The cost of resistance and the maintenance of genetic polymorphism in host–pathogen systems

JANIS ANTONOVICS AND PETER H. THRALL

Department of Botany, Duke University, Durham, North Carolina 27708, U.S.A.

SUMMARY

By using models which incorporate both numerical and gene-frequency dynamics, we investigate the conditions for a stable polymorphism in host disease resistance when there is a genetically uniform pathogen. We show that polymorphism is more likely when the difference in resistance conferred by alternative alleles is large rather than small. This conforms with the frequent observation of major gene effects on resistance. Moreover, when allelic differences in resistance are large, polymorphism is possible over a wide range of costs, including situations where costs approach values close to zero. The actual resistance cost that can be sustained in such polymorphic populations is dependent on the transmission mode and the intensity of disease-independent population regulation. Expectations regarding resistance costs in any particular host–pathogen system will be dependent on knowledge of the epidemiological and genetic characteristics of that system.

1. INTRODUCTION

It has frequently been argued that for there to be a stable genetic polymorphism in disease resistance the fitness of the resistant genotypes should be less than that of the susceptible genotypes in the absence of the disease (Leonard 1977; Vanderplank 1984; Parker 1992). The argument is intuitively obvious: without such a cost, an allele for resistance should continue to increase in frequency as long as some disease is present. The expectation that a resistance cost should be present in systems whenever hosts are variable with regard to their resistance to pathogens or herbivores has led to a search for costs in both natural populations and agricultural systems (Fritz & Simms 1992). However, the difficulty of demonstrating such costs has led to an extensive debate regarding methodologies used to detect costs, whether appropriate ecological complexities leading to indirect costs have been considered, and whether the failure to find costs simply reflects the fact that many host–pathogen systems may not be in any kind of evolutionary equilibrium (Parker 1992; Simms 1992a, b).

The biological basis of the resistance cost may be quite varied (Simms 1992a). Most simply, the cost could be a direct pleiotropic effect of the alleles for resistance: the biochemical, morphological or phenological features conferring resistance may have detrimental effects on fitness in the absence of the disease. Alternatively, the detrimental effects could manifest themselves indirectly through increased susceptibility to another pathogen genotype, or another species of pathogen or predator. Clearly, these costs may themselves evolve. Costs present at the mutational origin of a resistance gene may themselves be subsequently reduced as a result of selection for genes that modify the way the costs are expressed (Lenski 1988).

Much of the discussion of resistance costs in host–pathogen systems has been based on purely genetic models (Mode 1958, 1960; Jayakar 1970; Leonard & Czochor 1980). However, one can envisage that the outcome might depend critically on the disease dynamics. For example, the spread of an allele for resistance might itself affect host population size and disease prevalence, and this in turn may affect the relative fitness of resistant and susceptible genotypes. Gillespie (1975), in one of the first papers to integrate epidemiological and genetic approaches, showed that the rate of spread of a resistance allele with costs depended not only on allele frequency but also on population density.

Here we report results from the analysis of simple models that incorporate numerical as well as gene-frequency dynamics. In particular, we investigate the conditions for a protected polymorphism of alleles that confer disease resistance but have pleiotropic effects on fitness in the absence of disease; we assume the pathogen is genetically uniform. We compare two transmission modes. In one case, disease transmission is dependent on the fraction of infectives in the population: such ‘frequency-dependent’ transmission is expected in sexually and vector-transmitted diseases where the disease-carrying agents actively search out their hosts (Getz & Pickering 1983; Thrall et al. 1993, 1994; Antonovics et al. 1994). The other transmission mode we consider is the case where the probability of a healthy host becoming diseased is a function of the density of infectives: this is likely to be characteristic of diseases where transmission is by random contact or passive scattering of disease agents into the environ-
2. THE MODELS

We assume that the pathogen is a microorganism, such that the diseased and healthy hosts can be represented without recourse to explicit modelling of pathogen numbers (Anderson & May 1981). We assume that there are two alleles at one locus determining the level of resistance in the host, that the host is haploid, and that the pathogen is genetically uniform. In addition, we assume that, although the disease infects the two types of hosts differentially, diseased hosts of both types transmit the disease equally. A pleiotropic cost of resistance is incorporated as a reduced birth rate of the more resistant genotype, both in the presence and absence of the disease. For simplicity, we model the situation where population growth is continuous, but with the realization that we may be overestimating the stability of model equilibria relative to the situation when generations are discrete.

