Theoretical Population Biology **53**, 75–84 (1998)
Article No. TP971342

TPB

Epistatic Selection in a Multi-locus Levene Model and Implications for Linkage Disequilibrium

Thomas Wiehe and Montgomery Slatkin

Department of Integrative Biology, University of California, Berkeley, California 94720-3140

Received January 2, 1997

We analyze a multiple-locus extension of the Levene (1953) model of population subdivision. We show that stable or quasistable linkage disequilibrium between two selected loci can be maintained even with free recombination, provided that there is a strong enough epistatic interaction. We then consider the dynamics of a third neutral locus and show that its approach to linkage equilibrium depends on the recombination rates *and* the selection intensities. There is an embedding or hitchhiking effect that extends the time during which a neutral locus which is closely linked to one of the selected loci remains in disequilibrium with both selected loci. Therefore, strong disequilibrium between two loci does not necessarily indicate that those loci are themselves selected, but it does indicate that there is strong selection acting at least on nearby loci. This property implies a warning that screening for linkage disequilibrium as a tool to identify functionally important sites in a genome can be misleading.

1. INTRODUCTION

The maintenance of persistent linkage disequilibrium between pairs of loci indicates the presence of some force opposing its decay through recombination. There is a long history in population genetics theory of examining models that predict permanent linkage disequilibrium. In fact, the term linkage disequilibrium was introduced by Lewontin and Kojima (1960) in describing the first such model. In the (diploid) Lewontin and Kojima model and its descendants (e.g., Karlin and Feldman, 1970, Feldman *et al.*, 1974), genetic polymorphism at each locus is maintained by overdominant selection, and permanent linkage disequilibrium will persist provided that there is sufficiently strong epistasis. Here, we show that a haploid model (therefore without any dominance effects) with population subdivision can have similar properties. Population substructure may be created by the interaction of infectious parasites with hosts, of which different individuals may be immune to different strains of the parasite. Selection, therefore, acts differently in the different environments. We show that this type of selection induces frequency–dependent selection, maintains polymorphism and, with appropriate epistatic interactions, linkage disequilibrium. The selection induced this way always favors rare types. The prototype of such a model has been introduced by Levene (1953). He showed that population substructure can lead to stable maintenance of polymorphism at one genetic locus, even in absence of overdominance. Gliddon and Strobeck (1975) derived necessary and sufficient conditions on the fitness parameters for the haploid analogue. Felsenstein (1981) used a three locus Levene model with a selection scheme similar to the one presented here as a model of speciation.

It has been suggested that linkage disequilibrium between loci would indicate which loci are antigenically active and responsible for maintaining strain differences in a parasite (Gupta *et al.*, 1996). That suggestion, which has considerable practical importance, is valid provided that linkage disequilibrium between a selected locus and a neutral locus would be expected to decay sufficiently rapidly so that there would be little chance to falsely

0040-5809/98 $25.00

identify the neutral locus as being selected. We show that selection does retard the decay of linkage disequilibrium between a neutral and both selected loci. In particular, if the neutral locus is closely linked to one of the selected loci, it will remain in linkage disequilibrium with both selected loci for much longer than would be expected based on the recombination rate alone. Therefore, the suggestion that linkage disequilibrium could serve as an indicator for functionally important sites is valid only when there has been time for linkage equilibrium with neutral loci to be established. The observation of strong disequilibrium would, however, indicate the presence of selection at nearby loci even before an equilibrium is reached.

2. THE MODEL

We first examine a two locus model. We assume a haploid species which, at some point in its life cycle, undergoes recombination. Selection and reproduction/recombination are temporally and, potentially, spatially separated. Such a scenario is adequate for a large number of infectious pathogens and viruses, such as *Plasmodium* or the Influenza virus. We model two biallelic loci (A/a and B/b) and assume the population to be sufficiently large so that genetic drift can be ignored. The distinguishing feature of the Levene (1953) model and its descendants (Maynard Smith and Hoekstra, 1980) is that selection takes place in separate selective environments ("niches" in the Levene's terminology). However, individuals from each environment randomly mate among the entire population with the contribution from each selective environment being determined by the proportion of that environment. The alternative model is one in which the contribution of each environment is in proportion to the mean fitness in each environment. The terms "hard" and "soft" selection are sometimes used to describe this difference (Wallace, 1968). "Soft selection" means that population regulation by selection occurs locally within subdivisions (Wade, 1985). The regulation of population density in each environment separately induces frequency–dependent selection that can prevent the loss of low frequency alleles. We will show that the same frequency dependence also maintains linkage disequilibrium. We consider three niches: selection operates in environments I and II ("immunized" hosts), environment III is neutral ("susceptible" hosts). In environments I and II, one of the haplotypes (AB in I and ab in II) has no fitness reduction. The fitnesses of the complementary haplotype (ab in I and AB in II) is $1-s$, where s is the intensity of selection ($0 \leqslant s \leqslant 1$). If $s=1$, then one of the haplotypes is completely eliminated when passing from one generation to the next. The amount of epistasis is regulated by a selection parameter γ of the recombinant haplotypes. The fitnesses of Ab and aB in both environments (I and II) are $1-\gamma$ ($0 \leqslant \gamma \leqslant 1$). In environment III all haplotypes have fitness 1. In the immunological framework used by Gupta *et al.* (1996), γ indicates the degree of *cross protection* of the host immune system against recombinant haplotypes of the parasite.

The proportions of the three environments are determined by the weights c_i ($\sum c_i = 1$). For simplicity, we assume that the two selective environments contribute equally, so that $c_1 = c_2 = c$. Table 1 lists the haplotypes and their fitnesses.

