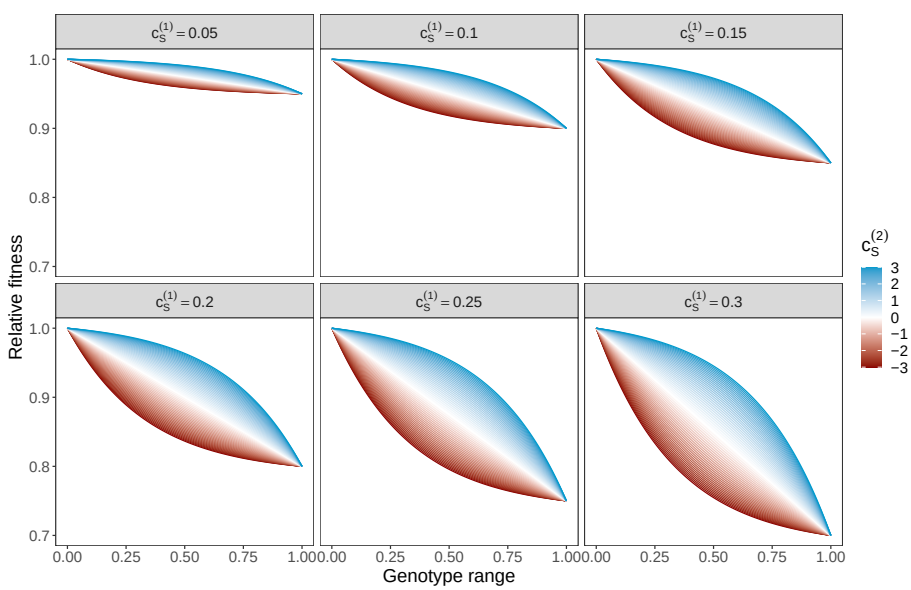

Figure S1. Fitness costs. Relationship between proportion of loci with a resistance (R-allele)/virulence (V-allele) and fitness for different combinations of $c_S^{(1)}$ and $c_S^{(2)}$. Here $c_S^{(1)}$ is the maximum reduction in fitness when a given genotype has all possible resistance or virulence alleles and $c_S^{(2)}$ governs the shape of the function.

Draft 0000000

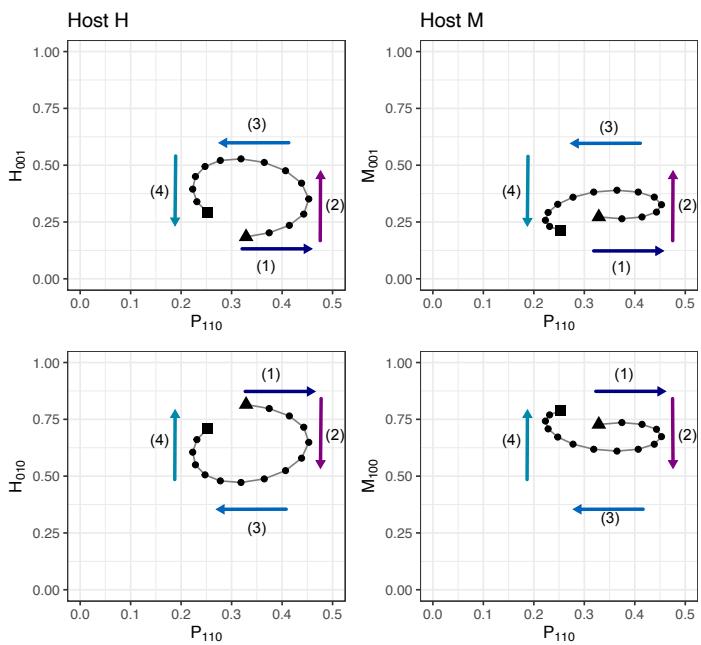

Figure S2. Phase plane plots for the changes of different host genotypes (y-axis) and the pathogen genotype P_{110} (x-axis) for the time interval 9120 to 9150. Subsequent time points are connected by segments. The first time point in the interval is marked by a triangle, the last one by a square. Results are shown for: $c_H^{(2)} = c_P^{(2)} = 3$, $c_H^{(1)} = c_P^{(1)} = 0.1$, $\phi = 0.5$, $\beta_H = \beta_P = 1$, seed $= 1,600$, $\sigma = 0.85$.