Because host–pathogen coexistence is not possible in the frequency-dependent transmission case when there is no disease-independent population regulation (Getz & Pickering 1983), we include density-dependent population regulation in the models. Specifically, we represent population regulation by a linear dependence of host birth rate (b) on total population density (N) (see Getz & Pickering 1983):

\[ b = b_1 - \gamma N, \]  

(1)

where \( b_1 \) is the intrinsic birth rate of the \( i \)th host, and \( \gamma \) is a constant representing the strength of density-dependence.

We have shown previously (Thrall et al. 1993) that when there is frequency-dependent transmission the conditions for host–pathogen coexistence are least restrictive if density-dependent forces act more strongly on the healthy class, if the disease is sterilizing but does not affect mortality, and if there is no host recovery. These features are characteristic of many sexually transmitted diseases (Holmes 1983; Felman 1986; Alexander & Antonovics 1988; Thrall et al. 1993), and we include them also in models incorporating density-dependent transmission for comparative purposes.

We can represent the numerical dynamics of a haploid host–pathogen system by using the following system of differential equations:

\[ \frac{dX_i}{dt} = X_i (b_1 - \mu - \gamma N) - P_i X_i \]  

(2)

\[ \frac{dX_3}{dt} = X_3 (b_1 - \mu - \gamma N) - P_3 X_3 \]  

(3)

\[ \frac{dY}{dt} = P_1 X_1 + P_2 X_2 - \mu Y, \]  

(4)

where \( X_i \) represents the numbers of the \( i \)th host genotype, \( Y \) is the number of infectious (infected) hosts, i.e. the two infected host genotypes do not differ in their ability to transmit the disease, \( N = X_1 + X_2 + Y \), \( \mu \) is the disease-independent host mortality rate, and \( P_i \) is the force of infection (a measure of the likelihood that the \( i \)th host genotype will become infected). If disease transmission is density dependent, then \( P_i \) will take the form

\[ P_i = \beta_i Y, \]  

(5)

where \( \beta \) is the disease transmission parameter. Alternatively, if transmission depends on the frequency of diseased individuals, then \( P_i \) can be represented by

\[ P_i = \beta_i Y/N. \]  

(6)

In the analyses that follow, we let \( X_i \) represent the numbers of the more susceptible host, and \( X_3 \) the numbers of the more resistant host; we therefore assume \( \beta_1 > \beta_2 \) and \( b_1 > b_2 \). We will henceforth term \( X_1 \) ‘susceptible’ and \( X_3 \) ‘resistant’, remembering that these terms refer to relative, not absolute, degrees of resistance. We make the further assumption that the relation between cost of resistance (measured as decreased birth rate) and the benefit acquired due to the lower probability of becoming infected can be represented by a ratio (\( \theta \)), such that

\[ \theta = (b_1 - b_2)/(\beta_1 - \gamma). \]  

(7)

In other words, \( \theta \) is the amount by which fitness of the resistant host is reduced for a given amount of ‘protection’ from disease.

In the case of frequency-dependent disease transmission, it can be shown that a necessary condition for a protected polymorphism is that

\[ \langle \beta_2 - \mu \rangle / \beta_2 < \theta < \langle \beta_1 - \mu \rangle / \beta_1. \]  

(8)

If inequality (8) is met, there will be a single valid internal equilibrium given by

\[ X_1^* = \frac{\langle \beta_2 (\theta - 1) + \mu \rangle (b_2 - \gamma)}{\langle \beta_2 - \beta_1 \rangle \gamma}, \]  

(9)

\[ X_3^* = \frac{\langle \beta_1 (\theta - 1) + \mu \rangle (b_2 - \gamma)}{\langle \beta_2 - \beta_1 \rangle \gamma}, \]  

(10)

\[ Y^* = \theta \langle b_2 - \mu - \beta_2 \theta \rangle / \gamma. \]  

(11)

It is straightforward to show that, under inequality (8), these equilibrium values are always stable (see Appendix 1 for details of the analysis).

For the case in which disease transmission is density dependent, a necessary condition for existence of a protected polymorphism is given by

\[ \langle \beta_2 (b_1 - \mu) - \mu \gamma \rangle / \beta_2 (\beta_1 + \gamma) < \theta < \langle \beta_1 (b_2 - \mu - \mu \gamma) \rangle / \beta_1 (\beta_2 + \gamma). \]  

(12)

It can be shown that, if inequality (12) is met, there will be a single valid and stable internal equilibrium given by

\[ X_1^* = \frac{\langle \beta_2 (b_2 - \mu - \beta_2 \theta) (1 + \gamma) \rangle - \mu \gamma}{\langle \beta_2 - \beta_1 \rangle \gamma}, \]  

(13)

\[ X_3^* = \frac{\langle \beta_1 (b_2 - \mu + \theta (b_2 + \gamma)) + \mu \gamma \rangle / \langle \beta_2 - \beta_1 \rangle \gamma}{\langle \beta_2 - \beta_1 \rangle \gamma}, \]  

(14)

\[ Y^* = 0. \]  

(15)

Details of the stability analysis for the density-dependent transmission cases are given in Appendix 2.