The change of frequencies of the four haplotypes is described by the difference equation

$$\Delta x_i = (x_i^S - x_i) + \eta_i r\, D(\mathbf{x}^S), \tag{1}$$

where Δ indicates the change in one generation, the superscript S marks the respective variables after selection and before recombination. $\eta_1 = \eta_4 = -1$, $\eta_2 = \eta_3 = 1$, $1 \leqslant i \leqslant 4$, and the population is censused after recombination. The coefficient of linkage disequilibrium in Eq. (1), D^S, is

$$D(\mathbf{x}^S) = x_1^S x_4^S - x_2^S x_3^S \tag{2}$$

and

$$x_i^S = \sum_{j=I}^{III} c_j \frac{x_i w_i^j}{\bar{w}^j}. \tag{3}$$

Assuming that selection acts on viability, a sketch of the (haploid) life cycle is depicted in Fig. 1.

As is well known, selection in the different types of niches induces frequency–dependent selection when regarded from the perspective of the whole population,

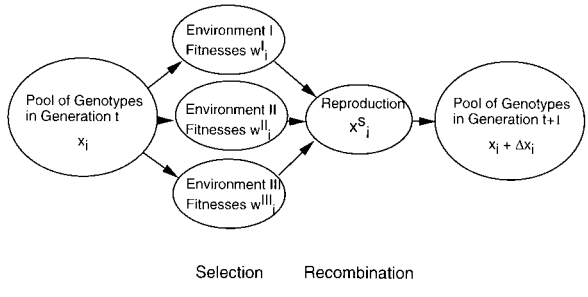

FIG. 1. Haploid life cycle. Selection and recombination are separated.

TABLE 1

Selection Scheme for the Two Locus Model

	Haplotype	Frequency	Fitness in environment I	Fitness in environment II	Fitness in environment III
(1)	AB	x_1	$w_1^I = 1$	$w_1^{II} = 1-s$	$w_1^{III} = 1$
(2)	Ab	x_2	$w_2^I = 1-\gamma$	$w_2^{II} = 1-\gamma$	$w_2^{III} = 1$
(3)	aB	x_3	$w_3^I = 1-\gamma$	$w_3^{II} = 1-\gamma$	$w_3^{III} = 1$
(4)	ab	x_4	$w_4^I = 1-s$	$w_4^{II} = 1$	$w_4^{III} = 1$

even though selection in each niche is not frequency–dependent. This can be seen by defining

$$u_i = c_1 w_i^I \bar{w}^{II} + c_2 w_i^{II} \bar{w}^I + c_3 w_i^{III} \bar{w}^I \bar{w}^{II}, \qquad (4)$$

$\bar{u} = \sum x_i u_i$ and replacing x_i^S in Eq. (3) by $x_i^S = (x_i u_i)/\bar{u}$. Thus, the relative fitness of each haplotype depends on the x_i. In particular, one finds that u_i decreases as x_i increases, which means that the individual haplotypes are advantageous when rare. One can understand the frequency–dependence in biological terms by noting that, for example, when AB individuals are in low frequency they will contribute disproportionately to the next generation because they have an advantage in a fixed fraction c_1 of the niches.

3. RESULTS

3.1. Equilibrium Analysis

It is useful to work with allele frequencies A and B and the coefficient of linkage disequilibrium D instead of haplotype frequencies. The relationship between both sets of variables is

$$\begin{aligned} x_1 &= AB + D \\ x_2 &= A(1-B) - D \\ x_3 &= (1-A)\, B - D. \end{aligned}$$

With this transformation the difference equations simplify to

$$\begin{aligned} \Delta A &= A^S - A \\ \Delta B &= B^S - B \\ \Delta D &= (1-r)\, D^S - D \end{aligned} \qquad (5)$$

(again, $(.)^S$ means the variables after selection). For general parameters the equations are coupled. For special cases they may simplify. For instance, with $s = 1$ and $\gamma = 1$ they are (valid in the interior of their domain of definition)

$$\begin{aligned} \Delta A &= c(1-2A) \\ \Delta B &= c(1-2B) \\ \Delta D &= (1-r)(c\,(1-2\,c)(1-A-B+2AB) \\ &\quad + (1-2c)\, D + c^2) - D. \end{aligned} \qquad (6)$$

For any parameters c, r, s or γ there is a polymorphic equilibrium $(\bar{A}, \bar{B}, \bar{D})$ with $\bar{A} = \bar{B} = 1/2$. The stability properties and $\bar{D}$ itself depend on the parameters c, r, s and γ. The general expression $\bar{D}$ is given in the Appendix (A.1). We note that linkage disequilibrium $\bar{D}$ increases as the intensity of selection (s) decreases (if γ is held fixed). This is because smaller s effectively favors the coupling haplotypes and therefore generates linkage disequilibrium in *each* subpopulation.

We examined the stability of the solution in the entire selection parameter space $0 \leqslant s, \gamma \leqslant 1$. The local stability can be determined by means of the Jacobian matrix. Evaluated at $(\bar{A}, \bar{B}, \bar{D})$, the Jacobian has the form

$$J = \begin{pmatrix} j_1 & j_2 & 0 \\ j_2 & j_1 & 0 \\ 0 & 0 & (1-r)\, j_6 \end{pmatrix}. \qquad (7)$$

