Evolutionary Ecology, 1993, **7**, 45–75

Coevolutionary genetics of plants and pathogens

STEVEN A. FRANK
Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92717, USA

Summary

The genetic polymorphism maintained by host-pathogen coevolution is analysed in a multilocus model. The model assumes gene-for-gene interactions of the type commonly observed between host plants and their fungal pathogens. Unstable (epidemic) systems maintain more resistance genes, fewer virulence genes, and less overall genetic diversity than stable (endemic) diseases. The stability of the system depends primarily on demographic parameters, such as the pathogen's intrinsic rate of increase, rather than genetic parameters, such as the costs of resistance and virulence. At equilibrium the model predicts that the number of resistance alleles in each host plant follows a binomial distribution that depends on the cost to the pathogen for carrying virulence alleles. Similarly, the number of virulence alleles in each pathogen spore follows a binomial distribution that depends on one minus the cost to the host for carrying resistance alleles. Data from wild populations match the predicted binomial distributions.

Keywords: plant disease; genetic polymorphism; gene-for-gene; epidemiology

Introduction

In this paper I analyse the forces that control genetic polymorphism for resistance and virulence in plants and their pathogens. Plant-pathogen coevolution maintains complex multilocus polymorphisms that vary over space and probably over time as well. Considerable effort has been devoted to both formal and speculative theories that explain the observed patterns of polymorphism.

The complexity of plant-pathogen coevolution has maintained a wide gulf between the assumptions of formal models and the observed genetics and population dynamics. In particular, past work has typically assumed that: (1) only one or perhaps two genetic loci are involved in the interaction; (2) population sizes of the host and pathogen are fixed in spite of changing disease frequencies; and (3) spatial subdivision and migration do not play an important role in the dynamics of populations and genotype frequencies.

I present a formal model that has realistic multilocus genetics, that allows population sizes to vary with changes in the frequency of disease, and that simulates the effects of immigration by occasionally introducing a randomly chosen genotype.

Background

Observations

I am mainly concerned with polymorphism in natural populations, but I introduce the problem by first summarizing studies of agricultural systems. These applied studies are rich in genetic detail and theoretical speculation when compared with the few analyses of natural populations that are presently available (Burdon, 1987a).

0269–7653

Agricultural scientists control disease by planting genetically resistant crop varieties. This method has been used for over 80 years, with roughly the same results in each case (Agrios, 1988).

First, by extensive breeding programs, a new crop variety is developed with major gene resistance to one or more of the locally dominant pathogen races. Typically these pathogens are different races of what is apparently a single species. Second, the new variety is planted. The genetic resistance reduces disease for a few years, and yield improves. Third, new races of virulent pathogens arise, disease increases, and yield declines. The new races are introduced locally by immigration or mutation. Finally, the cycle begins again: new crop varieties, temporary reduction of disease, new pathogen races, . . .

In the 1940s Flor discovered a remarkably simple relationship between the resistance genes in different flax varieties and the virulence genes in different races of flax rust (Flor, 1956, 1971). At each locus in flax that affects resistance, the alternative alleles confer either a susceptible (S) or resistant (R) phenotype. For each such locus in the plant, of which there are many, the pathogen has a matching locus that confers either an avirulent (A) or a virulent (V) phenotype. A flax variety is resistant to a pathogen race if, at any one of the matching gene-for-gene loci, the flax phenotype is R and the rust phenotype is A. Otherwise an inoculation results in disease. The actual genetics in flax and other gene-for-gene systems is not quite so simple, either because of multiple resistance determinants at a single locus or tight linkage among resistance loci, but on the whole a gene-for-gene system with over 25 matching host-pathogen loci provides a good approximate description for the interaction in flax.

Flor's simple gene-for-gene hypothesis has great power in reducing complexity to a manageable level (Person, 1959). Consider that, for N loci, there are 2^N distinct resistance phenotypes of crop varieties in relation to the 2^N different virulence patterns among pathogens. Under the gene-for-gene hypothesis, the outcome for each of the 2^{2N} variety-by-pathogen interactions follows a simple and predictable pattern.

More than 25 different crop-pathogen systems have been reported to follow a gene-for-gene interaction (Burdon, 1987a). The importance of gene-for-gene interactions in agricultural systems and their prevalence in wild populations are controversial subjects. Some authors have suggested that, among the many types of qualitative and quantitative genetic interactions that occur, the design of breeding and screening programs strongly biases the inferred genetics toward single genes with large effect (Day, 1974; Barrett, 1985). In addition, nearly all well-documented cases of gene-for-gene genetics come from agricultural systems, raising the possibility that other genetic patterns are more important in wild populations (Burdon, 1987a, pp.67–9 summarizes this topic).

These cautionary remarks are reminders that other types of genetic resistance could, in theory, play a more important role than gene-for-gene interactions, even if there is little evidence at present. Against these cautions, however, it may be noted that: the prevalence of agricultural examples follows from plant pathologists interests in crop improvement rather than the biology of natural populations; there is a good deal of circumstantial evidence supporting gene-for-gene interactions in wild populations; and, in the few cases where detailed genetic analysis has been applied to a natural system, the results suggest a gene-for-gene pattern (Burdon, 1987a; Clarke *et al.*, 1990). In reality few systems will match exactly the idealized gene-for-gene pattern, but many systems do have complementary major gene effects between host and pathogen. I use the phrase 'gene-for-gene' as a shorthand for 'complementary major gene interactions between host and pathogen' when discussing observations.

The large number of loci that enter into each gene-for-gene system suggests the possibility of widespread polymorphism for resistance and virulence in natural populations. Alternatively,

although these loci can affect disease, it may be that polymorphism is rare in nature because most loci are fixed for a particular allele. Indeed, unless there is some opposing force, fixation for resistance (R) in hosts and virulence (V) in pathogens seems likely.

The evidence from natural populations suggests that considerable polymorphism for both resistance and virulence often occurs (Burdon, 1987a). More interestingly, the total amounts and the patterns of polymorphism vary considerably among systems (e.g. Burdon, 1987b; Jarosz and Burdon, 1990) and, within a particular plant-pathogen system, may vary widely from one location to another over distances of a few kilometres (e.g. Clarke *et al.*, 1990; Jarosz and Burdon, 1991).

Theories

The differences in observed polymorphism among systems, and among populations of a particular host-pathogen interaction, provide an excellent opportunity to test theories about the forces that control polymorphism. To date, theoretical analyses have been quite limited because of the inherent complexity of the genetics and population dynamics. Three classes of theory can be distinguished.

First, Vanderplank (1984) suggested that avirulence-virulence polymorphisms in pathogens are maintained by a pleiotropic cost of virulence. The cost model assumes that the avirulent strain has higher fitness than the virulent strain in the absence of matching resistance, whereas only the virulent strain can reproduce on a host with matching resistance. The logic of the model is that if costs were absent, virulent strains would always have a greater or equal fitness to avirulent strains, and therefore avirulence would rarely be observed. A mix of avirulence and virulence is, however, frequently observed, thus Vanderplank infers that a cost of virulence exists. A similar cost argument applies to the maintenance of resistance polymorphisms.

Another prediction of the cost model concerns the loss of unnecessary resistance or virulence within a local population (Vanderplank, 1984). If a particular resistance allele is absent in a region then, according to this hypothesis, the matching virulence allele should be rare because it has a fitness disadvantage relative to avirulence. Likewise, if virulence is common, then matching host resistance is predicted to be rare. This hypothesis has spawned a large literature of commentary and conflict, mainly concerning the interpretation of available data rather than the coherency of the seemingly obvious logic (e.g. Nelson, 1973; Crill, 1977; Parlevliet, 1981). Leonard and Czochor (1980) have, however, questioned whether the assumption of costs necessarily leads to the expected spatial pattern in a way that is easy to detect and interpret. The theoretical analyses presented below suggest that, for wild populations, the predicted match between host and pathogen gene frequency is weak or absent even when there are significant costs.

In a second class of theory, formal mathematical models have analysed the conditions for the maintenance of polymorphism (reviewed by Leonard and Czochor, 1980; Burdon, 1987a). The main conclusions are that an increasing cost of resistance for the host reduces the frequency of the matching virulence allele in the pathogen, and an increasing cost of virulence in the pathogen increases the frequency of the matching resistance allele in the host.

These previous models assume one or perhaps two loci are involved in the gene-for-gene interaction. They also typically assume that host and pathogen population sizes are constant rather than fluctuating in response to changing disease intensity and epidemics.

Third, Frank (1991a,b) has analysed single-locus models in which population sizes of hosts and pathogens fluctuate in response to changing patterns of disease frequency. These models showed that demographic parameters, such as birth, death, and migration rates, typically have a stronger effect on patterns of polymorphism than do the genetic parameters such as costs of resistance and virulence. The genetics in these models were, however, only loosely analogous to gene-for-gene

systems and, as in previous models, the multilocus character of observed polymorphisms was not analysed. The multilocus models presented below, which are considerably more realistic, also show the importance of demographic parameters for the control of genetic polymorphism.

The Model

Genetics of virulence and resistance

In this model haploid hosts and pathogens interact at N loci to determine whether an inoculation results in a resistant reaction or in disease. The virulence-resistance interactions follow the gene-for-gene model typical of many plant pathosystems (Flor, 1956, 1971; Day, 1974; Burdon, 1987a; Callow, 1987; Gabriel and Rolfe, 1990).

At each of the N loci an individual host or pathogen has either a 0 or 1 allele, so there are 2^N genotypes. Host genotypes i are the bit-string of 0's and 1's when i is written in base two, with $i = 0,1,2, \ldots, 2^N - 1$. For example, the bit-string genotype 110 has a 0 allele at the first locus and 1 alleles at the second and third loci; the genotype when written in base ten has a value of $i = 6$. The abundance of host genotype i is h_i. Pathogen genotypes j, with abundance p_j, are coded in the same way as host genotypes.

An inoculation develops into disease if the pathogen can either escape detection by the host or withstand the host's defenses. An empirically derived rule, the gene-for-gene interaction, is that the host can block pathogen growth only if both host and pathogen have matching 1 alleles at one or more loci. The alleles are coded such that: the 1 allele in a host can confer resistance, the 0 allele is universally susceptible; the 1 allele in the pathogen can lead to avirulence, the 0 allele is universally virulent.

Pathogen j is virulent on host i only if there are no matching 1 alleles at any of the N loci. Virulence can conveniently be described by bitwise intersection as (i AND j) = 0, where the genotypes are written as bit-strings.

Dynamics of the host-pathogen system: asexual model

The model tracks the total population sizes of all host and pathogen genotypes rather than just the relative frequencies of genotypes. Epidemic fluctuations in population sizes and disease frequencies can therefore be studied along with gene frequency dynamics. The model assumes discrete Lotka-Volterra dynamics, modified from May's (1974) Equation 3.20, given by:

$$\Delta h_i = h_i \left[r_i - \sum_{k=0}^{2^N-1} r_k h_k - m \sum_{k=0}^{2^N-1} \lambda_{ik} p_k \right]$$

$$\Delta p_j = p_j \left[-s + b_j \sum_{k=0}^{2^N-1} \lambda_{kj} h_k \right] \qquad (1)$$

The difference equation Δh_i is interpreted as follows. The h_i individuals of host genotype i can potentially increase their numbers by $r_i h_i$ after reproduction. The parameter r_i has two components: r, the maximum rate of increase for the species, and A_i, the total cost of resistance alleles carried by genotype i. The total cost A_i will be defined in terms of the cost per allele, a, as described later in the section on Mathematical analysis.

The next term in the equation for Δh_i describes how offspring of genotype i compete for limited resources among themselves and the offspring of the other host genotypes. The reduction

in growth caused by intraspecific competition is given in the term $\Sigma r_k h_k$. The carrying capacity for the host population is standardized to one, and all other values are scaled accordingly.

Finally, the increase in h_i is further reduced through pathogen infection by the morbidity and mortality of disease, m, multiplied by the abundance of virulent pathogens specific for that host type, where $\lambda_{ij} = 1$ if pathogen j is virulent on host i (i AND $j = 0$), and $\lambda_{ij} = 0$ if the host is resistant (i AND $j \neq 0$).