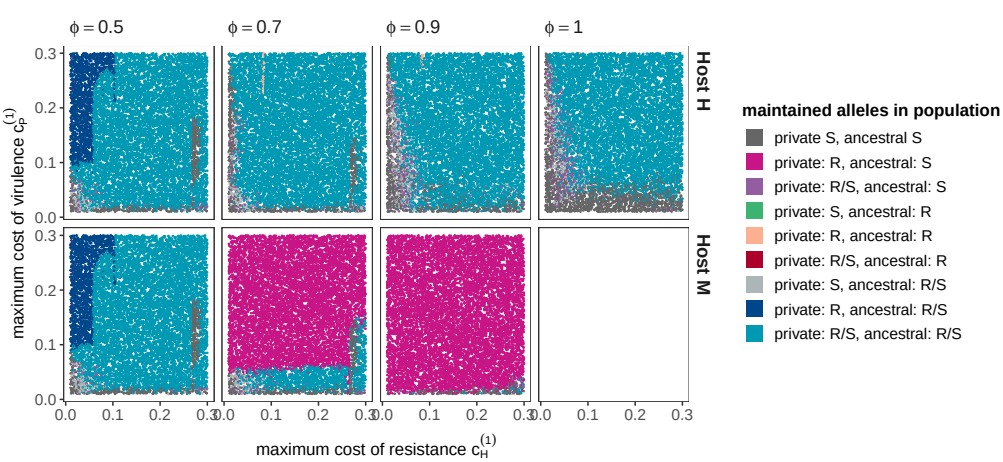

Figure S3. Maintained alleles for different combinations of the maximum costs of resistance ($c_H^{(1)}$) and virulence ($c_P^{(1)}$) based on 10,000 randomly drawn combinations of $c_H^{(1)} \in \mathcal{U}(0.01, 0.3)$ and $c_P^{(1)} \in \mathcal{U}(0.01, 0.3)$ and initial genotype frequencies for $c_H^2 = c_P^2 = 3$. Other parameters are fixed to: $\beta_H = \beta_P = 1$,$\sigma = 0.85$.

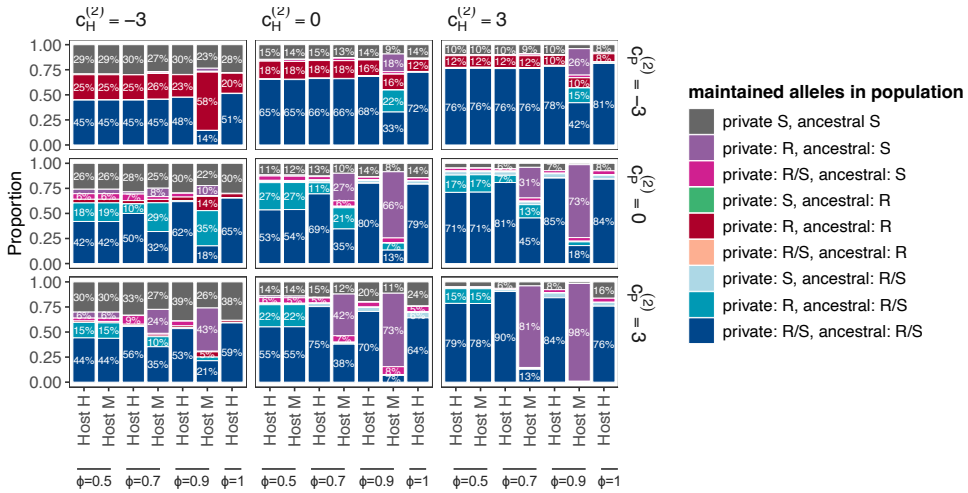

Figure S4. Maintained alleles for different combinations of the maximum costs of resistance ($c_H^{(1)}$) and virulence ($c_P^{(1)}$) based on 10,000 randomly drawn combinations of $c_H^{(1)} \in \mathcal{U}(0.01, 0.3)$ and $c_P^{(1)} \in \mathcal{U}(0.01, 0.3)$ and initial genotype frequencies for $c_H^2 = c_P^2 = 3$. Each facet corresponds to a single combination of the shape parameters $c_H^{(2)}$ (columns) and c_P^2 (rows). The relative proportion of host H is increasing from the left to the right. Other parameters are fixed to: $\beta_H = \beta_P = 1$,$\sigma = 0.85$.

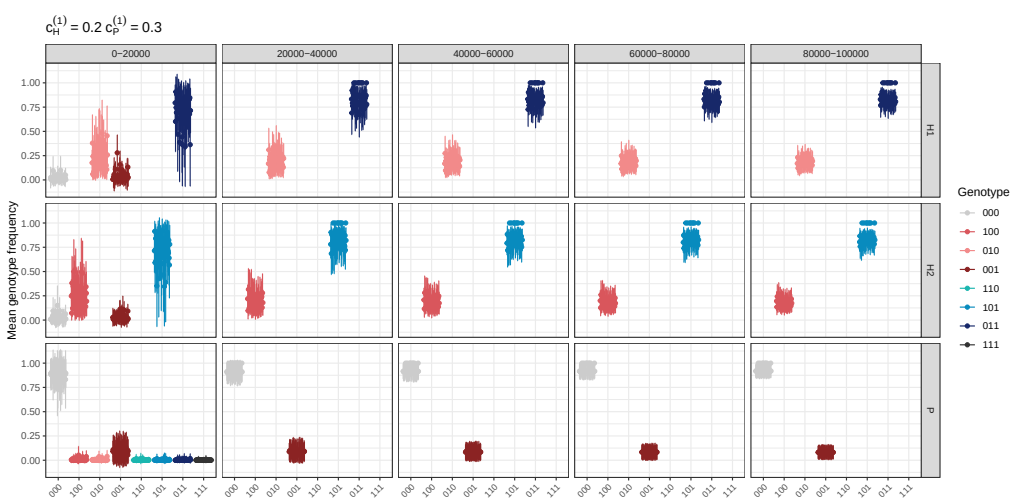

Figure S5. Figure showing the effect of initial genotype frequencies on the maintained genotypes. Shown are the results for 50 simulations for $c_H^{(1)} = 0.2$, $c_P^1 = 0.3$, $c_H^{(2)} = c_P^{(2)} = 0$ and $\phi = 0.5$ with random initialization of genotype frequencies, but otherwise all other parameters fixed to the same value. Each dot represents the average genotype frequency of a given genotype in a single simulation in the given time interval. Each facet shows the result for a single time interval (columns) for each species (rows).