The j_i are given in the Appendix (A.4). The modulus of the third eigenvalue λ_3 (it coincides with the lower right entry of J) is less than 1 for all parameter choices. For the other eigenvalues, one finds $|\lambda_1| < 1$, if $s < 2\gamma$ ("positive additive" or "diminishing" (Kimura and Maruyama, 1966) epistasis), independently of c and r. $|\lambda_2| < 1$, if s is larger than a rather complicated expression which depends on γ, r and c. The eigenvalues λ_i are given in the Appendix (A.2) for the special case $r = 1/2$ and $c = 1/2$. The two conditions, $|\lambda_1| < 1$ and $|\lambda_2| < 1$, define a subset of the s–γ–unit square where the equilibrium $(\bar{A}, \bar{B}, \bar{D})$ is stable. This subset depends on the recombination rate r and the weight parameter c. For the extreme case of free recombination ($r = 1/2$) the region of stability is shaded

in Fig. 2. Combined, the regions which are shaded in light and dark gray indicate the case when $c = 1/2$ (i.e., absence of the neutral niche III). For arbitrarily small, but positive, c the domain of stability does not vanish: it is the region shaded in dark gray. It is surprising, that the population can settle at an equilibrium with non-zero linkage disequilibrium, even if the environment is almost entirely neutral, but if selection in the scarcely populated niches I and II is sufficiently strong—as could be the case at the onset of an infectious disease, when the major part of the host population is susceptible to the infection (see Fig. 2 when $c \to 0$, s and γ large). This scenario is qualitatively different from one in which selection is present in all niches, but weak (Fig. 2 when $c = 1/2$, s and γ small). For any recombination rate, the part of the unit-square where $s > 2\gamma$ (i.e., when epistasis is "reinforcing") remains unstable. Only if epistasis is positive additive ($s < 2\gamma$), the equilibrium $(\bar{A}, \bar{B}, \bar{D})$ may be stable. This property is equivalent to a threshold condition on the parameter γ of cross protection found by Gupta *et al.* (1996). Within the set $s < 2\gamma$, the region of stability becomes larger as recombination becomes less frequent. For example, the dotted line in Fig. 2 defines, together with the line $s = 2\gamma$, the region of stability of $(\bar{A}, \bar{B}, \bar{D})$ for $r = 1/20$ and $c = 1/2$. Examples which show that $(\bar{A}, \bar{B}, \bar{D})$ is in general not unique with the property $D \neq 0$ can be constructed. For instance, letting $\gamma = 1$, $r = 1/2$, $c = 1/2$, then it can be seen from a bifurcation diagram (s treated as bifurcation parameter, see Fig. 3) that there are at least three different polymorphic equilibria, if $s \in (0.67, 0.74)$.

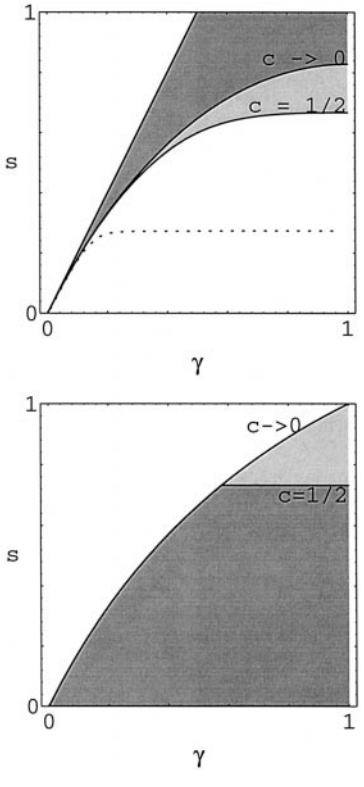

FIG. 2. (a) Whether the polymorphic equilibrium $(\bar{A}, \bar{B}, \bar{D})$ is locally stable depends on the parameters. The shaded areas indicate the stable regions if recombination between $\mathscr{A}$ and $\mathscr{B}$ is free ($r = 1/2$): dark gray if c is small (preponderance of the neutral environment III), and light together with dark gray if $c = 1/2$ (absence of neutral environment). The dotted line, together with $s = 2\gamma$, circumscribes the region where $(\bar{A}, \bar{B}, \bar{D})$ is stable when $c = 1/2$ and $r = 1/20$. (b) Region of stability (shaded) for the non–polymorphic equilibria (1) and (5) from Table 2 ($r = 1/2$).

All equilibria with $D = 0$ are non–polymorphic (except, when $s = 2\gamma$; then, $(1/2, 1/2, 0)$ is an equilibrium). A complete list is given in Table 2. Therefore, for any steady state, the statements "both loci are polymorphic" and "$D \neq 0$" are equivalent (non–zero linkage disequilibrium always implies polymorphism).

Equilibria with $D = 0$ are independent of the weight parameter c and the recombination rate r. However, their local stability depends on parameter choices for c, r, s and γ. For instance, the $s - \gamma$ parameter region for which equilibria (1) and (5) are stable is indicated as shaded region in Fig. 2b. The border of this region is determined by the functions $s = (2\gamma)/(1 + \gamma)$ (case $c \to 0$) and $s = \min((2\gamma)/(1 + \gamma), \sqrt{3} - 1)$ (case $c = 1/2$). Equilibria (2) and (4) are stable if $s > 2\gamma$. Equilibria (3), (6), (7) and (8) exist only if $(2\gamma)/(1 + \gamma) < s < 2\gamma$. There, none of them is stable.

Finally, we note that transient stability (quasistability) is possible for positive additive epistasis (see Fig. 4).

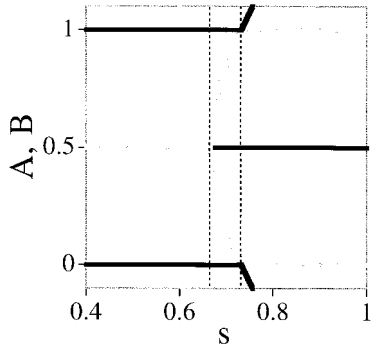

FIG. 3. Bifurcation diagram for equilibria of the allele frequencies A and B vs. selection intensity s. Stable and unstable equilibria are indicated as black and gray lines, respectively. The polymorphic equilibrium $\bar{A} = \bar{B} = 1/2$ is unstable if $s < 0.67$ and undergoes a subcritical pitchfork bifurcation at $s \approx 0.67$, where three polymorphic equilibria branch off. $\bar{A} = \bar{B} = 1/2$ is stable for $s > 0.67$, the other two branches are unstable. The monomorphic equilibria ($A = B = 0$ and $A = B = 1$) are stable if $s < 0.74$, undergo a transcritical bifurcation at $s \approx 0.74$ and are unstable for $s > 0.74$. Other parameters: $\gamma = 1$, $r = 1/2$, $c = 1/2$.