The difference equation Δp_j is interpreted in a similar manner. A proportion of pathogen individuals s dies at the end of each time period. The number of pathogens that successfully attack a host is proportional to p_j multiplied by the abundance of susceptible hosts. Each successful pathogen produces b_j progeny that survive to the next time period and land on a host. The value of b_j has two components: b, the maximum birth rate for the species, and V_j, the total cost of virulence alleles carried by genotype j. The total cost V_j will be defined in terms of the cost per allele, v, as described later in the section on Mathematical analysis.

Introduction and loss of genetic variability: extinctions and recolonizations

Two parameters are used to control the extinction and reintroduction of genetic variability in the local population. For extinctions, if the abundance of a genotype falls below a truncation point set at 0.1 percent of the carrying capacity of the host population, then the abundance of that genotype is set to zero. This process is applied to all host and pathogen genotypes in each time-step. For reintroductions, in each time-step a uniformly distributed random number between zero and one is chosen, and if this number is less than α, then one of the 2^N possible genotypes is chosen randomly and, if its abundance is less than the truncation point for extinctions, then the abundance of that genotype is set to twice the truncation value. This process, which is applied independently to the hosts and pathogens in each time-step, simulates either new mutations or the arrival of occasional immigrants in the local population. I will refer to α as the migration rate, although the process can include both migration and mutation. Newly introduced genotypes enter into the local population at the same stage in the life cycle as newly recruited progeny born in the local population.

Sexual model and summary of life cycle

The sexual model is identical to the asexual model except that, in each time-step, haploid genotypes pair to form diploid genotypes, which immediately undergo meiosis to produce haploid progeny. Mating is random, and the recombination fraction is 0.5 between adjacent loci. In the models studied here, both host and pathogen are asexual or both are sexual. The sequence in the life cycle is: infection, birth and death processes described in Equation 1; truncation of rare genotypes; collection of statistics; mating followed by meiosis, if the model is sexual; and immigration.

Mathematical analysis

Mathematical results throughout this paper are derived for the asexual model. The computer simulations presented below show that many of the predictions from the asexual model also apply to the sexual model.

The dynamical system in Equation 1 has 2^{N+1} dimensions and is too complex to analyse. In this section I use a few simplifying assumptions to reduce the dimensionality of the system. In the following sections I derive predictions about biologically interesting quantities such as the abundance of hosts and pathogens, the frequency of resistant reactions, and the burden of disease

on the growth of the hosts. I then present a series of computer simulations that test how well these predictions actually describe the dynamical behaviour of the full system.

To simplify the system I classify a host genotype by its number of resistance alleles and a pathogen genotype by its number of virulence alleles. Specifically, I assume that the total cost of resistance for the ith host genotype depends only on the number of resistance alleles $X(i)$ that a host carries, and the total cost of virulence for the jth pathogen genotype depends only on the number of virulence alleles $Y(j)$ that a pathogen carries. Then, at equilibrium, host genotypes fall into one of the $N + 1$ categories with 0,1, . . . , N resistance alleles, and, likewise, pathogen genotypes fall into one of $N + 1$ categories defined by the number of virulence alleles. Thus, at equilibrium, the asexual system given by Equation 1 can be reduced to $2(N + 1)$ dimensions

$$\begin{aligned} \Delta h_x &= h_x \left[r_x - \sum_{k=0}^{N} r_k h_k - m \sum_{k=0}^{N-x} \binom{N-x}{k} \tilde{p}_{k+x} \right] = 0 \\ \Delta p_y &= p_y \left[-s + b_y \sum_{k=0}^{y} \binom{y}{k} \tilde{h}_k \right] = 0 \end{aligned} \tag{2}$$

where x is the number of resistance alleles in a host, and y is the number of virulence alleles in a pathogen. The total abundance of hosts with x resistance alleles is $h_x = \binom{N}{x}\tilde{h}_x$, where $\tilde{h}_x$ is the abundance of each of the $\binom{N}{x}$ different genotypes that have x resistance alleles. Similarly, $p_y = \binom{N}{y}\tilde{p}_y$ is the abundance of pathogens that have y virulence alleles.

A number of simple predictions emerge from Equation 2 if the costs of resistance and virulence are multiplicative across loci: $r_x = r(1 - a)^x$ and $b_y = b(1 - v)^y$, where a and v are the costs of resistance and virulence per locus, respectively. I discuss these predictions in the following sections.

I have not obtained any mathematical results on the stability of Equation 2. The simulations show that stability usually decreases with an increase in the potential for epidemics, determined by the potential growth rate of the pathogen population, $b - s$. Further, high immigration rates, α, tend to stabilize the system by reducing the lag time before the arrival of a favoured genotype. The dependence of stability in this model on the potential growth of the pathogen population and on immigration is consistent with previous mathematical and computer results on a similar host-pathogen model (Frank, 1991a).

Computer analysis

Overview

The results are divided into three main sections: numerical dynamics of populations, simple measures of genetic polymorphism, and complex measures of genetic polymorphism and coevolution. On first reading it may be sufficient to study the introduction for each section, skim the results to get a sense of the general patterns, and then turn to the Discussion where the key points are emphasized. Many details of the simulation results are summarized in the figures; one may wish to study some of the more complex figures only to see the actual data or to follow the chain of reasoning that led to a particular conclusion.

Summary statistics and methods of analysis

Each run of the computer model has particular values assigned to all of the parameters (Table 1). The model begins with an empty patch containing no hosts or pathogens. The patch is colonized during the following 2000 generations by the immigration process described above.

Table 1. Parameters varied in the computer analysis

Parameter	Description
r	Intrinsic rate of increase of host with no resistance alleles
a	Cost of resistance: host with x resistance alleles has an intrinsic rate of increase of $r(1 - a)^x$
m	Proportional to morbidity and mortality per successful infection
b	Birth rate of pathogen with no virulence alleles
v	Cost of virulence: pathogen with y resistance alleles has a birth rate of $b(1 - v)^y$
s	Death rate of pathogen
α	Rate at which new alleles are introduced by migration or mutation

Each generation is one turn through the life cycle described in the section Sexual model and summary of life cycle, which includes one application of Equation 1 to the abundances of hosts and pathogens. Because Equation 1 can be interpreted in a variety of ways, the average life span of an individual can be longer than one turn through the life cycle but, in this paper, I ignore issues that concern overlapping generations and age structure.

No statistics are collected during the initial 2000 generations of each run. In each of the following 5000 generations, statistics are collected on the abundance of hosts and pathogens, the frequency of pathogen attacks to which the hosts are resistant, the diversity of genotypes, and a variety of other measures described later. The run is completed at the end of these 5000 generations and, for each statistic, there are 5000 values that form a temporal distribution for that run. Each distribution is summarized by its 5th, 25th, 50th, 75th, and 95th percentiles – I often refer to the 50th percentile as the median and the difference between the 95th and 5th percentiles as the range.

Each design is a set of runs with parameters defined by a factorial combination of values. For example, if the parameters s, v, b, and r were each varied over three levels, then the design would have 3^4 runs. The output for this design would consist of 3^4 values for each of the five percentile levels for each of the statistics collected. Consider as an example the analysis of host abundance. The parameter r, as it turns out, has no effect on host abundance (see below). This is revealed by the fact that r explains essentially none of the variance in any of the five percentile levels across the 3^4 runs of the design. Thus, I summarize these data by presenting the median and other percentiles as functions of s, v, and b, when averaged over the three values of r. All output is analysed and summarized in this manner. The parameters for each factorial design are presented in the figure legends.

The simulations were written in the C programming language. The source code is available on request.

Numerical dynamics of populations

Epidemiology and the forces that drive numerical changes in populations have traditionally been analysed separately from the forces that influence genetic variability (Frank, 1992). However, recent empirical (Burdon and Jarosz, 1992) and theoretical (Frank, 1991a) studies show that the temporal dynamics of population sizes may play a key role in shaping patterns of genetic variability in plant-pathogen systems. This conclusion is supported by the simulations described

here. The results for numerical dynamics are therefore presented first, and the later sections on genetic variability are interpreted in light of the results on numerical dynamics.

The main conclusions about population dynamics are: (1) The stability of both host and pathogen populations is determined mainly by the pathogen's maximum intrinsic rate of increase, $b - s$ (see Table 1 for definition of parameters). Slowly growing pathogens lead to endemic disease and stable population sizes. Rapidly growing pathogens cause epidemics and fluctuating population sizes. (2) The migration rate, α, influences the pattern of population fluctuations when population sizes are unstable. Low rates yield higher extinction probabilities of local populations and longer times before recolonization. In addition, low migration rates cause greater population fluctuations because of longer waiting times for the introduction of an advantageous genotype.

Abundance of hosts

In this section I analyse the factors that influence the total size of the host population. I discuss the results for the different resistance classes, h_x, in a later section on Number of resistance alleles per host individual.

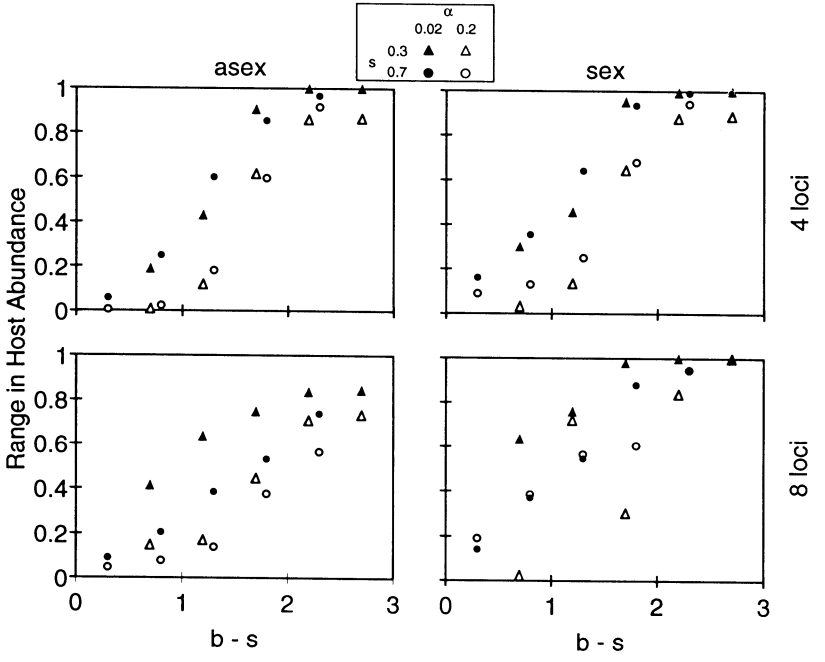

Figure 1. Factors controlling fluctuations in host abundance. The range is the difference between the 95th and 5th percentiles over 5000 generations. Filled symbols are for $\alpha = 0.02$, open symbols for $\alpha = 0.2$. Triangles represent $s = 0.3$, circles $s = 0.7$. Top panels are models with four loci; bottom panels are models with eight loci. Left panels are asexual models; right panels are sexual models. The parameters are: $b = 1,1.5,2,2.5,3$; $s = 0.3,0.7$; $v = 0.025,0.05,0.1$; $a = 0.025,0.05,0.1$; $r = 0.75,1.25$; $m = 0.02,0.2$. For the lower-right panel, parameters for b,s as in the other panels, and $a = v = 0.05$, $r = 1$, and $m = 0.02\sqrt{10}$. Plotted values are averages over v,a,r,m. These parameters had relatively little effect on the range, except for v, which explained a few percent of the observed variance, probably because of its influence on the pathogen's intrinsic rate of increase.

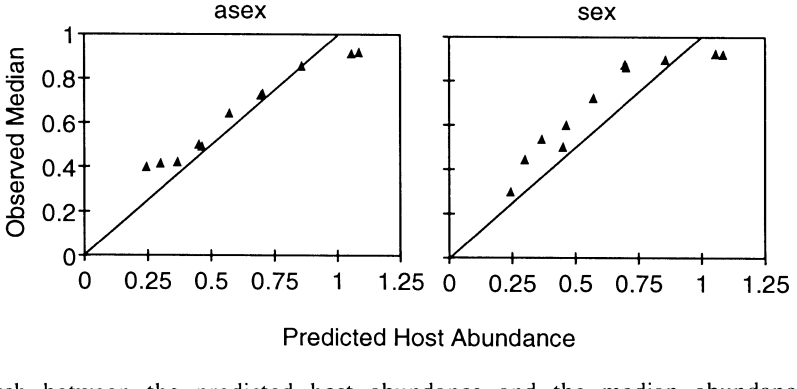

Figure 2. Match between the predicted host abundance and the median abundance observed in simulations. The left and right panels are asexual and sexual models, respectively, each with eight loci. The parameters are: $b = 1.0,1.5$; $s = 0.3,0.7$; $v = 0.025,0.05,0.1$; $a = 0.025,0.05,0.1$; $r = 0.75,1.25$; $m = 0.02,0.2$; $\alpha = 0.1$. The points shown are averaged over the values for r, m, and a. The parameters r and m had negligible effect, and a complex interaction between a, b, s, and v, explained less than 10% of the observed variance. For parameter combinations where the prediction is greater than one, the observed values clustered near 0.9. For four loci and the same parameters, the match between predicted and observed values was nearly prefect for predictions between zero and one.