It follows from inequalities (8) and (12) that the valid range for a protected polymorphism is greatest when the difference in disease susceptibility of the genotypes is large. Moreover, it can be seen (figure 1) that the range of costs that can be sustained in a polymorphic population are far greater for values of \( \beta_2 \)
Figure 1. Phase plane diagrams showing regions of stable equilibria in a host–pathogen system with a genetically variable host and uniform pathogen. (a) Frequency-dependent transmission. (b) Density-dependent transmission. Cross-hatched regions show fixation of the susceptible genotype, shaded regions show stable polymorphism, and open regions show fixation of the resistant genotype. For resistant genotypes the transmission rate (β) and the birth rate (b) vary over the regions shown. For susceptible genotypes, β = 1.0 in the frequency-dependent case, and β = 0.01 in the density-dependent case; in both cases b = 0. Equilibrium regions are based on existence conditions (8) and (12) in text, where μ = 0.2 and γ = 0.01. Note that, at the boundary condition of b = 0 (i.e. zero resistance costs), either the resistant type goes to fixation or the polymorphism is neutral if the disease is lost from the population.

as they approach zero (complete resistance) than for values of b close to b (highly susceptible). Indeed, completely resistant individuals can be maintained in a polymorphic state with highly susceptible individuals over a broad range of costs and always if resistance incurs some small cost.

For the frequency-dependent transmission case, the condition for increase in a monomorphic population is simply β > μ. It can be shown by substitution of equation (6) into (4) that a necessary condition for increase of the disease class is that the mean transmission rate, (Xβ + Xβ)/μ, be greater than μ. Provided that the disease spreads in a population of susceptibles (μ > μ), genetic polymorphism is possible in cases when β > b (i.e. the disease in a monomorphic population would lead to population extinction) and β < μ (i.e. the disease would fail to spread in a monomorphic population).

For the density-dependent transmission case, the condition for disease increase in a monomorphic population is that Xβ > μ. It can be shown, following substitution of equation (5) into (4), that the necessary condition for increase of the diseased class, Y, is Xβ + Xβ > μ. From this it follows that, provided that Xβ > μ, genetic polymorphism is also possible even when Xβ < μ (i.e. when the disease would fail to spread in a monomorphic population). Therefore, under both transmission modes, the conditions for disease and population persistence are broadened by the existence of a genetic polymorphism; completely resistant individuals can coexist with completely susceptible ones.

In the case of density-dependent transmission, but not with frequency-dependent transmission, host–pathogen coexistence as well as polymorphism (and the corresponding range of permissible costs) depend on the parameter γ which is a measure of the strength of density dependence and hence the population carrying capacity. The condition for disease spread when the pathogen is rare (with uniform host and pathogen) is given by β > μγ/(μ−μ). When γ is large, equilibrium population size, (μ−μ)/γ, may be too small to allow disease spread. Given that γ is sufficiently small to allow disease spread, the conditions allowing polymorphism become less restrictive as γ decreases and the carrying capacity in the absence of disease increases.

The numerical models outlined above were modified to represent changes in gene frequencies alone (i.e. assuming a constant population density). This was done by making the assumption that X and Y represent frequencies (i.e. sum to 1) in equations (2)–(4); to obtain recursions in frequencies we then divided each of these equations by their sum. It is then straightforward to show that no polymorphic equilibrium exists under these conditions. Jayakar (1970), by using a haploid genetic model, obtained an identical result for the case where one genotype was completely resistant.

3. DISCUSSION

The results presented here justify the intuitive expectation that for there to be a genetic polymorphism in level of disease susceptibility disease resistance has to be associated with some cost. However, the expectation that highly resistant individuals should incur a correspondingly high cost to coexist with susceptibles is not borne out. Completely resistant individuals can theoretically coexist with highly susceptible ones even when the cost of resistance approaches values that are close to zero. The reason is that the presence of the resistant hosts reduces the incidence of the disease to a lower frequency or density, and hence reduces the ‘force of infection’, such that susceptible hosts can invade when at low frequency. Conversely, when susceptibles are common, the ‘force of infection’ is high and resistants can invade. This creates a broad region for polymorphism. However, when the two genotypes are very similar in their disease susceptibility, the conditions for a protected polymorphism become restricted. This is because introduction of an alternative allele that differs
only slightly in resistance has little effect on the disease incidence. An allele conferring a small gain in resistance can therefore only be maintained in a polymorphic state with an alternative allele if the cost precisely balances the gain in the fitness due to the increased resistance. If there is no cost then the resistant type will increase in frequency until the disease is lost; subsequently, susceptible and resistant types coexist in neutral stability. The general shape of the region for polymorphism (i.e. wide and abutting the boundary condition of zero costs when the resistance was extreme relative to the susceptible genotype, yet more restricted and not abutting the boundary of zero costs when resistance was less extreme) was the same regardless of the values of the various model parameters, although the actual size of the region was dependent on specific parameter values.