TABLE 2

Steady States with Linkage Equilibrium ($D=0$)

	Frequency of A	Frequency of B
(1)	0	0
(2)	0	1
(3)	0	$\frac{2\gamma-s-\gamma s}{2\gamma(\gamma-s)}$
(4)	1	0
(5)	1	1
(6)	1	$\frac{(1-\gamma)(s-2\gamma)}{2\gamma(\gamma-s)}$
(7)	$\frac{2\gamma-s-\gamma s}{2\gamma(\gamma-s)}$	0
(8)	$\frac{(1-\gamma)(s-2\gamma)}{2\gamma(\gamma-s)}$	1
(9)[a]	$\frac{1}{2}$	$\frac{1}{2}$

Note. Equilibria which may be stable for suitable parameters are boldfaced.

[a] Only if $s=2\gamma$.

However, based on the linearized system, it is not possible to determine the time duration, during which the system is quasistable (with $D\neq 0$ and both loci polymorphic).

Choosing $s=\gamma=1$ (complete cross protection), the model becomes what is known as the matching alleles model (Frank, 1993). Then, in the immunological example, hosts in environment I and II resist when *at least one* of the relevant alleles of the infectious parasite is recognized by the immune system. The selection scheme is

	I	II
AB	1	0
Ab	0	0
aB	0	0
ab	0	1.

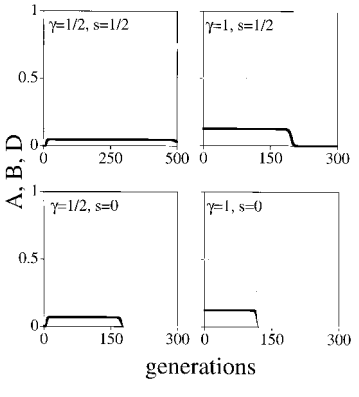

FIG. 4. For positive additive epistasis, the polymorphic equilibrium can be quasistable: polymorphism and non-zero linkage disequilibrium may be maintained for a long time before eventually decaying to zero. Plotted are allele frequencies A and B (gray lines) and linkage disequilibrium D (black line) vs. time. Parameters are $r=1/2$ and $c=1/2$, initial conditions are $A(0)=0.99$, $B(0)=0.01$, $D(0)=0$.

For some values of c and r, linkage disequilibrium is

c	r	$\bar{D}$
0.5	0.5	0.125
	0	0.25
0.05	0.5	0.023
	0	0.25 .

Obviously, if linkage is tight ($r=0$), then strong linkage disequilibrium is permanently maintained, even if the neutral environment dominates ($c=0.05$).

3.2. Dynamic Analysis

For a neutral two locus model linkage equilibrium is approached at a rate which is determined by the recombination rate, $D(\tau)=(1-r)^{\tau}\approx e^{-r\tau}$, τ denoting time in generations. For the two locus Levene model, linkage disequilibrium may be approached much more rapidly, depending on γ, s and the c_i. It is in general not the recombination rate that dominates the speed of convergence. The convergence rates for A, B and D are obtained from the eigenvalues of $J-E_3$ (E_3 is the 3×3 unit matrix). These eigenvalues are λ_i-1. In the linearized system, linkage disequilibrium is decoupled from the allele frequencies (cf. the structure of J above). Therefore, the convergence speed of D is determined by $|\lambda_3-1|$ and the convergence speed for the allele frequencies, A and B, is ultimately determined by the smaller of the two eigenvalues, $\min(|\lambda_1-1|, |\lambda_2-1|)$. In the special case of complete cross protection ($s=\gamma=1$), D approaches ($\bar{A}$, $\bar{B}$, $\bar{D}$) at a rate

$$\rho_D = 2c + r(1-2c). \tag{8}$$

Allele frequencies go somewhat more slowly, at a rate

$$\rho_{A,B} = 2c, \tag{9}$$

independently of r.

To approximate the rates for arbitrary parameters, one may use a Taylor series expansion. For example, the first order approximation of ρ_D in γ (around $\gamma=1$) and for $s=1$ is

$$\rho_D \approx \left| \frac{16c(1-r)(1-\gamma)(2c(1-r)+r)^2}{(4c(1-r)+r)^2} - (2c(1-r)+r) \right| \tag{10}$$

(see Fig. 5 for a plot of ρ_D and $\rho_{A,B}$). The rate ρ_D depends strongly on c. Furthermore, it depends strongly on the recombination rate (r) when c is small, but strongly on γ when c is large. The rate $\rho_{A,B}$ depends strongly on γ and on c, weakly on r.

3.3. Three Loci

To simplify notation, let the three loci be arranged in the fixed order $\mathscr{A}-\mathscr{N}-\mathscr{B}$ and assume that $\mathscr{A}$ and $\mathscr{B}$ are selected loci and $\mathscr{N}$ is neutral. Recombination rates are r_1 and r_2 for recombination between $\mathscr{A}-\mathscr{N}$ and $\mathscr{N}-\mathscr{B}$, respectively. For $r_1=r_2=1/2$ recombination between each pair of loci is free, in particular, it is free between both ends of the region ($\mathscr{A}-\mathscr{B}$). The relationship with r is

$$r = r_1(1-r_2) + (1-r_1)\, r_2,$$

when there is no interference in recombination. There are eight haplotypes, numbered as in Table 3. The three locus model reduces to the former two locus model when alleles at $\mathscr{N}$ are pooled. The right hand sides of the three locus recombination–selection equations contain four summands, one describing selection and three describing the possible recombination events. The equation for haplotype frequency y_1 is given in the Appendix (A.3), the others are analogous.