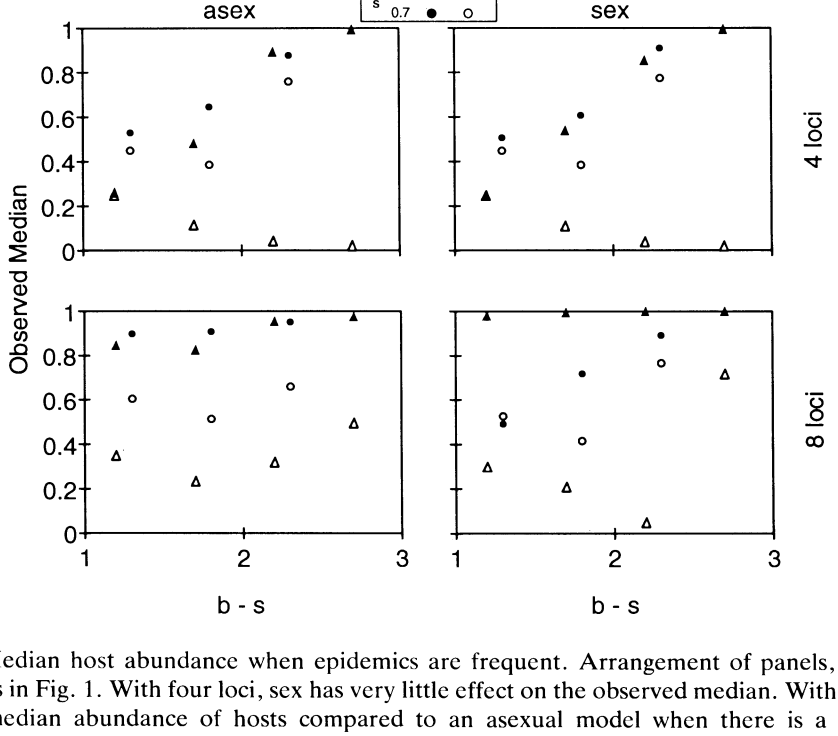

Figure 3. Median host abundance when epidemics are frequent. Arrangement of panels, symbols, and parameters as in Fig. 1. With four loci, sex has very little effect on the observed median. With eight loci, sex lowers the median abundance of hosts compared to an asexual model when there is a relatively low immigration rate ($\alpha = 0.02$) and high pathogen death rate ($s = 0.7$).

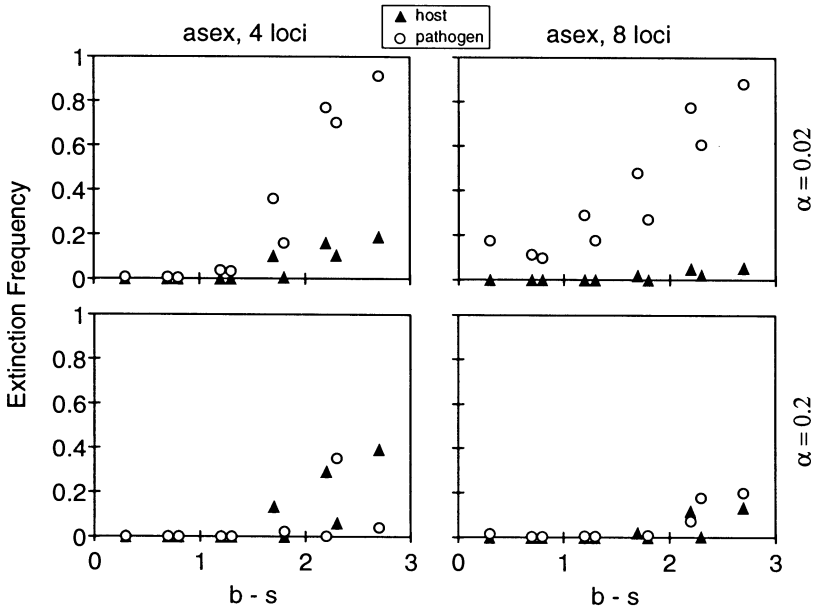

Figure 4. The frequency of generations in which hosts or pathogens are locally extinct. Triangles are the frequency at which all hosts are absent from the local patch. Circles are the frequency at which no pathogens successfully infect hosts, given that some hosts are present. The qualitative trends for sexual models with four and eight loci were very similar to the trends shown for asexual models when matched for number of loci. Parameters are the same as in the upper-left panel of Fig. 1.

The predicted equilibrium abundances of host genotypes with x resistance alleles, h_x^*, and all hosts, H^*, are obtained by solving the system in Equation 2, yielding:

$$h_x^* = \binom{N}{x} \frac{sv^x}{b(1-v)^x} \tag{3}$$

$$H^* = \sum_{x=0}^{N} h_x^* = \frac{s}{b(1-v)^N} \tag{4}$$

The simulations show that, in both sexual and asexual models, stability decreases as the potential rate of growth of the pathogen population, $b - s$, rises. Stability increases as the immigration rate, α, rises (Fig.1). Many loci and sex each decrease stability relative to systems with, respectively, few loci and no sex, but the effects are small compared with the impact of $b - s$.

The host abundance predicted by Equation 4 provides a good match with the median abundance observed in the simulations when there are eight loci and $b - s < 1.3$ (Fig. 2), even though considerable fluctuations in abundance may occur for these parameters (Fig. 1). With four loci the match is nearly perfect (see caption for Fig. 2).

The system becomes characterized by bouts of intense disease and subsequent extinction of pathogens as $b - s$ increases. Under these circumstances the median host abundance is controlled by the explosiveness of the pathogens, $b - s$, the rate at which lost genotypes are reintroduced, α, and the number of loci involved in the interaction, N (Fig. 3). Cases in which the host abundance is near its carrying capacity correspond to cases in which the pathogen population is often locally extinct (Fig. 4).

Abundance of pathogens

This section summarizes the factors that affect the size of the pathogen population, P^*. The equilibrium abundances of pathogen genotypes with y virulence alleles, p_y^*, and all pathogens, P^*, are

$$p_y^* = \binom{N}{y} (r/m)(1-a)^y a^{N-y} \qquad y = 0,1,\ldots,N-1 \tag{5}$$

$$p_N^* = (r/m)(1-a)^N(1-H^*) \tag{6}$$

$$P^* = \sum_{y=0}^{N} p_y^* = (r/m)\left(1 - H^*(1-av)^N\right) \tag{7}$$

The complex interactions among parameters can be reduced to four general attributes: scale, stability, median abundance under stability, and median abundance under repeated epidemic-extinction cycles.

Scale. A surprising fact of Lotka–Volterra dynamics is that the total morbidity and mortality imposed on the host population is independent of the morbidity and mortality per infection, m/r. This occurs because the size of the pathogen population is proportional to r/m, or, put another way, the pathogen population grows until it causes a particular amount of morbidity and mortality, independently of the severity of each infection. For the parameter combinations given in the caption to Fig. 1, the scaling of the pathogen population size to r/m occurs independently of whether the pathogen population is stable or fluctuating.

Stability. Fluctuations in pathogen abundance follow the same pattern as fluctuations in host abundance (Fig. 1) when the migration rate is relatively high, $\alpha = 0.2$. When the introduction of genetic novelty is relatively rare, $\alpha = 0.02$, then the range of pathogen abundance increases steadily as $b - s$ increases, until $b - s$ is between 1.5 and 2.0, at which point the range declines rapidly. This occurs because, with a capacity for explosive growth, the pathogen population often blows up while causing a severe epidemic, then crashes to extinction and must wait for new migrants of a virulent genotype to arrive and recolonize the host population. These results are based on the same parameter combinations given in Fig. 1.

Abundance. The prediction for the total size of the pathogen population, Equation 7, gives a close match to the observed median when $b - s < 1.5$ and there are four loci, $N = 4$. This is true for both sexual and asexual models for the parameters given in the caption of Fig. 1. For $N = 8$ in an asexual model, the observed median depends on the migration rate, α. Relatively high migration, $\alpha = 0.2$, yields a scatter of observed medians about a line that is approximately 0.7 multiplied by the equilibrium value, with no observed medians above the predicted value. For relatively low migration, the observed medians are scattered between zero and the predicted value.

With larger fluctuations, $b - s > 1.5$, the observed medians are consistently lower than predicted. When migration is relatively rare, the pathogen population is often locally extinct (Fig. 4, top row).

The burden of disease

At equilibrium, the fitness burden of disease – the reduction in host growth rate caused by infection – is

$$B = \sum_{i=0}^{2^N-1} \sum_{j=0}^{2^N-1} (h_i^*/H^*)(m/r_i)\lambda_{ij}p_j^*$$

$$= 1 - H^*(1 - av)^N \left(\frac{1 - a(1 - v)}{1 - a} \right)^N \quad (8)$$

When neither av nor N are too large, then $B \approx 1 - H^*$. When host abundance is stable or fluctuates over a small range, then Equation 8 provides a good description for how sick the host is. In general, the increasing fluctuations in the burden of disease, which are a function of $b - s$ and α, follow the pattern for host abundance in Fig. 1, except that burden fluctuates more widely because it depends on the product of host and pathogen abundances.

Simple measures of genetic polymorphism

The complex patterns of genetic variability in plant-pathogen interactions can be summarized in many different ways. The following sections analyse three simple measures that have been used in studies of natural populations. These measures have been chosen because the data are relatively easy to collect and because they reflect the complex patterns of genetic variability that result from the coevolutionary process. Some preliminary data from natural populations are presented in the Discussion.

In the first section I analyse the frequency of host resistance for an inoculation between randomly sampled host and pathogen genotypes from a single population. The frequency of resistance tends to be low when the system is stable. The following sections explain why a low resistance rate typically occurs in a stable coevolutionary interaction. In particular, the second section shows that the number of resistance alleles carried by each host is generally low at equilibrium, whereas the third section shows that the number of virulence alleles carried by each pathogen is generally high at equilibrium.

Frequency of resistance

The frequency of inoculations that lead to a resistant reaction at equilibrium is

$$R = 1 - \sum_{i=0}^{2^N-1} \sum_{j=0}^{2^N-1} \lambda_{ij}(h_i^*/H^*)(p_j^*/P^*)$$

$$= 1 - \sum_{y=0}^{N} \sum_{x=0}^{y} \binom{N}{y} \binom{y}{x} h_x^* p_y^*/H^*P^*$$

$$= \frac{1 - (1 - av)^N}{1 - H^*(1 - av)^N} \quad (9)$$

The stability of resistance frequency over time is controlled by the same factors that influence the stability of population sizes: the pathogen's maximum rate of increase, $b - s$ and the migration rate, α. High pathogen growth and low migration each increase the fluctuations in resistance (Fig. 5).

The frequency of resistance is stable under the parameters that promote the strongest ecological stability (right column of Fig. 6). In this case the frequency of resistance tends to be

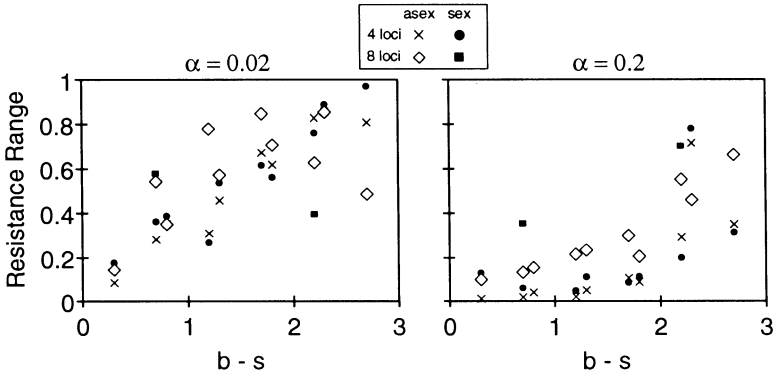

Figure 5. Fluctuations in the frequency of resistance. The parameter combinations are as in the upper-left panel of Fig. 1, with each point averaged over parameters other than α,b,s. The × is for an asexual model with four loci; the circle for a sexual model with four loci; the diamond is for an asexual model with eight loci; and the square for a sexual model with eight loci (only two squares in each panel).

low for the parameters used in the simulations. Under relatively low immigration, $\alpha = 0.02$, the observed median for resistance frequency is higher than predicted, and sex and many loci cause an increase in the frequency of resistance relative to no sex and few loci (left column of Fig. 6).

