GENE FLOW AND SELECTION IN A TWO-LOCUS SYSTEM[1]

MONTGOMERY SLATKIN

Department of Biophysics and Theoretical Biology, The University of Chicago, 920 East 58th Street, Chicago, Illinois 60637

Manuscript received April 29, 1974
Revised copy received June 4, 1975

ABSTRACT

A model of gene flow and selection in two linked loci is analyzed. The problems considered are the effects of linkage on the clines in frequencies at the two loci and the role of gene flow in producing linkage disequilibrium between the loci. Also, the possible significance of linkage as a mechanism for permitting a population of "track" spatial changes in the environment is considered. The results are that when the recombination fraction between the loci is of the same order of magnitude as the selection coefficients or smaller, then linkage is important in determining the gene frequencies and a substantial amount of linkage disequilibrium is present in the cline. Depending on the spatial pattern of selection on the two loci, linkage can either decrease or increase a population's response to local selection.

GENE flow between local populations can be an important mechanism in determining their genetic composition. Depending on the amount of gene flow, a population may be genetically distinct, a part of a cline, or a part of a panmictic population. Much work has been done on the interaction of gene flow with spatially varying selection on a single locus (see ENDLER 1973; and SLATKIN 1973). I will present here the results from a model of gene flow and selection on two linked loci.

There are two questions of interest which arise in this problem: (1) What is the effect of linkage on the gene frequencies in clines predicted on the basis of the one-locus theory? (2) How much linkage disequilibrium between the loci can be produced by gene flow? The first question relates to the effect that linkage has on a population's ability to respond to local environmental conditions. I will show that, depending on the patterns of environmental change, linkage can either increase or decrease the response to those patterns. The second question pertains to the interpretation of linkage disequilibrium in natural populations. LEWONTIN (1974) says that presence of linkage disequilibrium is a sensitive measure of additive epistasis between loci. However, I will show that a large amount of linkage disequilibrium can be generated by gene flow in a cline, even in the absence of epistasis. This same point has been made by LI and NEI (1974), by FELDMAN and CHRISTIANSON (1975) and by PROUT (1973) for other models of gene flow.

[1] This research has been supported by AEC contracts Nos. AT(11-1)-1437 and -2472.

THE MODEL

I will consider a genetic model with only two loci with two possible alleles at each locus (A_1, A_2 and B_1, B_2) and assume that there are a large number of colonies or demes each identified by a single spatial coordinate, z. Random mating and selection take place independently in each colony. After selection acts, individuals disperse before mating. The model is of species which have non-overlapping generations and which are subject to density-independent selection. Assume that there is no position effect (i.e., the fitness of the A_1B_1/A_2B_2 and A_1B_2/A_2B_1 genotypes are the same), there are nine fitnesses which must be specified at each location. They can be represented by a matrix $W_{ij}(z)$ ($i = 0,1,2$ and $j = 0,1,2$), which represents the fitness of a genotype with i A_1 alleles and j B_1 alleles. The recombination fraction between the two loci is r.

Since density effects are ignored, the state of the population in each colony can be represented by the frequencies of the four gametes. If we define $x_1(z,t)$, $x_2(z,t)$, $x_3(z,t)$ and $x_4(z,t)$ to be the frequencies of A_1B_1, A_1B_2, A_2B_1, and A_2B_2 respectively in colony z in generation t, then after mating and selection but before dispersal, the gametic frequencies in each colony are

$$
\begin{aligned}
x_1^*(z,t) &= \frac{x_1(z,t)\,\bar{W}_1(z,t)}{\bar{W}(z,t)} - \frac{rW_{11}(z,t)}{\bar{W}(z,t)}D(z,t) \\
x_2^*(z,t) &= \frac{x_2(z,t)\,\bar{W}_2(z,t)}{\bar{W}(z,t)} + \frac{rW_{11}(z,t)}{\bar{W}(z,t)}D(z,t) \\
x_3^*(z,t) &= \frac{x_3(z,t)\,\bar{W}_3(z,t)}{\bar{W}(z,t)} + \frac{rW_{11}(z,t)}{\bar{W}(z,t)}D(z,t) \\
x_4^*(z,t) &= \frac{x_4(z,t)\,\bar{W}_4(z,t)}{\bar{W}(z,t)} - \frac{rW_{11}(z,t)}{\bar{W}(z,t)}D(z,t)
\end{aligned} \tag{1}
$$