We concentrate on properties of the linkage functions D_{AB}, D_{BN} and D_{AN}. We transform variables into allele

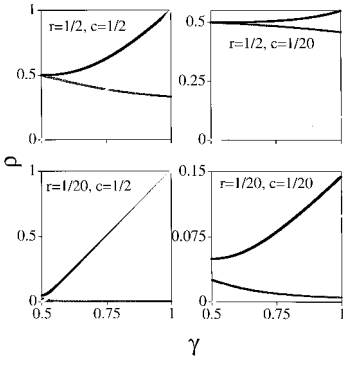

FIG. 5. Convergence rates $\rho_{(.)}$ versus cross protection γ. Black lines: decay rate $\rho_{D_{AB}}$ of linkage disequilibrium between the selected loci. Dark gray lines: decay rate $\rho_{D_{AN}}$ and $\rho_{D_{NB}}$ of linkage disequilibria between selected and neutral loci. Light gray lines: rate with which the allele frequencies (A, B) approach their equilibrium values. Here, $r_1=r_2$ and $r=r_1(1-r_2)+r_2(1-r_1)$. Selection $s=1$.

TABLE 3

Enumeration of the Haplotypes for the Three Locus Model

	Haplotype	Frequency	polymorphic equilibrium $\bar{y}$
(1)	ANB	y_1	$\bar{N}(\frac{1}{4}+\bar{D}_{AB})$
(2)	ANb	y_2	$\bar{N}(\frac{1}{4}-\bar{D}_{AB})$
(3)	AnB	y_3	$(1-\bar{N})(\frac{1}{4}+\bar{D}_{AB})$
(4)	Anb	y_4	$(1+\bar{N})(\frac{1}{4}-\bar{D}_{AB})$
(5)	aNB	y_5	$\bar{N}(\frac{1}{4}-\bar{D}_{AB})$
(6)	aNb	y_6	$\bar{N}(\frac{1}{4}+\bar{D}_{AB})$
(7)	anB	y_7	$(1-\bar{N})(\frac{1}{4}-\bar{D}_{AB})$
(8)	anb	y_8	$(1-\bar{N})(\frac{1}{4}+\bar{D}_{AB})$

Note. $\bar{N}$, the equilibrium frequency of the N allele depends on initial conditions. $\bar{D}_{AB}$, linkage disequilibrium between the selected loci $\mathscr{A}$ and $\mathscr{B}$, is given in Eq. (A.1).

frequencies (A, N, B), second and third order linkage disequilibria: $D_{AB}=[AB]-A\cdot B$, $D_{AN}=[AN]-A\cdot N$, $D_{NB}=[NB]-N\cdot B$, and $D_{ANB}=[ANB]-A\cdot N\cdot B-A\cdot D_{NB}-N\cdot D_{AB}-B\cdot D_{AN}$ (the symbols in brackets are haplotype frequencies and single letters are allele frequencies). Then, the difference equations simplify to

$$\begin{aligned}
\Delta A &= A^S - A \\
\Delta N &= N^S - N \\
\Delta B &= B^S - B \\
\Delta D_{AN} &= (1-r_1)\, D^S_{AN} - D_{AN} \\
\Delta D_{NB} &= (1-r_2)\, D^S_{NB} - D_{NB} \\
\Delta D_{AB} &= (1-r_1-r_2+2r_1r_2)\, D^S_{AB} - D_{AB} \\
\Delta D_{ANB} &= (1-r_1)(1-r_2)\, D^S_{ANB} - D_{ANB}.
\end{aligned} \tag{11}$$

Properties of equilibria are analogous to the two locus case. However, a single equilibrium of the two locus model will in the three locus model generally become a manifold of equilibria, which can be parametrized in terms of the neutral allele frequency N. In particular, there is a polymorphic equilibrium such that the selected loci are in linkage disequilibrium and each selected locus is in linkage equilibrium with the neutral one. It is

$$\begin{pmatrix} \bar{A} \\ \bar{N} \\ \bar{B} \\ \bar{D}_{AN} \\ \bar{D}_{NB} \\ \bar{D}_{AB} \\ \bar{D}_{ANB} \end{pmatrix} = \begin{pmatrix} 1/2 \\ * \\ 1/2 \\ 0 \\ 0 \\ (\text{see Eq. (A.1)}) \\ 0 \end{pmatrix}, \tag{12}$$

where $*$ indicates that $\bar{N}$ may take any value (which one depends on the initial conditions). Obviously, only $\bar{D}_{AB}$ depends on the parameters of the model. Its stability properties are those of the two locus model. To determine the convergence rates we use again the Jacobian matrix. Evaluated at the above equilibrium the Jacobian has the structure

$$\begin{pmatrix} j_1 & 0 & j_2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ j_2 & 0 & j_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1-r_1)\,j_4 & (1-r_1)\,j_5 & 0 & 0 \\ 0 & 0 & 0 & (1-r_2)\,j_5 & (1-r_2)\,j_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & j_3 \\ 0 & 0 & 0 & 0 & 0 & (1-r_1-r_2+2r_1r_2)\,j_6 & 0 \\ & & & & & 0 & (1-r_1)(1-r_2)\,j_7 \end{pmatrix}. \tag{13}$$