A low immigration rate causes a high median resistance frequency when the system fluctuates widely (Fig. 7). An increase in the number of loci causes a greater increase in resistance than does a switch from asexual to sexual reproduction (Fig. 7).

A note on method: to calculate the median and other percentiles of resistance for each run of 5000 generations, a generation is not counted if either the host population or the pathogen population is extinct.

Number of resistance alleles per host individual

The abundance of hosts with x resistance alleles is, at equilibrium, given by Equation 3. Dividing the abundance of each class by the total abundance, $\hat{h}_x^* = h_x^*/H^*$, yields the frequency distribution for the number of alleles per host,

$$\hat{h}_x^* = \binom{N}{x} v^x (1 - v)^{N-x} \tag{10}$$

This is a binomial distribution that depends only on the number of loci involved in the host-pathogen interaction, N, and the cost of virulence alleles in the pathogen, v. Frank (1992) provided an intuitive explanation for the surprising result that the number of resistance alleles per host is independent of the cost of resistance.

Statistics on the frequency of hosts in each class, x, were gathered in each of the 5000 generations of a run. The 5, 25, 50, 75, 95 percentiles of these frequencies were calculated for each class x over the 5000 generations, yielding a distribution of observed values for each class.

Four factors have a strong influence on the number of resistance alleles per host: N and v, as shown by the equilibrium formula, and $b - s$ and the immigration rate, α, for their effect on the overall stability of the system. These effects are summarized in Figs 8 and 9.

Three general conclusions can be drawn from Figs 8 and 9. First, when the system is nearly stable and is asexual (row three, left column in each figure), the binomial prediction closely

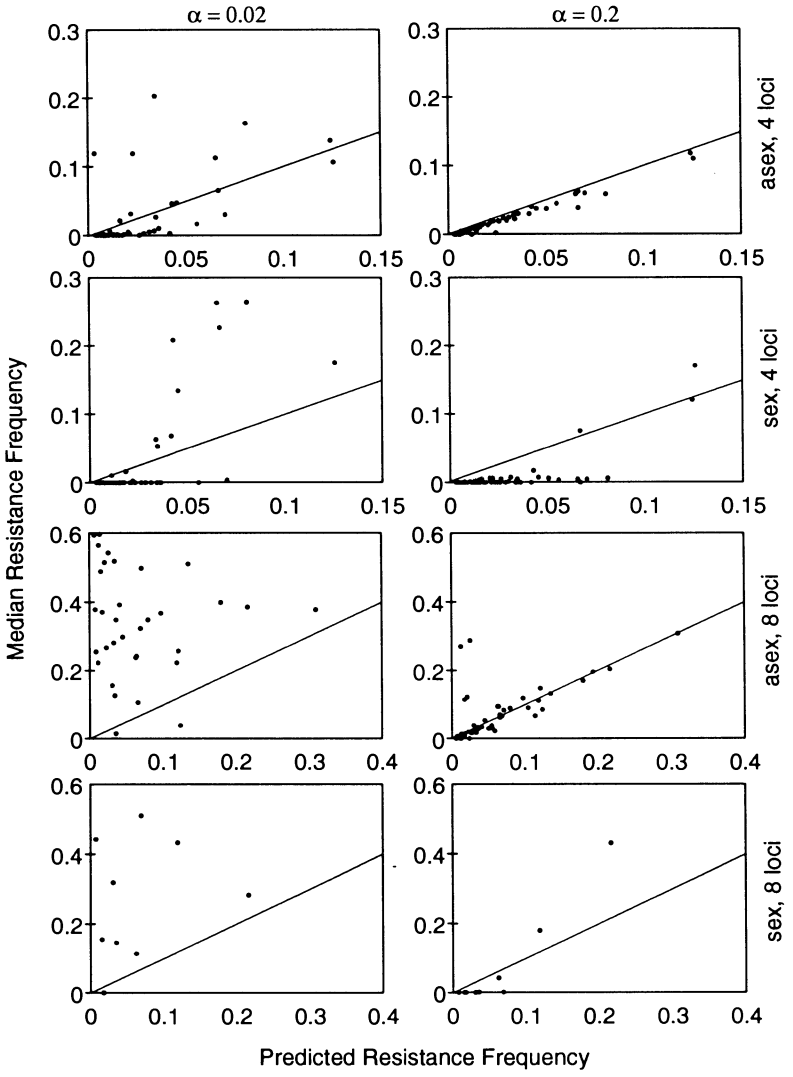

Figure 6. Predicted frequency of resistance and observed median values for low values of pathogen growth rate. Parameters are as in the upper-left panel of Fig. 1, except that $b - s \leq 1.3$ in the left column and $b - s \leq 1.8$ in the right column. For eight loci with sex, $s = 0.3$ and $b = 1.0$ are shown, averaged over the same parameter combinations as the other rows. Note the variation in scale in the top *versus* bottom set of panels.

matches the observed median – the difference is usually less than five percent. Second, sex has a small effect on the distributions when there are four loci (Fig. 8). When there are eight loci, sex causes a moderate increase in the amount of polymorphism (Fig. 9). Third, fluctuations in the distribution of resistance loci can occur even when the total abundance of hosts remains nearly constant. This can be seen by comparing the moderate fluctuations in the abundance of resistance alleles in the third row and second column of Fig. 9 with the narrow range in host abundance for the same parameter values marked by the open triangle at $b - s = 0.7$ in the lower-right panel of Fig. 1.

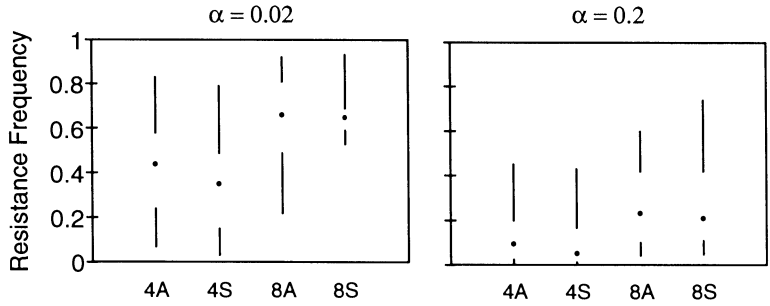

Figure 7. Distribution of the resistance frequency for high values of pathogen growth rate. The percentiles of the resistance frequency are calculated for each run; the plotted values are averages of these percentiles over a set of runs. The circle shows the median, the line above the circle stretches from the 75th to the 95th percentile, and the line below stretches from the 5th to the 25th percentile. Values are averages over the parameters in the upper-left panel of Fig. 1, except that $b - s > 1.3$ in the left panel and $b - s > 1.8$ in right panel. Categories 4A, 4S, 8A, and 8S for four loci with no sex and with sex, and eight loci with no sex and with sex, respectively. For 8S, $s = 0.3$, $b - 2.5$, averaged over the same parameters as the other cases.

Number of virulence alleles per pathogen individual

The abundance of pathogens with y resistance alleles is, at equilibrium, given by Equation 5 and Equation 6. The frequency distribution of each class is obtained by dividing the abundance of each class by the total abundance, $\hat{p}_y^* = p_y^*/P^*$, yielding an approximately binomial distribution with parameters N, the number of loci, and $(1 - a)$, one minus the cost of resistance:

$$\gamma_y = \binom{N}{y} (1 - a)^y a^{N-y} \qquad y = 0,1,\ldots,N - 1 \tag{11}$$

$$\gamma_y = (1 - a)^y (1 - H^*) \qquad y = N \tag{12}$$

$$\hat{p}_y^* = \gamma_y / \sum_{y=0}^{N} \gamma_y \tag{13}$$

where H^* was given in Equation 4. Frank (1992) provided an intuitive explanation for the fact that the number of virulence alleles per pathogen is nearly independent of the cost of virulence.

The equilibrium theory provides some guidance when the immigration rate is relatively high (lower two rows in Figs 10 and 11), although the predicted qualitative trends disappear in a model with sex and a relatively high cost of virulence (lower two panels in the right column of Fig. 11).

The intrinsic rate of increase of hosts, r, and the morbidity and mortality per infection, m, do not influence the results, which is consistent with the predictions in the equations for $\hat{p}_y^*$. The effects of three other factors can be seen by examining Figs 10 and 11.

First, relatively high costs of resistance for the host, a, tend to reduce the number of virulence alleles carried by the pathogen. This trend is predicted by the equilibrium theory.

Second, the cost of virulence, v, interacts in a complex way with other factors. When immigration is relatively rare, v has only a small effect (top two rows in Figs 10 and 11). When immigration is relatively common, increasing v tends to reduce the frequency of individuals carrying the maximum number of virulence alleles (e.g. third row in each figure). This latter effect can be explained by examining Equation 6: an increase in v increases the overall abundance

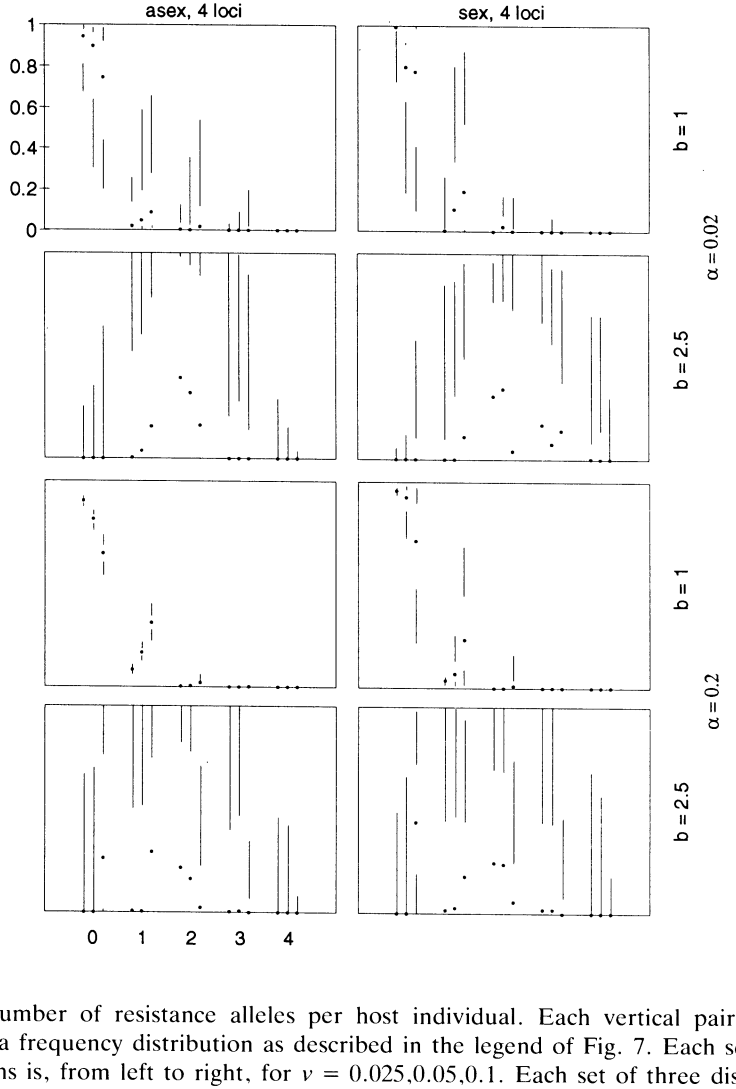

Figure 8. The number of resistance alleles per host individual. Each vertical pair of lines and the centrepoint show a frequency distribution as described in the legend of Fig. 7. Each set of three closely spaced distributions is, from left to right, for $v = 0.025, 0.05, 0.1$. Each set of three distributions is for a particular number of resistance alleles per host, x, ranging from zero to four as shown in the lower left panel. The top and bottom pairs of rows have immigration rates $\alpha = 0.02$ and $\alpha = 0.02$, respectively. Parameters are the same as in the upper-left panel of Fig. 1, except that $s = 0.3$ in all cases. Percentile values are averaged over runs with different parameter combinations for a, m, and r.

of hosts, H^*, as shown in Equation 4, and thus reduces the abundance of pathogens with N virulence alleles. There is an exception, however, shown in the bottom row of Fig. 10. Here an increase in v causes a small increase in the median number of virulence alleles carried by each pathogen and a reduction in the variation among individuals in the population. These changes probably occur because a higher cost of virulence reduces the pathogen's capacity to cause epidemics and thus stabilizes the host-pathogen interaction.