where

$$
\begin{aligned}
\bar{W}_1(z,t) &= x_1(z,t)W_{22}(z,t) + x_2(z,t)W_{21}(z,t) + x_3(z,t)W_{12}(z,t) + x_4(z,t)W_{11}(z,t) \\
\bar{W}_2(z,t) &= x_1(z,t)W_{21}(z,t) + x_2(z,t)W_{20}(z,t) + x_3(z,t)W_{11}(z,t) + x_4(z,t)W_{10}(z,t) \\
\bar{W}_3(z,t) &= x_1(z,t)W_{12}(z,t) + x_2(z,t)W_{11}(z,t) + x_3(z,t)W_{02}(z,t) + x_4(z,t)W_{01}(z,t) \\
\bar{W}_4(z,t) &= x_1(z,t)W_{11}(z,t) + x_2(z,t)W_{10}(z,t) + x_3(z,t)W_{01}(z,t) + x_4(z,t)W_{00}(z,t) \\
\bar{W}(z,t) &= x_1(z,t)\bar{W}_1(z,t) + x_2(z,t)\bar{W}_2(z,t) + x_3(z,t)\bar{W}_3(z,t) + x_4(z,t)\bar{W}_4(z,t)
\end{aligned} \tag{2}
$$

and

$$D(z,t) = x_1(z,t)x_4(z,t) - x_2(z,t)x_3(z,t) \ .$$

(See LEWONTIN and KOJIMA 1960 for a derivation of these equations.)

We consider only the simplest kind of dispersal of individuals and assume that there is no mortality during dispersal and no habitat selection. In that case dispersal affects each of the gametes independently, regardless of whether the dispersal occurs during the adult stage or the gamete state of the life cycle. I will

assume that dispersal occurs after selection. A slightly different model would be required for a situation in which dispersal occurs at the seed or larval stage, but there are no significant differences between the two approaches unless very strong selection is considered. Dispersal is modeled by a function $M(z,z')$, which is the probability that an individual will leave colony z' to settle and mate in colony z. If the colonies are discrete, then M is a matrix (commonly called the migration matrix; cf. MALÉCOT 1948; or KIMURA and WEISS 1964) and if there is assumed to be a continuum of colonies then M is a function of two variables and is called the migration function (MALÉCOT 1948; or SLATKIN 1973).

To compute the x's in the next generation in one colony, the contributions from all other colonies must be summed (or integrated, depending on the model). Therefore,

$$x_i(z,t+1) = \int M(z,z')x_i^*(z',t)dz' \quad i=1,2,3,4 \tag{3}$$

and we have a complete specification of the model in equations (1), (2) and (3). In this paper, I will consider only the equilibrium solutions to this system so $x_i(z,t+1) = x_i(z,t)$ and assume also that the equilibrium solutions are stable. This assumption is supported by the numerical iteration of the basic equations.

By sufficiently restricting the model we can gain some insight about the solutions. For mathematical convenience, I will consider the continuum model on a linear, infinite region which is homogeneous with respect to dispersal. The assumption that the region is infinite in extent is not restrictive. The effects of finite boundaries are not important as long as the boundaries are far from the regions of significant changes in the gene frequencies. Genetic drift is assumed to be unimportant so the properties of the model do not depend on the total size of the region. Since dispersal is assumed to be homogeneous and symmetric, $M(z,z')$ must be a function of $|z-z'|$ only.

EXAMPLES

Initially, we write the relative fitness, W_{ij}, in the form shown in Table 1. The parameters s and t are measures of the selection acting at each of the loci; e is a measure of the deviation from additive selection between the loci; and $\gamma(z)$ is a function which describes spatial variation in the relative fitness values of the alleles. In this model, the geographic changes are assumed to occur at the same location for both of the loci. The function $\gamma(z)$ is defined so that it has a maximum value of 1; thus, the spatial variation in this selection is separated from the

TABLE 1

Selection matrix when environment changes at the same location for both loci

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	$1+(s+t)\gamma(z)+e$	$1+t\gamma(z)$	$1+(t-s)\gamma(z)-e$
B_1B_2	$1+s\gamma(z)$	1	$1-s\gamma(z)$
B_2B_2	$1+(s-t)\gamma(z)-e$	$1-t\gamma(z)$	$1-(t+s)\gamma(z)+e$

strength of selection, measured by s and t, and the epistasis measured by e. The fitnesses are written in this way to facilitate comparisons with the one-locus model of gene flow and selection and with the two-locus models without gene flow. If the fitnesses are additive between the loci, then $e = 0$; if they are multiplicative, then $e = st$.

We will consider a model of a step change in environmental conditions separating two regions: in one, A_1 and B_1 are favored and in the other, A_2 and B_2 are favored. Thus $\gamma(z) = +1$ for $z>0$ and -1 for $z<0$. We expect to find a cline in frequencies at both of the loci and, if the loci were independent, one-locus theory could be used to predict the shape of the cline. We consider here the effect of epistasis and linkage on the clines predicted using one-locus theory.