The expressions j_i are given in the Appendix (A.4). As can be seen from the structure of J, allele frequencies A and B (corresponding to the first and third rows in J) are—in the linearized system—decoupled from any linkage disequilibria. Furthermore, D_{AB} and D_{ANB} (rows 6 and 7) are decoupled, and D_{AN} and D_{NB} (rows 4 and 5) are coupled only among each other. N (row 2) is coupled only to D_{ANB}. The eigenvalues are

$$\mu_1 = 1$$

$$\mu_2 = j_1 - j_2 \quad \text{(identical to } \lambda_1\text{)}$$

$$\mu_3 = j_1 + j_2 \quad \text{(identical to } \lambda_2\text{)}$$

$$\mu_4 = \frac{\left[(2-r_1-r_2)\,j_4 - \sqrt{(r_1-r_2)^2\,j_4^2 + 4(1-r_1)(1-r_2)\,j_5^2}\right]}{2} \tag{14}$$

$$\mu_5 = \frac{\left[(2-r_1-r_2)\,j_4 + \sqrt{(r_1-r_2)^2\,j_4^2 + 4(1-r_1)(1-r_2)\,j_5^2}\right]}{2}$$

$$\mu_6 = (1-r_1-r_2+2r_1r_2)\,j_6 \quad \text{(analogous to } \lambda_3\text{)}$$

$$\mu_7 = (1-r_1)(1-r_2)\,j_7$$

and, obtained from these, the convergence rates are

$$\begin{aligned} \rho_{A,B} &= \min(|\mu_2 - 1|, |\mu_3 - 1|) \\ \rho_{D_{AN}, D_{NB}} &= \min(|\mu_4 - 1|, |\mu_5 - 1|) \\ \rho_{D_{AB}} &= |\mu_6 - 1|. \end{aligned} \tag{15}$$

Two time scales determine the change of the variables A, B, D_{AN} and D_{NB}. Ultimately important is the slow change characterized by the two minima above. Initially, however, the change is fast and characterized by the maximum of the respective eigenvalues (cf. Fig. 7).

In the special case of complete cross protection, the rates become

$$\rho_{A,B} = 2c \tag{16}$$

$$\rho_{D_{AB}} = 2c + (r_1 + r_2 - 2\,r_1 r_2)(1-2c). \tag{17}$$

Choosing $c = 1/2$ the (slow) rate for D_{AN} and D_{NB} is

$$\rho_{D_{AN}, D_{NB}} = \frac{2r_1 r_2}{2 - r_1 - r_2 + 2r_1 r_2}. \tag{18}$$

The (transient) fast rate in this case is $\rho_{D_{AN}, D_{NB}} = 1$, independently of r_1 or r_2. However, the ultimate approach to linkage equilibrium with the neutral locus is governed by the slow rate. Hence, linkage equilibrium might be established very slowly. In contrast, D_{AB} will approach its equilibrium value very rapidly and, what is surprising, independently of the recombination rates. For comparison, in the absence of selection ($c = 0$ or $\gamma = s = 0$) or in the absence of epistasis ($s = 2\gamma$), equations for D_{AN} and D_{NB} are decoupled ($j_5 = 0$). Therefore, the rates with which linkage equilibrium is approached, are then

$$\begin{aligned} \rho_{D_{AN}} &= r_1 \\ \rho_{D_{NB}} &= r_2 \\ \rho_{D_{AB}} &= r_1 + r_2 - 2r_1 r_2, \end{aligned}$$

in accordance with standard theory.

However, with positive additive epistasis present, both quantities, D_{AN} and D_{NB}, are coupled and their approach of linkage equilibrium is (i) retarded with respect to the neutral (decoupled) case and (ii) not independent of each other. If the three loci are not equidistantly spaced then the smaller of the two recombination rates essentially determines $\rho_{D_{AN}, D_{NB}}$, whereas the larger one determines $\rho_{D_{AB}}$. In particular, tight linkage of the neutral locus to one selected locus also retards the approach of linkage equilibrium between the neutral and a more distant selected locus (see Fig. 7). Notice that linkage disequilibrium between the neutral and a selected

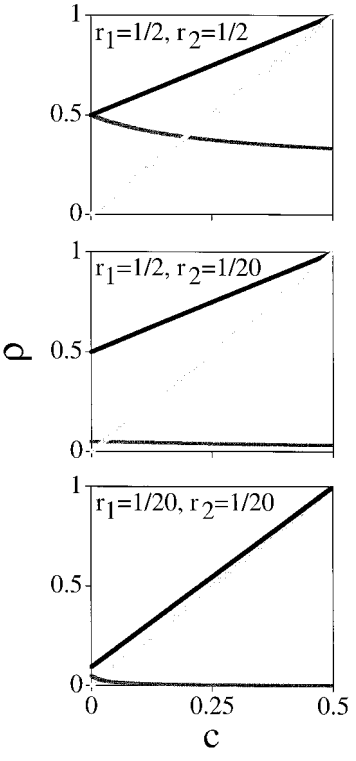

FIG. 6. Convergence rates vs. c. Legend as in Fig. 5. Selection $s = 1$ and cross protection $\gamma = 1$.