Third, the sexual system interacts with the rate of immigration, α, and the cost of virulence, v. When immigration is relatively rare, and the distribution is spread over a number of classes, sex

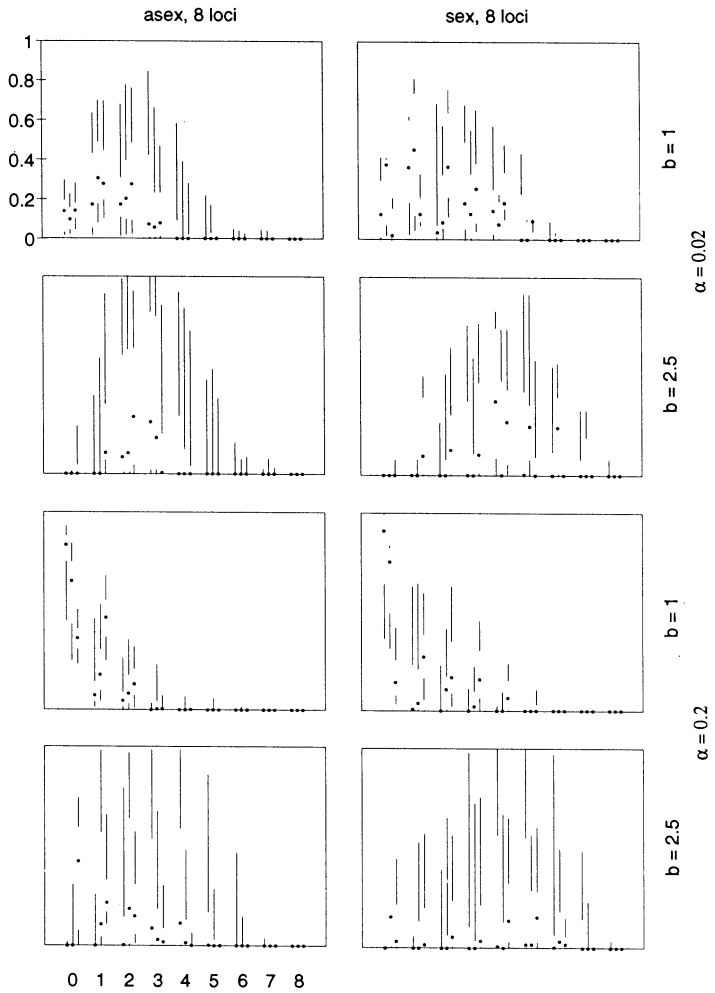

Figure 9. The number of resistance alleles per host individual. Same as Fig. 8 except that each model has eight loci. Parameters are the same as those in Fig. 8.

tends to narrow the distribution. When immigration is relatively common, pathogens in an asexual model tend to carry many virulence alleles even when virulence is relatively costly (high v). In a sexual model, an increase in v reduces the number of virulence alleles carried by pathogens.

Complex measures of genetic polymorphism and coevolution

The measures in the previous sections reduce the complex multilocus genetics to simple summaries of phenotype or genotype. The following sections analyse more detailed descriptions of multilocus genetics and the processes that drive the coevolutionary dynamics. The first section addresses patterns of genotypic diversity as measured by the Shannon index, which gives a picture of both the number of different genotypes present in a population and the evenness of the

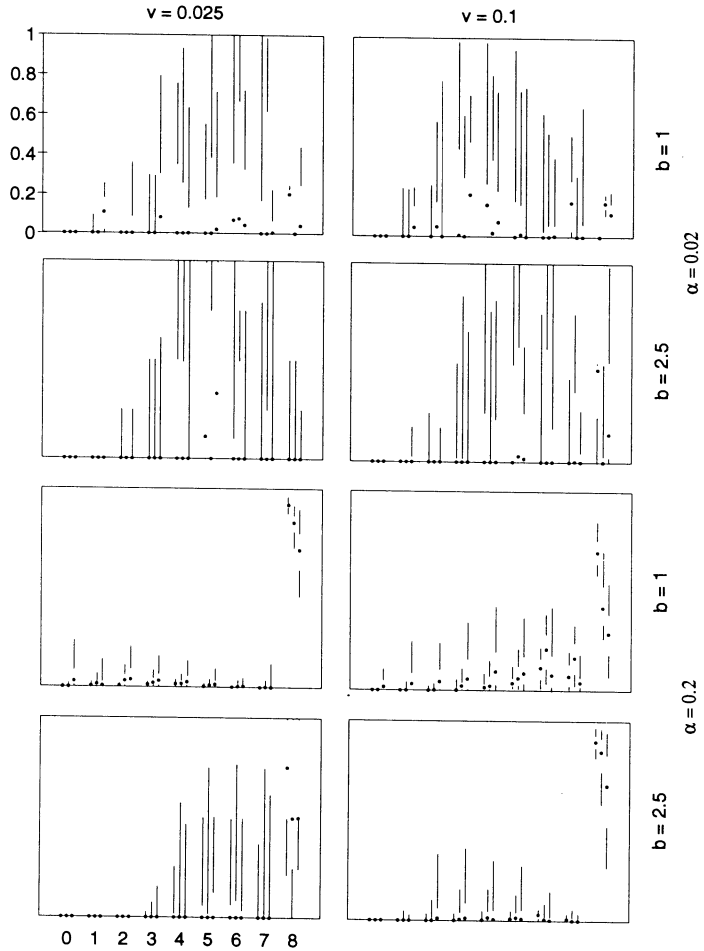

Figure 10. The number of virulence alleles per pathogen individual in an asexual model with eight loci. Rows have the same parameter definitions as in Fig. 8. Each set of three closely spaced distributions is, from left to right, for $a = 0.025,0.05,0.1$. The cost of virulence is $v = 0.025$ in the left column and $v = 0.1$ in the right column. Other parameters are the same as in the upper-left panel of Fig. 1, except that $s = 0.3$ in all cases.

distribution (see Magurran, 1988, for a comparison of diversity measures). The second section provides a detailed analysis of the match between the frequency of resistance and virulence alleles at each locus. The pattern of matching frequencies is often used to test the assumption that costs of virulence or resistance are of central importance in maintaining the observed polymorphisms.

Genetic diversity

I chose to analyse the Shannon measure of diversity in detail because several authors have used it to summarize the genotypic diversity of natural populations and to interpret the forces that maintain diversity (Groth and Roelfs, 1982; Groth and Roelfs, 1987; Jarosz and Burdon, 1991; Burdon and Jarosz, 1992). The Shannon index is $-\sum_i \beta_i \log_2 \beta_i$, where β_i is the frequency of the ith genotype. The general conclusions from this analysis are summarized in the Discussion.

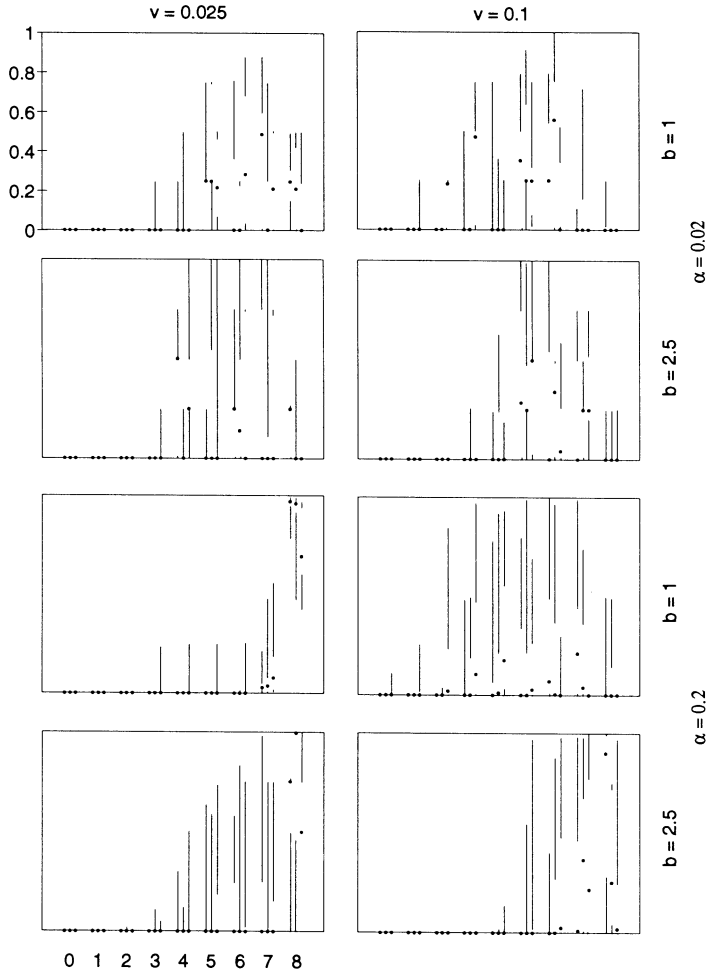

Figure 11. The number of virulence alleles per pathogen individual in a sexual model with eight loci. Parameters are the same as in Fig. 10.

Hosts. The equilibrium theory outlined above suggests how various parameters affect genotypic diversity in an asexual model: diversity is expected to increase with an increasing cost of virulence, v, and with an increasing number of loci, N (see Equation 3). In addition, higher immigration rates, α, may be expected to increase diversity, and a greater tendency for epidemics (high $b - s$) may be expected to reduce diversity by increasing extinctions of genotypes and by increasing the intensity of selection. I designed a set of simulations to analyse the effects on host diversity of these parameters (b,s,v,α,N), the cost of resistance, a, and the sexual system.

The results from these simulations confirm the trends predicted by the equilibrium theory for v and N as well as the two conjectures about the roles of $b - s$ and α (Fig. 12). The top row shows a relatively stable system with $b = 1.5$ and $s = 0.3$. Under these conditions, diversity increases rapidly with an increasing cost of virulence, v, as shown by the increase in diversity within each set of four closely spaced distributions (see figure legend). Each paired set of four distributions shows the same parameter combinations with low immigration on the left, $\alpha = 0.02$, and high

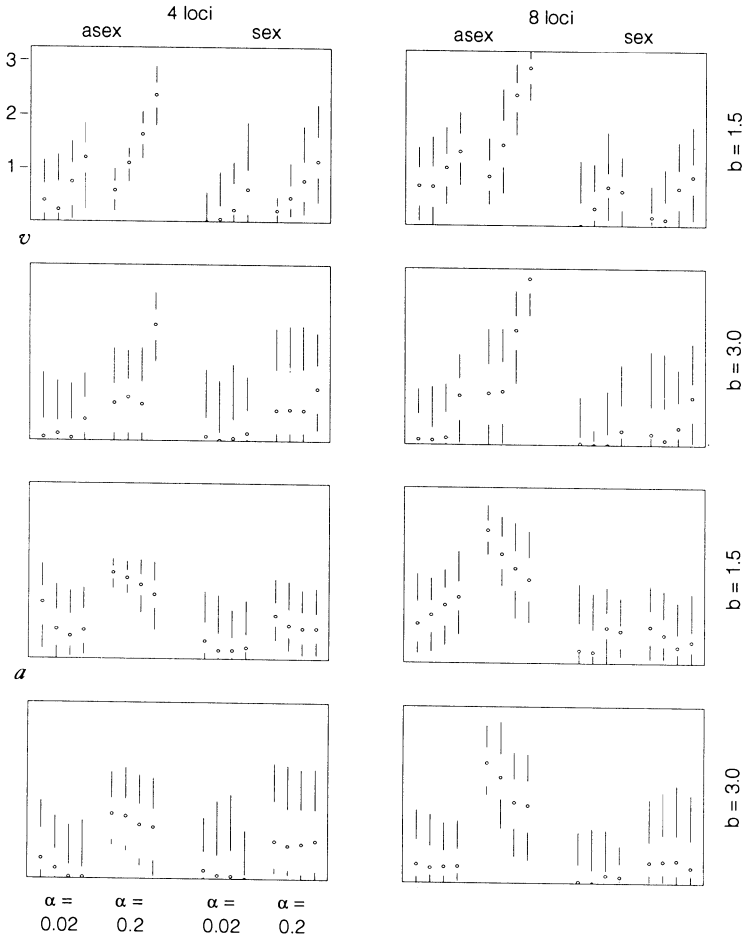

Figure 12. Genotypic diversity of hosts as measured by the Shannon index. The scale for the Shannon index measurements is given in the upper left panel. Each vertical pair of lines plus the centrepoint show a distribution as described in Figs 7–11. Each set of four closely spaced distributions shows increasing costs of virulence, v (top two rows), or resistance, a (bottom two rows). Parameter values for the design are b = 1.0,1.5,2.0,2.5,3.0; s = 0.3,0.7; a,v = 0.025,0.5,0.1,0.2; and α = 0.02,0.2. In the figure, only b = 1.5,3.0 and s = 0.3 are shown. Further details are in the text.

migration on the right, $\alpha = 0.2$. Increased immigration caused an increase in diversity. The left half of each panel shows a set of parameters in an asexual model, the right half shows the same parameters in a sexual model. In all cases, sex causes a decrease in diversity, which is particularly pronounced with eight loci and a high immigration rate. Finally, more loci cause a small or moderate increase in diversity in the asexual model but cause a small loss of diversity in the sexual model.