Several other studies have considered resistance polymorphisms in an epidemiological context. May & Anderson (1983) pointed out that when one genotype is universally resistant a cost of resistance is necessary for polymorphism, but they did not explore the conditions for this to occur. Gillespie (1975) and Kemper (1982) derived equilibrium conditions for situations where there was a uniform pathogen and a genetically variable host, one genotype was completely resistant (and resistance had a fitness cost), and where there was density-dependent transmission. Gillespie (1975) considered within-season epidemics but a constant population size over successive generations; Kemper (1982) ignored within-season epidemiology but assumed host recovery. Examination of the conditions for polymorphism in both of these cases also reveals that polymorphism is more likely over a wider range of costs when there are extremes of resistance and susceptibility. Jayaker (1970), Gillespie (1975) and Kemper (1982) also showed that conditions for allele frequency equilibria are qualitatively similar for haploid and diploid cases (assuming there is no heterozygote advantage). Therefore, it appears our results are general for a wide range of disease models.

Our theoretical results argue that if resistance variation is found in natural systems it may often be of an extreme nature, and indeed this expectation is borne out by a wide range of studies in natural and agricultural systems (Burdon 1987; Kennedy & Barbour 1992; Parker 1992). Such patterns have been shown by our own studies (see, for example, Alexander & Antonovics 1988; Alexander 1989, 1990; Antonovics & Alexander 1992; Alexander et al. 1993; Thrall & Jarosz 1994a, b) with the anther-smut disease (Ustilago violacea) of white campion (Silene alba), where there is genetic variation for extreme levels of disease resistance in the host, but little variation in the pathogen (Alexander 1989; Alexander et al. 1993; Thrall & Jarosz 1994a, b). Resistance is determined both physiologically and phenologically (Alexander et al. 1993), and these types of resistances appear to be independent (although sample sizes have precluded a rigorous test of this). Phenological resistance is largely the result of later onset of flowering, and this undoubtedly has a substantial fitness effect in terms of reduced reproduction over the season; no cost has been associated with the physiological resistance (Alexander et al. 1993). Clearly, the models presented in this paper show that both ‘costly’ and ‘cheap’ resistances could be maintained as stable polymorphisms.

In a recent review, Parker (1992) stated that the evidence from natural and agricultural systems suggests that ‘large fitness costs are not a general attribute of genes for disease resistance’, and infers that ‘it is unlikely that the cost of resistance theory can provide a general explanation for resistance polymorphism’. Our results argue that the failure to find substantial costs to resistance does not logically preclude the possibility of resistance polymorphisms; polymorphism of alleles with extreme effects is possible even when costs approach values that would only be detectable with difficulty in empirical studies. However, it is also true that, when resistance costs are very low, genes for susceptibility may be rare and difficult to detect.

In nature, it is possible that indirect costs (e.g. increased susceptibility to alternative pathogen genotypes or species) may be more important than direct physiological costs in maintaining polymorphisms in disease resistance (Parker 1992). However, even multi-locus gene-for-gene models make the assumption that there is some resistance (and virulence) cost, otherwise ‘all-purpose’ genotypes carrying suites of resistance genes should become fixed in populations (Vanderplank 1984; Burdon 1987). Although gene-for-gene models have rarely incorporated numerical dynamics (Leonard 1977; Barrett 1988; Seger 1988), a simulation model of a multi-locus gene-for-gene system incorporating numerical dynamics showed that polymorphisms could be maintained over a broad range of costs (Frank 1993).

It is interesting that the equilibrium numbers of diseased individuals in the density-dependent transmission case is simply the cost–benefit ratio and independent of \( \gamma \) (the degree of density-dependent population regulation). A similar result was derived by Gillespie (1975), by using the classical Kerack & McKendrick (1927) model for epidemic spread of a disease. From an empirical standpoint, if it can be established that disease transmission is density dependent then the cost of resistance is simply the difference in disease resistance (measured by the transmission coefficients) multiplied by the disease prevalence. Given the many oversimplifications implicit in such heuristic models, it is unlikely that this would provide an accurate estimate of the real resistance cost; however, such an estimate might be of value in determining the sample sizes needed to detect costs by direct fitness measures of genotypes.