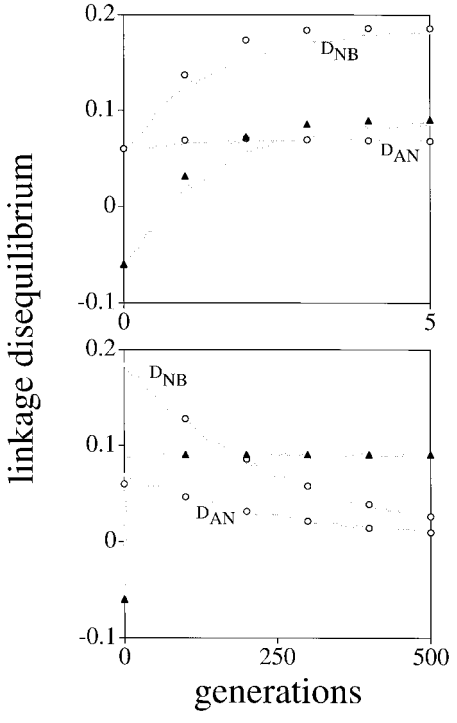

FIG. 7. D_{AN}, D_{NB} (open circles) and D_{AB} (filled triangles) vs. time (in generations), based on the discrete recursion Eq. (11). D_{AB} approaches its equilibrium (≈ 0.09) very rapidly. In contrast, D_{AN} and D_{NB} evolve on two time scales, corresponding to the eigenvalues μ_2 and μ_3 (see Eq. (14)). For this particular choice of parameters, there is an initial rapid increase (upper figure) and then a slow decay (lower figure) to their equilibrium (0). Numerical values of the rates, calculated from the eigenvalues of the Jacobian are $\rho_{D_{AB}} = 0.734$, $\rho_{D_{AN},D_{NB}}$ (slow) = 0.004, $\rho_{D_{AN},D_{NB}}$ (fast) = 0.726. The gray lines are plots of the function $\sigma_1 + \sigma_2 \exp(-\rho\tau)$, where ρ is the respective rate and σ_i are fitted parameters. Other parameters: $\gamma = 0.85$, $s = 0.95$, $c = 0.45$, $r_1 = 0.5$, $r_2 = 0.005$. Initial conditions: $A(0) = B(0) = 0.6$, $N(0) = 0.9$, $D_{AN}(0) = D_{NB}(0) = 0.06$, $D_{AB}(0) = -0.06$, $D_{ANB}(0) = -0.042$.

site may well exceed linkage disequilibrium between the selected sites for an appreciable amount of time. Generally, smaller recombination rates also retard convergence (see Fig. 6). Table 4 summarizes qualitatively the results outlined above.

Conditions on parameters such that the dynamical system will settle at an equilibrium with $D_{AB} = 0$ are

TABLE 4

Correlation between the Magnitude of Parameters and Convergence Rates

	Strongly		Weakly	
Allele frequencies ($\rho_{A,B}$)	cross protection γ	+	recombination	−
	selection s	+		
Linkage equilibrium ($\rho_{D_{AN},D_{NB}}$)	recombination	+	selection s	−
			cross protection γ	−
Linkage disequilibrium ($\rho_{D_{AB}}$)	selection s	+		
large c:	cross protection γ	+	recombination	+
small c:	recombination	+	cross protection γ	+

Note. + indicates positive, − negative correlation.

analogous to the two locus model. The equilibria listed in Table 2 and their stability properties carry over to the three locus model.

4. DISCUSSION AND CONCLUSIONS

Our analysis was motivated by the study of Gupta *et al.* (1996) but our model is quite different. We follow the frequencies of different haplotypes of a parasitic organism under the assumption that the pool of hosts and the properties of those hosts (which account for the selection schemata) are fixed, whereas Gupta *et al.* model the number of hosts that are resistant to different haplotypes. Although the biological bases of the two models are different, their dynamics are similar because in the Gupta *et al.* model the numbers of hosts of different types reflect the frequencies of different haplotypes of the parasite. Both models contain the same essential features, namely the selection imposed on the parasite by hosts resistant to different haplotypes, the importance of cross–protection to different haplotypes, and the importance of population subdivision in maintaining linkage disequilibrium. Our results show that the linkage disequilibrium results from the frequency–dependence created by population structure, provided that the contribution of each type of host to the reproductive pool of the parasite is fixed, as in the Levene (1953) model.

The relative simplicity of our model allows us to consider a third neutral locus and show that it approaches linkage equilibrium with the selected loci on a time scale determined by the selection intensities and recombination rates. There is an embedding or hitchhiking effect that retards the decay of linkage disequilibrium between the neutral locus and the two selected loci. Therefore, the observation of linkage disequilibrium between two loci does not necessarily imply that those two loci are themselves under selection, as was suggested by Gupta *et al.* (1996), although the observation of linkage disequilibrium would indicate the presence of selection at least at a nearby locus.

Selection on parasitic species seems to fall naturally into the framework of the well–known Levene model in population genetics. This model has the inherent property of favoring low frequency haplotypes, because of the separate regulation of population density in each subpopulation or "niche". The Levene model and its generalizations proved useful before to address questions in theoretical ecology (Maynard Smith and Hoekstra, 1980; Gillespie and Langley, 1976) or as a model of speciation (Felsenstein, 1981).

APPENDIX

A.1. The general expression for linkage disequilibrium, $\bar{D} = \bar{D}(r, c, s, \gamma)$, in the two locus model and when allele frequencies are $\bar{A} = \bar{B} = 1/2$ is

$$\bar{D} = \frac{\left[\begin{array}{c} r(4-s-2\gamma) \\ -\sqrt{r^2(4-s-2\gamma)^2 + 8c(1-r)(2c(1-r)+r)(s-2\gamma)^2} \end{array}\right]}{8(2c(1-r)+r)(s-2\gamma)}. \tag{A.1}$$

For the three locus model, the corresponding expression $\bar{D}_{AB}$ is obtained if r is replaced by $r_1 + r_2 - 2\,r_1\,r_2$.