The second row of Fig. 12 shows the same parameters as the top row but with $b = 3$, which shifts the system from moderately fluctuating abundances and gene frequencies to a system characterized by repeated bouts of epidemics followed by extinction of the pathogens. Under epidemic conditions diversity tends to be lower, and the cost of virulence, v, and the sexual system no longer have much effect on diversity. Other parameter combinations indicate that

diversity decreases fairly smoothly as $b - s$ increases (see legend in Fig. 12 for parameter values in this design).

The bottom two rows repeat the pattern of the upper two rows, but each set of four closely spaced distributions shows four different values for the cost of resistance, a, with all values averaged over the different levels for the cost of virulence, v. In sexual models, a has little effect, whereas in asexual systems a interacts with the number of loci and the immigration rate.

Pathogens. The equilibrium theory predicts that pathogen diversity rises with increases in the cost of resistance, a, the cost of virulence, v, and the number of loci, N (Equations 5 and 6). In addition, higher immigration rates, α, may be expected to increase diversity, and a greater tendency for epidemics (high $b - s$) may be expected to reduce diversity by increasing extinctions of genotypes and by increasing the intensity of selection.

Simulation results are presented in Fig. 13. The arrangement of the figure is the same as for host diversity in Fig. 12, except that the top two rows show results for the cost of resistance, a, and the bottom two rows show results for the cost of virulence, v. Notable results include:

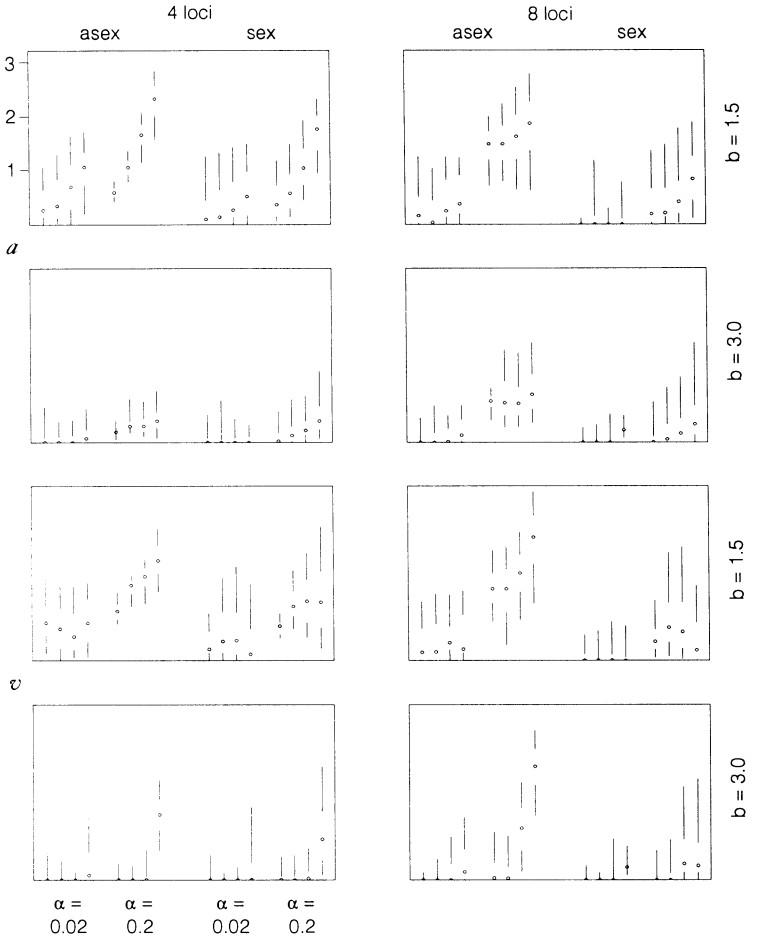

Figure 13. Genotypic diversity of pathogens. Structure and parameters are the same as Fig. 12, except that in the top two rows a varies, and in the bottom two rows v varies. Details are in the text.

increasing a causes a strong increase in diversity when epidemics are rare; frequent epidemics (high $b - s$) reduce diversity; sexuality causes a small or moderate reduction in diversity relative to asexuality; the number of loci has only a small effect that depends on the values of other parameters; and an increase in the cost of virulence has little effect when immigration is relatively rare but causes an increase in diversity when immigration is common.

Genic diversity and linkage disequilibrium. Sexual systems supported less genotypic diversity than asexual systems for both hosts and pathogens. This occurred in spite of the fact that, for the same level of genic polymorphism, sexual systems have higher genotypic diversity than asexual systems because linkage disequilibrium is much lower in sexual systems. Results from the design in the legend of Fig. 12 showed that substantial linkage disequilibrium was maintained in the asexual systems but very little linkage disequilibrium was maintained in the sexual systems. Thus genic polymorphism was much higher for asexual than for sexual systems. This is discussed further in the next section.

Detecting costs of resistance and virulence

Costs of resistance and virulence play a central role in all theories of plant-pathogen polymorphism. Broad surveys of pathogens that attack crops sometimes show that virulence alleles are rare in places where matching resistance is rare, and virulence is common where matching resistance is common. These data are cited as evidence supporting the assumption that virulence-avirulence polymorphisms are maintained because of costs (Vanderplank, 1984). Often, however, geographic and temporal surveys fail to show a match between high frequencies of resistance and high frequencies of virulence. Because clear matching patterns are often not observed, many authors have questioned whether costs affect polymorphism (Crill, 1977; Parlevliet, 1981).

Abstract of results. The simulations show that there is only a weak tendency for a match between the frequencies of resistance and virulence. The match can be moderate for isolated, sexual populations with little immigration, but immigration or a tendency for epidemics reduces the match. The strength of the match fluctuates widely over time. Higher costs of virulence or resistance actually reduce the strength of the match in many circumstances.

The analysis of frequencies at individual loci (genic polymorphism) complements the genotypic analysis in the previous section. Higher levels of genic polymorphism are maintained in asexual systems because selection acts directly on each genotype and, with no recombination, only indirectly on each locus. In sexual systems selection acts directly on genotypes but, because of frequent recombination, each locus responds to selection as if it were an independent entity living within the genetic background determined by the gene frequencies at other loci.

Correlation between matching loci. The results are summarized in Fig. 14. Each distribution (vertical pair of lines and centre point) shows the Spearman rank correlation between frequencies at matching host and pathogen loci. The Spearman correlations were calculated as follows. The frequencies of resistance at the eight host loci were ranked from one to eight, the frequencies of virulence at the eight pathogen loci were ranked from one to eight, and the correlation was calculated in each generation based on the paired ranks of gene-for-gene matching loci in the hosts and pathogens. The distributions shown are the set of correlation coefficients obtained over 5000 generations of a single run.

The median correlation in the asexual model was typically greater than zero, but 25 percent or more of the observed correlations were negative (Fig. 14). Higher levels of immigration (high α),

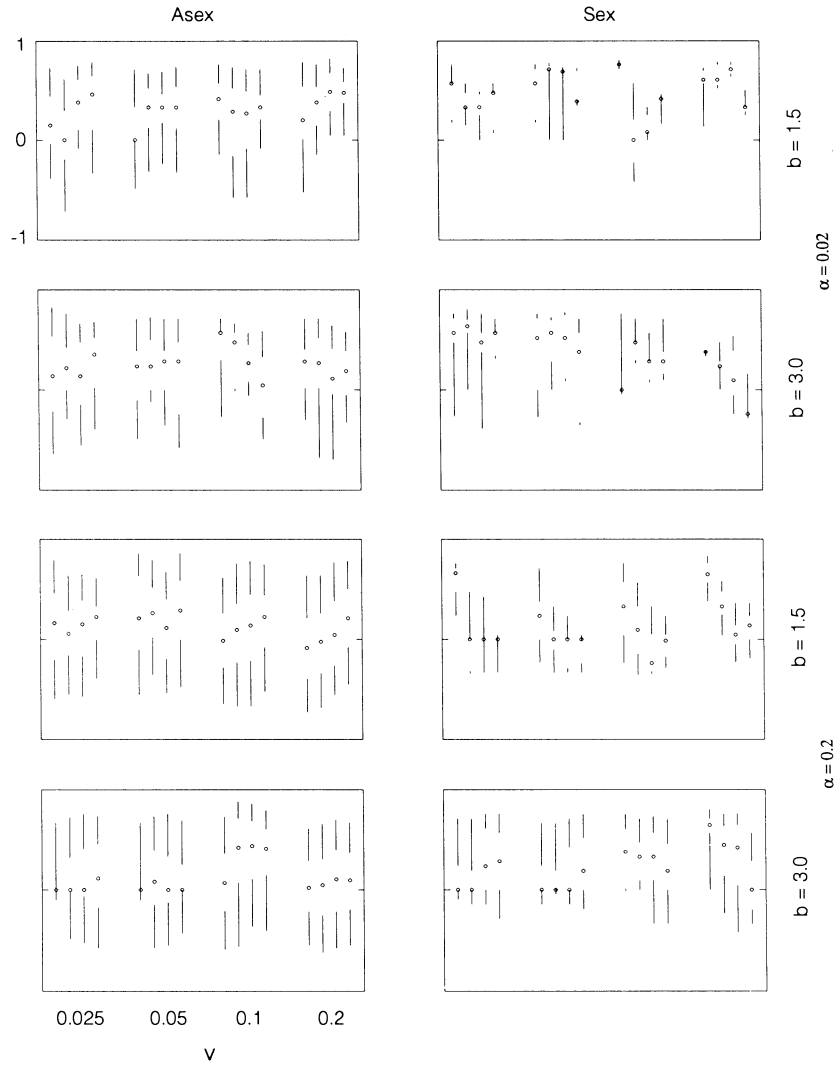

Figure 14. Match between frequency of resistance and virulence as measured by Spearman correlation coefficients. Each set of four closely spaced distributions shows an increasing cost of resistance from left to right of $a = 0.025,0.05,0.1,0.2$. Each distribution shows the values for one run of 5000 generations.

and a greater tendency for epidemics (high b) lower the correlations. The costs of resistance and virulence have only a small effect, and increased costs are as likely to reduce the correlation as they are to increase the correlation.

The correlations in the sexual model are strongly affected by the immigration rate. In isolated populations (low α) the correlations are typically greater than zero more than 95 percent of the time, although the correlations fluctuate considerably between zero and approximately 0.8. The correlation breaks down with more immigration, falling to zero or below 25 to 50 percent of the time.