Because single-locus haploid models are in many ways equivalent to two-species models (Read et al. 1994), it follows that the ‘resistant genotype’ could also represent a separate coexisting species. Our models would apply to this situation if the two species were regulated by similar factors (i.e. were equivalent with regard to \( \gamma \)). In the case of frequency-dependent transmission (e.g. by vectors or pollinators), the mere presence of a resistant species (provided that pollinators or vectors are equally attracted to it) may actually ‘protect’ a highly susceptible species from going to

---

extinction by reducing the ‘effective’ frequency of the disease. In the case of density-dependent transmission, the presence of a second resistant species, by effectively decreasing the density of the susceptible species, may also reduce the disease prevalence or even prevent the disease from persisting. These results show that ‘indirect’ interactions among species mediated through pathogens may be dependent on transmission mode as well as the details of their ecological interactions (Holt & Pickering 1985; Begon & Bowers 1994).

Host–pathogen systems present a range of life cycles, transmission modes, resistance mechanisms and ecologies. Clearly, ‘permissible’ resistance costs are likely to vary from case to case, depending on both the genetics and ecology of the interaction. Our results further emphasize the general point that if we are to understand the phenomenon of genetic variation in resistance, it is imperative to develop models that simultaneously incorporate gene frequency and numerical dynamics (May & Anderson 1983). Resorting to intuitions in complex interacting systems can be very misleading.

This work was supported by NSF grant DEB-9119626, a Guggenheim Fellowship to J.A., and a Royal Society Guest Research Fellowship on behalf of J.A. to M. P. Hassell. We thank E. L. Simms for helpful comments on the manuscript.

REFERENCES


which is satisfied for $F = b_y - \beta_2 \theta [F^*]$. The equilibrium values given by (9)–(11) are arrived at by substitution of $F^*$ into (A 1)–(A 3). Positivity of $X^*_1, X^*_2,$ and $Y^*$ is guaranteed under inequality (8).

(b) Stability of the internal equilibrium

The characteristic equation $(C(\lambda))$ calculated from the Jacobian matrix is given by the cubic

$$C(\lambda) = -X^* + a_2 X^2 + a_3 X + a_4,$$

(A 5)

where

$$a_2 = (\theta - 1) (b_2 - \mu - \beta_4 \theta),$$

$$a_3 = -\theta \mu (b_2 - \mu - \beta_4 \theta) - [\beta_4 (1 - \theta) - \mu] [\beta_4 (1 - \theta) - \mu],$$

and

$$a_4 = \theta [\beta_4 (1 - \theta) - \mu] [\beta_4 (1 - \theta) - \mu] (b_2 - \mu - \beta_4 \theta).$$

The coefficients $a_2, a_3,$ and $a_4$ are all negative, therefore the real parts of all eigenvalues of (A 5) must also be negative, and the internal equilibrium is always stable when it exists. Although we have not been able to prove this, numerical evaluation of equations (2), (3) and (4) for the frequency-dependent case show that the eigenvalues consist of one real and a pair of complex conjugates; the moduli of the complex roots must be larger than that of the real root because all simulations showed convergence to be oscillatory.

APPENDIX 2. Density-dependent transmission

As in the frequency-dependent transmission case, we derive the equilibrium values for $X^*_1, X^*_2$ and $Y^*$ by substitution of the variable $F$ and the expression for the cost–benefit ratio $\theta$ into equations (2)–(4), but with $P^i$ taking the form given in equation (5)), and the cost–benefit ratio $\theta$, and then solving the resulting expressions to arrive at:

$$X^*_1 = (F - \mu) [\beta_4 \theta (\beta_2 - \beta_3) + \beta_2 (F + \beta_1 - \beta_4 - \beta_4 \mu)]$$

(A 1)

$$X^*_2 = (F - \mu) [\theta (\beta_1 - \beta_4) + F + \beta_1 - \beta_4 - \mu] / (\beta_1 (\beta_2 - \beta_4) \gamma),$$

(A 2)

$$Y^* = (F - \mu) [\theta (\beta_2 - \beta_4) + F - \beta_4] / (\beta_2 \gamma).$$

(A 3)

Substitution of the equilibrium values (A 1), (A 2) and (A 3) into equation (3) results in an equilibrium expression in terms of $F$,

$$\beta_1 (\beta_2 - \beta_4) (b_2 - \beta_4 \theta - F) = 0,$$

(A 4)