A.2. Eigenvalues of the Jacobian J of system (5), for $r = 1/2$ and $c = 1/2$ are

$$\begin{aligned} \lambda_1 &= \frac{16(1-\gamma)}{3(4-2\gamma-s)+a} \\ \lambda_2 &= \frac{\left[\begin{array}{c} 16(6(2\gamma(1-\gamma)(2-s) - s(2-2s+\gamma s)) \\ -s^2(1-\gamma) + (2\gamma - s - \gamma s)\,a) \end{array}\right]}{(2\gamma-s)(a+3(4-2\gamma-s))^2} \\ \lambda_3 &= \frac{64(1-\gamma)(2-s)}{3(4-2\gamma-s)+a}, \end{aligned} \tag{A.2}$$

where $a = \sqrt{9(2\gamma-s)^2 + 8(2(1-\gamma) - s + \gamma s)}$. To make $(\bar{A}, \bar{B}, \bar{D})$ stable it is necessary to satisfy $|\lambda_2| < 1$. Solving $\lambda_2 = 1$ for s defines, together with $s = 2\gamma$, the boundary of the domain of stability, shaded in light gray in Fig. 2. For small γ ($\gamma < 1/4$), the bounding function is well approximated by

$$s_{c=1/2}(\gamma) \approx \frac{2\gamma}{1+\gamma}.$$

If $c \to 0$, the boundary is given by

$$s_{c\to 0}(\gamma) = 2(\sqrt{1+2\gamma-\gamma^2} - 1)\,.$$

A.3. The selection–recombination difference equation for the three locus model (here, for haplotype 1; see Table 3) is

$$\begin{aligned} \Delta y_1 = (y_1^S - y_1) \\ &+ r_1(1-r_2)(y_5(y_2+y_3+y_4) - y_1(y_6+y_7+y_8)) \\ &+ (1-r_1)\,r_2(y_2(y_3+y_5+y_7) - y_1(y_4+y_6+y_8)) \\ &+ r_1 r_2\,(y_3(y_2+y_5+y_6) - y_1(y_4+y_7+y_8)) \end{aligned} \tag{A.3}$$

A.4. The entries of the Jacobian matrix, evaluated at $\bar{A} = 1/2$, $\bar{B} = 1/2$, $\bar{D}_{AN} = \bar{D}_{NB} = \bar{D}_{ANB} = 0$ are

$$j_1 = \frac{\left[\begin{array}{l} 4(2-\gamma)(2-\gamma-s)+32(1-c) \\ \quad \times \bar{D}(2\gamma-\gamma^2-s)+(1-2c)\,b_1 \end{array}\right]}{b_2^2}$$

$$j_2 = \frac{8c((2-\gamma)\,\gamma+4\,\bar{D}\gamma(\gamma-s)-s)}{b_2^2}$$

$$j_3 = \frac{8c(2\gamma-s)}{b_2}$$

$$j_4 = \frac{(4-2\gamma-s)+4\bar{D}(1-2c)(2\gamma-s)}{b_2} \qquad \text{(A.4)}$$

$$j_5 = \frac{2c(2\gamma-s)}{b_2}$$

$$j_6 = \frac{\left[\begin{array}{l} 16(1-\gamma)-8s+4(1+2c)\,\gamma s \\ \quad +(1-2c)(32\,\bar{D}(\gamma(2-\gamma)-s)+b_1+4\gamma^2) \end{array}\right]}{b_2^2}$$

$$j_7 = \frac{\left[\begin{array}{l} 16(1-\gamma(1-2c^2 s))-8s+(1-2\,c) \\ \quad (32\,\bar{D}(\gamma(2-\gamma)-s)+b_1+b_3) \end{array}\right]}{b_2^2}.$$

$\bar{D} = \bar{D}_{AB}$ has to be replaced by the respective expression given in Eq. (A.1). Furthermore, $b_1 = 16\,\bar{D}^2(2\gamma-s)^2 + s^2 + 8\,\bar{D}s^2$, $b_2 = 4 - 2\gamma - s + 4\bar{D}(2\gamma - s)$ and $b_3 = 4\gamma(1+2c)(\gamma+s) + 2c(s^2 - 16\bar{D}^2(2\gamma-s)^2)$.

ACKNOWLEDGMENTS

This research was supported by a research grant from the German Academic Exchange Service (DAAD) to T. W. and by a grant from the U. S. NIH (GM40282) to M. S. We thank Sunetra Gupta and Yannis Michalakis for valuable comments on an earlier draft of the manuscript.

REFERENCES

Feldman, M. W., Franklin, I., and Thomson, G. J. 1974. Selection in complex genetic systems. I. The symmetric equilibria of the three–locus symmetric viability model, *Genetics* **76**, 135–162.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**, 124–138.

Frank, S. A. 1993. Specificity versus detectable polymorphism in host–parasite genetics, *Proc. R. Soc. London B* **254**, 191–197.

Gillespie, J., and Langley, C. 1976. Multilocus behavior in random environments. I. Random Levene models, *Genetics* **82**, 123–137.

Gliddon, C., and Strobeck, C. 1975. Necessary and sufficient conditions for multiple–niche polymorphism in haploids, *Am. Nat.* **109**, 233–235.

Gupta, S., Maiden, M. C. J., Feavers, I. M., Nee, S., May, R. M., and Anderson, R. M. 1996. The maintenance of strain structure in populations of recombining infectious agents, *Nat. Med.* **2**(4), 437–442.

Karlin, S., and Feldman, M. W. 1970. Linkage and selection: Two locus symmetric viability model, *Theor. Pop. Biol.* **1**, 39–71.

Kimura, M., and Maruyama, T. 1966. The mutational load with epistatic gene interactions in fitness, *Genetics* **54**, 1337–1351.

Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available, *Am. Nat.* **87**, 331–333.

Lewontin, R. C., and Kojima, K. 1960. The evolutionary dynamics of complex polymorphisms, *Evolution* **14**, 458–472.

Maynard Smith, J., and Hoekstra, R. 1980. Polymorphism in a varied environment: How robust are the models? *Gen. Res. Cambridge* **35**, 45–57.

Wade, M. J. 1985. Soft selection, hard selection, kin selection and group selection, *Am. Nat.* **125**, 61–73.

Wallace, B, 1968. Polymorphism, population size and genetic load, *in* "Population Biology and Evolution" (R. C. Lewontin, Ed.), pp. 87–108, Syracuse Univ. Press, Syracuse, NY.

Printed in Belgium