Analysis of genic polymorphism at matching loci. The frequencies of matching resistance and

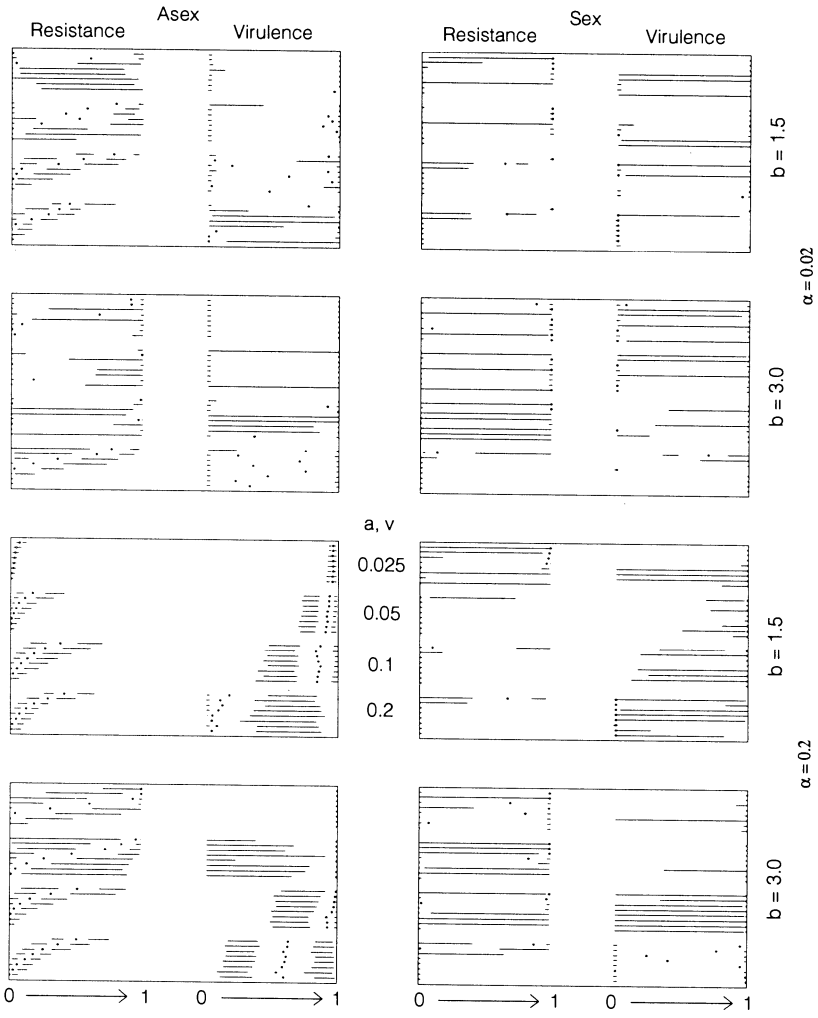

Figure 15. Frequencies of resistance and virulence at matching host and pathogen loci.

virulence alleles are shown in Fig. 15. The panel in the third row, left column is the easiest to read. Each set of eight closely spaced distributions shows the allele frequencies at the eight matching host and pathogen loci. In each generation the host loci are ranked according to their frequency of the resistance allele. On the left, the highest distribution in a set shows the distribution for the frequency of the resistance allele for the highest ranking locus. The next seven distributions show the next seven ranks. On the right the frequency distributions are shown for the matching pathogen loci. The highest distribution shows the frequency of virulence for the pathogen locus that matches the host locus with the highest frequency of resistance, and so on for the next seven lower ranks.

The panel in the third row, left column shows that resistance and virulence are correlated (Fig. 15). For example, the bottom set of distributions ($a = v = 0.2$) shows that as resistance declines virulence decreases. The small changes in the pathogen distribution show, however, that the correlation is weak. The distribution of correlation coefficients is shown in the matching panel

(third row, left column) of Fig. 14, in the distribution farthest to the right. The correlation is positive about 70% of the time but, in general, the correlation is weak.

Sex influences genetic polymorphism in many ways. This can be seen by comparing panels in the left and right columns of row three in Fig. 15. The frequency at any one locus tends to be very high or very low when there is sex, whereas linkage in asexual systems maintains much higher genic polymorphism by constraining frequencies between narrow intermediate ranges. The reason is that, with sex, selection can act relatively independently on each locus, whereas in asexual systems the genotype as a whole is affected directly and selection is much weaker on each locus. The wider swings in genic frequencies under sex provide more opportunity for a match between resistance and virulence frequencies. In the top set of distributions of the right panel, the median frequency of resistance is near one for the top four host loci, and the matching pathogen locus is nearly fixed for virulence. The distribution of correlation coefficients for this parameter set can be found in the matching panel in Fig. 14, the farthest distribution to the left. Note that the correlation is much more strongly positive for small costs ($a = v = 0.025$) than for large costs ($a = v = 0.2$) in the same panels.

Discussion

Polymorphism

Resistance polymorphisms. At equilibrium, the number of resistance alleles carried by each host follows a binomial distribution that depends on the number of loci, N, and the cost of

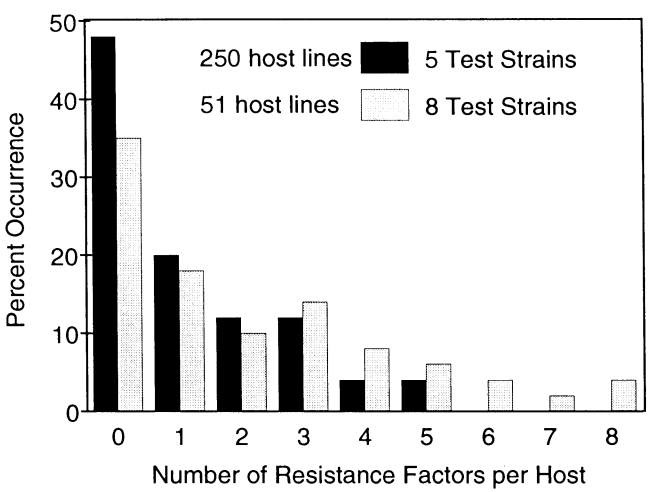

Figure 16. Distribution of the number of resistance factors per plant. The host is groundsel (*Senecio vulgaris*), a naturally occurring weedy species, and the pathogen is a powdery mildew (*Erysiphe fischeri*). The number of factors was inferred from the 28 distinct resistance phenotypes observed in 250 plant lines tested against five pathogen strains (black bars); the maximum number of detectable phenotypes under a gene-for-gene system with five test strains is $2^5 = 32$. Of these 250 lines, 51 were tested against the original five pathogen strains plus three additional strains (shaded bars), yielding 29 distinct resistance phenotypes. The different distributions result mainly from the different number of test strains rather than different genotypes of the plants. Limited genetic analysis supports the hypothesis that each factor is a single locus of a gene-for-gene interaction between host and pathogen (Clarke *et al.*, 1990). Sixteen of the 250 plant lines were derived from a primarily outcrossing race, the rest were derived from a race that reproduces almost entirely by self-fertilization. Data from Harry and Clarke (1986).

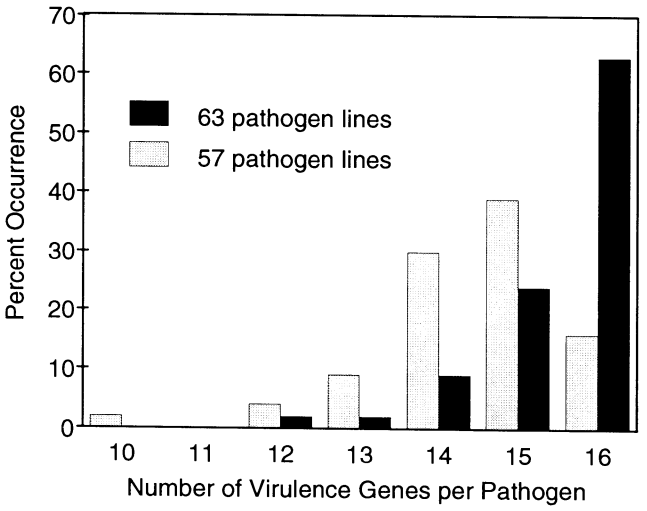

Figure 17. Distribution of the number of virulence genes per pathogen. The powdery mildew pathogens (*Erysiphe graminis*) were isolated from populations of wild barley (*Hordeum spontaneum*). The pathogen isolates were tested against barley cultivars of known resistance genotype; disease genetics for cultivars and pathogens follow the gene-for-gene pattern. Pathogens were obtained from five separate locations in Israel; among these five locations, the most skewed distribution (black bars, 63 isolates) and the least skewed distribution (shaded bars, 57 isolates) are shown. Data from Dinoor and Eshed (1987). One caveat: the pathogen population may be influenced by the use of resistant cultivars and migration between pathogen populations that attack cultivated and wild hosts.

virulence, v (Equation 10). This equilibrium holds exactly for asexual systems and is an excellent approximation when both hosts and pathogens are sexual (Fig. 8).

The cost of virulence is likely to be less than 25 or 30%, and may be quite small. Most hosts will therefore carry, at equilibrium, only a small number of resistance factors relative to the total number possible – the number of loci, N. As the system changes from stable to one characterized by epidemics, the number of resistance alleles per host increases significantly (Figs 8 and 9).

I was able to find only one data set from a natural population that reported the number of resistance factors per host. In the original article (Harry and Clarke, 1986) the data were presented in a table that showed inferred phenotype for each sample. I reduced the data to the number of loci with a resistance phenotype for each sample, and plotted these reduced data to show the distribution of the number of resistance factors per host (Fig.16). The data have the shape of a binomial distribution with few resistance alleles per individual as predicted for populations that are nearly stable (Equation 10).

Virulence polymorphisms. At equilibrium, the number of virulence alleles carried by each pathogen is approximately binomial and depends on the number of loci, N, and one minus the cost of resistance, $1 - a$ (Equation 13). This equilibrium is nearly exact for stable asexual systems (Fig. 10), and holds for sexual systems with small values of v (Fig. 11).

The cost of resistance is likely to be small. At equilibrium, most pathogens will carry nearly the maximum number of virulence factors, N. As the system changes from stable to unstable, the number of virulence factors per pathogen decreases significantly (Figs 10 and 11).

I was able to find only one data set from a natural population that reported the number of

virulence factors per pathogen. The original article (Dinoor and Eshed, 1987) presented a table with the inferred distribution of the number of loci with virulent phenotype per individual (Fig.17). The data have the shape of a binomial distribution with many virulence alleles per individual as predicted near equilibrium.

Frequency of resistance. The theoretical results show that the frequency of resistance per inoculation is typically low in stable systems and fluctuates widely in epidemic systems (Equation 9, Figs 5, 6 and 7). The low level of resistance in the stable coevolutionary system occurs because hosts carry few resistance alleles and pathogens carry many virulence alleles. Data from three wild populations support the conclusion that the frequency of resistance is often low (Parker, 1985; Clarke *et al.*, 1990; Jarosz and Burdon, 1990).

Genetic diversity. System stability and immigration each have simple effects on diversity. When the system is stable, diversity increases in the manner described by the equilibrium theory: rising costs of virulence, v, cause an increase in host diversity (top row of Fig. 12), and rising costs of resistance, a, cause an increase in pathogen diversity (top row of Fig. 13). As fluctuations increase, diversity decreases because more loci become fixed for particular alleles and because selection is, in general, more intense. Increased immigration enhances diversity by reducing the time between the loss of an allele and its subsequent reintroduction into the population.

Sex and recombination also affect diversity. In general, asexual models supported considerably more genotypic diversity (Figs 12 and 13), in spite of the fact that, for a given level of genic diversity, recombination causes sexual systems to support a higher level of diversity (greater evenness among genotypes) than asexual systems. The explanation for this difference is that asexual systems maintained much higher levels of genic diversity than sexual systems (Fig. 15) because, in asexual systems, only the genotype as a whole can respond to selection, and selection is much weaker on individual loci (Burdon, 1987a, pp. 148–51). Put another way, it appears that an unnecessary resistance or virulence allele is lost more rapidly from sexual than asexual populations.

Groth and Roelfs (1982) have reported the only large data set that compares diversity of sexual and asexual populations. In their study of cereal rusts, sexual populations of the pathogen were more diverse than asexual populations, which contradicts the prediction of the model. However, their study analysed an agricultural system, in which host genotypes are controlled, rather than a natural coevolutionary system.

Conditions for stability. The stability of the host-pathogen interaction has important effects on patterns of polymorphism, as illustrated in the previous paragraphs. The factors controlling stability are summarized in Figs 8–11, and were described in detail earlier. Two factors have the most prominent influence on stability and deserve special mention: the pathogen's intrinsic rate of increase, $b - s$, and the immigration rate, α.