First, we analyze the case with $s = t$. From the symmetry of the problem, it is reasonable to assume there is a solution which is unchanged if A_1 is replaced by B_1, A_2 by B_2 and z by $-z$. For this solution $x_2(z) = x_3(z) = x_2(-z)$ and $x_1(z) = x_4(-z)$. T. NAGYLAKI (personal communication) has shown that for this model, such a solution always exists. While there may be other equilibrium solutions not having these properties, in the numerical calculations there was no evidence that there are other stable equilibrium solutions for which all alleles are present.

The basic equations (1)–(3) can be written

$$\begin{aligned}
x_1(z) &= \int_{-\infty}^{\infty} \frac{M(z-z')}{\bar{W}(z')} \{x_1(z')[1+x_1(z')(2s\gamma(z')+e) \\
&\quad + 2sx_2(z')\gamma(z')] - rD(z')\} \quad dz' \\
x_2(z) &= x_3(z) = \int_{-\infty}^{\infty} \frac{M(z-z')}{\bar{W}(z')} \{x_2(z')[1+s\gamma(z')(x_1(z')-x_4(z')) \\
&\quad - ex_2(z')] + rD(z')\} \quad dz' \\
x_4(z) &= \int_{-\infty}^{\infty} \frac{M(z-z')}{\bar{W}(z')} \{x_4(z')\ [1-2s\gamma\ (z')\ x_2(z') \\
&\quad - x_4(z')\ (2s\gamma(z')-e)] - rD(z')\} \quad dz'
\end{aligned} \tag{4}$$

in this special case. We notice that the combination $x_1(z) - x_4(z)$ appears frequently enough that it may be a more useful variable than one of the original ones. Since there are now only two independent variables, we write

$$B(z) = x_1(z) - x_4(z) \tag{5}$$

and find the equations for $B(z)$ and $x_2(z)$. We note

$$\begin{aligned}
D(z) &= \frac{1-B^2(z)}{4} - x_2(z) \\
x_1(z) &= \frac{1+B(z)}{2} - x_2(z) \\
x_4(z) &= \frac{1-B(z)}{2} - x_2(z)
\end{aligned} \tag{6}$$

to find that

$$B(z) = \int_{-\infty}^{\infty} M(z-z') \{B(z') + \frac{s\gamma(z')}{\bar{W}(z')} [1-B^2(z') - 2x_2(z')] + \frac{e}{\bar{W}(z')} B(z')(1-B^2(z'))/2\} \quad (7)$$

and

$$x_2(z) = \int_{-\infty}^{\infty} M(z-z') \{x_2(z') - \frac{s\gamma(z')B(z')x_2(z')}{\bar{W}(z')} - \frac{ex_2(z')}{2\bar{W}(z')} (1+B^2(z') - 2x_2(z')) + \frac{r}{4\bar{W}(z')} (1-B^2(z') - 4x_2(z'))\} \, dz' \quad (8)$$

where

$$\bar{W}(z) = 1+2s\gamma(z)B(z) + \frac{e}{2} (1+B^2(z) - 4x_2(z)) \ .$$

Equations (7) and (8), in general, cannot be solved analytically, but there are some conclusions which we can reach by comparing them with previously solved cases. Those conclusions will be verified by numerical solutions.

The epistasis appears in each equation as a term multiplied by e. Therefore it seems that there is no complex effect of epistasis in this problem as there is in other models (the two-locus, heterotic model, for example). When the fitnesses are approximately multiplicative (e of the same order of magnitude as s^2), then for weak selection, $e \ll s$ and the solutions to (7) and (8) are almost independent of e. To be more precise, the solutions to (7) and (8) could be expressed as perturbation series in e and to the lowest order, the gene frequencies in the cline would be independent of e. This is in contrast to the two-locus heterotic models in which the result depends strongly on the relative values of e and r. Of course, when e is on the same order of magnitude as s, or when s is nearly 1, the epistasis may be important and this argument cannot be used.

If we let $e=0$ (considering only the lowest order terms and ignoring epistasis) and reduce (7) and (8) to differential equations by using the same techniques as in my previous paper (SLATKIN 1973), we get

$$\frac{-l^2}{2} \frac{d^2B(z)}{dz^2} = s\gamma(z) \ (1-B^2(z) - 2x_2(x)) \quad (9)$$

$$\frac{-l^2}{2} \frac{d^2x_2(z)}{dz^2} = -s\gamma(z)B(z)x_2(z) + \frac{r}{4} (1-B^2(z) - 4x_2(z)) \quad (10)$$

where

$$l^2 = \int_{-\infty}^{\infty} M(z) z^2 dz$$

Equations (9) and (10) are correct to lowest order in s. For the symmetric solution, it is sufficient to consider only the region $0<z<\infty$. The parameter

l is approximately the dispersal distance. The boundary conditions are that $B(0)=x'_2(0)=0$ and $B(z)\to 1$ and $x_2(z)\to 0$ as $x\to\infty$. The reduction of the integral to the differential equations is equivalent to assuming that gene flow can be modeled as a diffusion process and is valid when the higher moments of $M(z-z')$ are not too large. That assumption is satisfied for most migration functions which have been measured.