C. Supplementary Tables

Table 2: Description of all parameters and variables in the extended model.

Name	description	formula
n	number of interacting parasite/host loci	
$\boldsymbol{x}$	a genotype in host species H	$\boldsymbol{x} = \{0, x_2, x_3\},\ \ x_2 \in \{0,1\}, x_3 \in \{0,1\}$
$\boldsymbol{u}$	a genotype in host species M	$\boldsymbol{u} = \{u_1, 0, u_3\},\ \ u_1 \in \{0,1\}, u_3 \in \{0,1\}$
$\boldsymbol{y}$	parasite genotype	$\boldsymbol{y} = \{y_1, y_2, y_3\},\ \ y_i \in \{0,1\}$
$f_H(\boldsymbol{x})$	frequency of host H-genotype $\boldsymbol{x}$	
$f_M(\boldsymbol{u})$	frequency of host M-genotype $\boldsymbol{u}$	
$f_P(\boldsymbol{y})$	frequency of pathogen genotype $\boldsymbol{y}$ in P	
$w_H(\boldsymbol{x})$	relative fitness of host H-genotype $\boldsymbol{x}$	
$w_M(\boldsymbol{u})$	relative fitness of host M-genotype $\boldsymbol{u}$	
$w_P(\boldsymbol{y})$	relative fitness of pathogen genotype $\boldsymbol{y}$	
$\lvert\boldsymbol{x}\rvert$	resistance range of host genotype $\boldsymbol{x}$	$\lvert\boldsymbol{x}\rvert = \sum_{i=1}^{n} x_i/(n-1)$
$\lvert\boldsymbol{u}\rvert$	resistance range of host genotype $\boldsymbol{u}$	$\lvert\boldsymbol{u}\rvert = \sum_{i=1}^{n} u_i/(n-1)$
$\lvert\boldsymbol{y}\rvert$	infectivity range of pathogen genotype $\boldsymbol{y}$	$\lvert\boldsymbol{y}\rvert = \sum_{i=1}^{n} y_i/n$
d_{xy}	number of effective resistance alleles in an interaction between host genotype $\boldsymbol{x}$ and pathogen genotype $\boldsymbol{y}$	$d_{xy} = \sum_{i=1}^{n} x_i(1-y_i)$
d_{uy}	number of effective resistance alleles in an interaction between host genotype $\boldsymbol{u}$ and pathogen genotype $\boldsymbol{y}$	$d_{uy} = \sum_{i=1}^{n} u_i(1-y_i)$
σ	parameter governing how the number of effective resistance alleles relates to the probability of infection	
$Q(\boldsymbol{x}, \boldsymbol{y})$	probability of infection in an interaction between host genotype $\boldsymbol{x}$ and pathogen genotype $\boldsymbol{y}$	$Q(\boldsymbol{x}, \boldsymbol{y}) = \sigma^{d_{xy}}$
$Q(\boldsymbol{u}, \boldsymbol{y})$	probability of infection in an interaction between host genotype $\boldsymbol{u}$ and pathogen genotype $\boldsymbol{y}$	$Q(\boldsymbol{u}, \boldsymbol{y}) = \sigma^{d_{uy}}$
β_H	effect of a successful infection on host fitness	
β_H	effect of a successful infection on parasite fitness	
$C_H(\lvert\boldsymbol{x}\rvert)$	relative fitness of genotype $\boldsymbol{x}$ with range $\lvert\boldsymbol{x}\rvert$ in the absence of infection	see Equation A 1

Continued on next page

Draft 0000000

Table 2: Description of all parameters and variables in the extended model.

Name	description	formula
$C_H(\|\boldsymbol{u}\|)$	relative fitness of genotype $\boldsymbol{u}$ with range $\|\boldsymbol{u}\|$ in the absence of infection	see Equation A 1
$C_P(\|\boldsymbol{y}\|)$	relative fitness of parasite genotype $\boldsymbol{y}$ with infectivity range $\|\boldsymbol{y}\|$	see Equation A 1
$c_H^{(1)} = c_{H_1}^{(1)} = c_{H_2}^{(1)}$	parameter for Equation A 1 for H_1 and H_2	
$c_H^{(2)} = c_{H_1}^{(2)} = c_{H_2}^{(2)}$	parameter for Equation A 1 for H_1 and H_2	
$c_P^{(1)}$	parameter for Equation A 1 for the parasite	
$c_P^{(2)}$	parameter for Equation A 1 for the parasite	