Fluctuations in population sizes and gene frequencies rise as the pathogen's rate of increase rises. Explosive pathogens cause repeated bouts of epidemics followed by population crashes of hosts and pathogens; under these conditions more resistance alleles per host and fewer virulence alleles per pathogen are found than at equilibrium. Slow pathogens tend to be endemic, causing only moderate fluctuations in gene frequencies and population sizes.

Immigration rate also has a strong effect on stability. When immigration is rare, the system is less stable because the dynamics are driven in part by the chance extinction of rare genotypes and the highly variable times before the introduction of advantageous genotypes. Increasing immigration tends to stabilize the system.

Detecting costs of resistance and virulence

Theories of polymorphism invariably assume that resistance has a cost relative to susceptibility, and virulence has a cost relative to avirulence (Leonard and Czochor, 1980). It has, however, been difficult to determine whether such costs are really universal features of agricultural or natural systems (Burdon, 1987a; Parker, 1990). Vanderplank (1984) suggested that, in agricultural systems, a cost of virulence would lead to a positive association, across time and space, between the frequency of resistance and virulence alleles at matching gene-for-gene loci. The success of this method has been debated extensively by plant pathologists (e.g. Nelson, 1973; Crill, 1977; Parlevliet, 1981).

I examined whether a positive association develops between the frequencies of matching resistance and virulence alleles in a coevolutionary interaction. The results show that a weak association sometimes exists, but the association fluctuates widely and is often negative rather than positive (Fig. 14). Higher costs often lead to weaker rather than stronger associations. According to this model, costs of resistance and virulence in natural populations cannot be determined by spatial analysis of matching gene frequencies.

Magnitude of costs

How large must costs be in order to explain the observed patterns of genetic diversity? This question cannot be answered exactly because we know too little in practice about the actual patterns of diversity that occur in nature and too little in theory about how various processes generate diversity. The mathematical and simulation results suggest that costs on the order of 2% to 5% per locus are sufficient to generate considerable genotypic diversity. The cost per locus required to generate extensive diversity may decline as the number of loci in the interaction increases (Equations 3 and 5). This negative relationship between required cost per locus and number of loci agrees with a previous analysis on a different coevolutionary system (Frank, 1989).

Assumptions

The model presented here includes epidemiological interactions and coevolution between multilocus genetic systems in hosts and pathogens. These assumptions provide much greater realism than previous theories of gene-for-gene systems (Leonard and Czochor, 1980; Burdon, 1987a, Frank, 1992).

Many simplifying assumptions remain, however, and one can only guess at their relative influence on aspects of polymorphism. Some of the more important assumptions are: haploid rather than diploid genetics; symmetry in costs, that is, the same cost of resistance is assumed for each host locus, and the same cost of virulence is assumed for each pathogen locus; costs accumulate multiplicatively across loci; the breeding system was either of two extremes – asexual or outcrossing sexual; population subdivision and migration was simulated by assuming immigration from a patch whose genetic composition is uncorrelated with the local population.

It seems likely that most quantitative and qualitative aspects of the system will not be greatly affected by realistic departures from these assumptions. In the case of the breeding system, the two extremes of a continuum were analysed, and moderate levels of inbreeding should fall within the bounds established.

The model's spatial and temporal assumptions about epidemiology are difficult to evaluate at present. Lotka–Volterra dynamics assume uniform spatial mixing of host and pathogen genotypes. The model also assumes no immunological memory, that is, resistance characteristics of a host do not change in response to previous infection as in vertebrates. The discrete time steps

used in the model correspond to discrete generations in the sexual case and to a rhythm of birth and death in the asexual case. The time metric affects the amplification of feedback between changes in the host and pathogen populations (May, 1974; Leonard and Czochor, 1980; Seger, 1988), and therefore affects the overall stability of the system. Many different temporal assumptions are possible. On the whole, the discrete steps used here probably enhance feedback and instability relative to other reasonable assumptions that may be used.

The good match between the equilibrium predictions of the model and observations from wild populations suggest that, at least for some cases, the assumptions of the model provide a reasonable approximation for natural systems (Figs 16 and 17).

Testing the theory

The best method for testing the theory outlined in this paper is comparison among different host-pathogen systems. Some examples are given in the following three comparisons.

1. Endemic diseases, with low rates of pathogen increase, $b - s$, differ from epidemic systems, which have a relatively high rate of pathogen increase. Epidemic systems are expected to have relatively higher numbers of resistance factors per host, lower numbers of virulence factors per pathogen, and less genotypic diversity in both hosts and pathogens.
2. Isolated populations have a greater tendency toward epidemics than do mosaics of subdivided patches connected by occasional gene flow.
3. Sexual systems tend to be relatively less stable than asexual systems. Breeding system has only a small effect on the number of resistance and virulence factors, but in sexual systems the relative reduction of genotypic diversity can be significant. In particular, sexual systems are expected to have less genic diversity and a greater evenness in genotypic distribution than asexual or inbreeding systems.

Relatively higher or lower costs of resistance or virulence also shift patterns of polymorphism, but these costs are so difficult to estimate that this comparison is unlikely to be useful.

The methods that have been used to detect polymorphism can cause difficulties when testing these comparative hypotheses. For example, epidemic systems are predicted to have higher numbers of resistance factors in the hosts and lower genotypic diversity in both hosts and pathogens than endemic systems. Estimates of number of resistance factors and diversity may be confounded, however, because the detection of each resistance factor is made easier when greater genotypic diversity is available.

A second problem is that, when attributes of a system fluctuate over time and space, each estimate from a single population must be treated as one sample from a potentially broad distribution. Repeated samples over space and time are needed before any confidence can be attached to a particular conclusion (Thompson, 1988; Burdon *et al.*, 1989, 1990; Frank, 1989, 1991b).

Broad spatial and temporal sampling is, at present, an overwhelming task. However, molecular probes for resistance and virulence are now being developed, which will make sampling considerably easier (Hamer, 1991). The comparative hypotheses outlined here provide a good starting point for broad surveys of natural populations.

Acknowledgements

I thank A.M. Jarosz, B.R. Levin and Y. Michalakis for helpful comments on the manuscript. My research is supported by NSF grant BSR–9057331 and National Institute of Health grants GM42403 and BRSG–S07–RR07008.

References

Agrios, G. N. (1988) *Plant Pathology*, 3rd edn, Academic Press, NY, USA.

Barrett, J. A. (1985) The gene-for-gene hypothesis: parable or paradigm. In *Ecology and Genetics of Host-Parasite Interactions*. (D. Rollinson and R. M. Anderson, eds) pp. 215–25. Academic Press, New York.

Burdon, J. J. (1987a) *Diseases and plant population biology*. Cambridge University Press, Cambridge, UK.

Burdon, J. J. (1987b) Phenotypic and genetic patterns of resistance to the pathogen *Phakopsora pachyrhizi* in populations of *Glycine canescens*. *Oecol.* **73**, 257–67.

Burdon, J. J. and Jarosz, A. M. (1992) Temporal variation in the racial structure of flax rust (*Melampsora lini*) populations growing on natural stands of wild flax (*Linum marginale*): local *versus* metapopulation dynamics. *Plant Pathol.* (in press).

Burdon, J. J., Jarosz, A. M. and Kirby, G. C. (1989) Pattern and patchiness in plant-pathogen interactions – causes and consequences. *Ann. Rev. Ecol. Syst.* **20**, 119–36.

Burdon, J. J., Brown, A. H. D. and Jarosz, A. M. (1990) The spatial scale of genetic interactions in host-pathogen coevolved systems. In *Pests, Pathogens and Plant Communities*. (J. J. Burdon and S. R. Leather, eds) pp. 233–47. Blackwell, Oxford, UK.

Callow, J. A. (1987) Models of host-pathogen interaction. In *Genetics and Plant Pathogenesis*. (P. R. Day and G. J. Jellis, eds) pp. 283–95. Blackwell, Oxford, UK.

Clarke, D. D., Campbell, F. S. and Bevan, J. R. (1990) Genetic interactions between *Senecio vulgaris* and the powdery mildew fungus *Erysiphe fischeri*. In *Pests, Pathogens and Plant Communities*. (J. J. Burdon and S. R. Leather, eds) pp. 189–201. Blackwell, Oxford, UK.

Crill, P. (1977) An assessment of stabilizing selection in crop variety development. *Ann. Rev. Phytopath.* **15**, 185–202.

Day, P. R. (1974) *Genetics of Host-Parasite Interaction*. W. H. Freeman, San Francisco, USA.

Dinoor, A. and Eshed, N. (1987) The analysis of host and pathogen populations in natural ecosystems. In *Populations of Plant Pathogens: Their Dynamics and Genetics*. (M. S. Wolfe and C. E. Caten, eds) pp. 75–88. Blackwell, Oxford, UK.

Flor, H. H. (1956) The complementary genic systems in flax and flax rust. *Adv. Genet.*, **8**, 29–54.

Flor, H. H. (1971) Current status of the gene-for-gene concept. *Ann. Rev. Phytopath.* **9**, 275–96.

Frank, S. A. (1989) The evolutionary dynamics of cytoplasmic male sterility. *Am. Nat.* **133**, 345–76.

Frank, S. A. (1991a) Ecological and genetic models of host-pathogen coevolution. *Heredity* 67, 73–83.

Frank, S. A. (1991b) Spatial variation in coevolutionary dynamics. *Evol. Ecol.* **5**, 193–217.

Frank, S. A. (1992) Models of plant-pathogen coevolution. *Trends Genet.* (in press).

Gabriel, D. W. and Rolfe, B. G. (1990) Working models of specific recognition in plant-microbe interactions. *Ann. Rev. Phytopath.* **28**, 365–91.

Groth, J. V. and Roelfs, A. P. (1982) The effect of sexual and asexual reproduction on race abundance in cereal rust fungus populations. *Phytopathology* **72**, 1503–7.

Groth, J. V. and Roelfs, A. P. (1987) Analysis of virulence diversity in populations of plant pathogens. In *Populations of Plant Pathogens: Their Dynamics and Genetics*. (M. S. Wolfe and C. E. Caten, eds) pp. 63–74. Blackwell, Oxford, UK.

Hamer, J. E. (1991) Molecular probes for rice blast disease. *Science* **252**, 632–3.

Harry, I. B. and Clarke, D. D. (1986) Race-specific resistance in groundsel (*Senecio vulgaris*) to the powdery mildew *Erysiphe fischeri*. *New Phytol.* **103**, 167–75.

Jarosz, A. M. and Burdon, J. J. (1990) Predominance of a single major gene for resistance to *Phakopsora pachyrhizi* in a population of *Glycine argyrea*. *Heredity* **64**, 347–53.

Jarosz, A. M. and Burdon, J. J. (1991) Host-pathogen interactions in natural populations of *Linum marginale* and *Melampsora lini*: II. Local and regional variation in patterns of resistance and racial structure. *Evolution* (in press).

Leonard, K. J. and Czochor, R. J. (1980) Theory of genetic interactions among populations of plants and their pathogens. *Ann. Rev. Phytopath.* **18**, 237–58.

Magurran, A. E. (1988) *Ecological Diversity and its Measurement*. Princeton University Press, Princeton, USA.

May, R. M. (1974) *Stability and Complexity in Model Ecosystems*, 2nd edn, Princeton University Press, Princeton, USA.

Nelson, R. R. (1973) The use of resistance genes to curb population shifts in plant pathogens. In *Breeding Plants for Disease Resistance*. (R. R. Nelson, ed) pp. 49–66. Pennsylvania State University Press, University Park, USA.

Parker, M. A. (1985) The pleiotropy theory for polymorphism of disease resistance genes in plants. *Evolution* **44**, 1872–5.

Parlevliet, J. E. (1981) Stabilizing selection in crop pathosystems: an empty concept or a reality? *Euphytica* **30**, 259–69.

Person, C. (1959) Gene-for-gene relationships in host:parasite systems. *Can. J. Bot.* **37**, 1101–30.

Seger, J. (1988) Dynamics of some simple host-parasite models with more than two genotypes in each species. *Phil. Trans. R. Soc. Lond. B* **319**, 541–55.

Thompson, J. N. (1988) Variation in interspecific interactions. *Ann. Rev. Ecol. Syst.* **19**, 65–87.

Vanderplank, J. E. (1984) *Disease Resistance in Plants*, 2nd edn, Academic Press, NY, USA.