We note that

$$D(z) = \frac{1-B^2(z)}{4} - x_2(z) \quad , \tag{11}$$

the linkage disequilibrium at z. Therefore (9) and (10) can be rewritten

$$\frac{-l^2}{2}\frac{d^2B(z)}{dz^2} = s\left(\frac{1-B^2(z)}{2} + 2D(z)\right) \tag{12}$$

$$\frac{-l^2}{2}\frac{d^2x_2(z)}{dz^2} = -sB(z)x_2(z) + rD(z) \quad . \tag{13}$$

We consider first the case where $r>>s$. As in the case of a cline in allele frequencies at one locus, the natural length scale in equation (12) is $l/\sqrt{s}$ (see SLATKIN 1973). Since the gametic frequencies sum to one, the gametic frequencies must all change on the same length scale. The natural length scale for equation (13) is not $l/\sqrt{s}$ unless D is of the same order of magnitude as s/r. Thus when $r>>s$ we would expect D to be small. Conversely, when r and s are of the same order of magnitude we would expect that D could be significantly larger. These predictions based on a qualitative analysis of the differential equations are confirmed by the numerical iteration of the exact equation as discussed later.

When $D(z)$ is small, it can be ignored in (12), which then has the solution

$$B(z) = -2+3\ \tanh^2\left(\frac{\sqrt{s/2}\,x}{l} + \tanh^{-1}\sqrt{\frac{2}{3}}\right) \quad , \tag{14}$$

and we find that the allele frequencies are the same as in the unlinked case (SLATKIN 1973). Equation (14) differs from equation (11) in SLATKIN (1973) by a factor of $\sqrt{2}$, as a result of an algebraic error contained in that paper and kindly pointed out by T. NAGYLAKI (personal communication). This is the expected result and the calculation merely shows the range of r for which linkage would be expected to be unimportant. As will be verified by the numerical results, when r is of the same order of magnitude as s or smaller, there would be a significant amount of linkage disequilibrium present in the cline. One consequence is that if s is reasonably large (.1 or .2), then some linkage disequilibrium is present for unlinked loci (r=.5) as a result only of the gene flow.

In the opposite extreme, for small values of r, the model is nearly the same as a model of a one-locus system with four alleles. In this case two of the "alleles" (A_1B_2 and A_2B_1) are at a disadvantage throughout the range and therefore should not be present at equilibrium. For r=0, we can easily show that the A_1B_2 and A_2B_1 gametes would be absent by noting that when r=0, the only possible solution of (13) with $x'_2(0)=0$ and $x_2(\infty)=0$ is $x_2(z)=0$, since $B(z)>0$ for $z>0$ in

the symmetric case. Thus, this system is equivalent to a two- allele system in which the selective different is $2s$. The maximum slope of the cline in frequencies at the two loci is $\sqrt{8s/3l^2}$, which is larger by a factor of $\sqrt{2}$ than in the case with large r. For intermediate values of r, we would expect the cline to be somewhat steeper than for large r, but the maximum possible increase in the slope is $\sqrt{2}$.

In my previous paper, I used the inverse of the maximum slope of the cline as a measure of the "characteristic length" of the system (SLATKIN 1973). I showed that the gene frequencies at equilibrium do not respond to spatial changes in the environment that occur on a scale less than the characteristic length. The above analysis indicates that linkage decreases the characteristic length whenever r is of the same order of magnitude as s and that the maximum decrease is $1/\sqrt{2}$ or about 30%. Therefore, tighter linkage between two loci will allow the gene frequencies at both loci to respond to smaller scale changes in the environment than they would in the absence of linkage, as long as the environmental changes occur in the same location at both loci. This suggests the possibility, then, of selection in favor of a modifier allele for tighter linkage in those cases when environmental changes are on a scale smaller than $l/\sqrt{s}$, the characteristic length in the unlinked case. This is consistent with KARLIN and MCGREGOR'S (1974) general theory of modifier alleles.

The predictions about the behavior of the exact model are based on a qualitative analysis of the approximate differential equations. I tested these predictions by directly iterating the discrete analog of equations (1)–(3) on a computer. Equation (3) was replaced by

$$x(i,t+1) = \sum_{j=1}^{n} x^*(j,t)M(i,j) \ .$$

where $M(i,j)$ is the migration matrix and i and j identify one of the n locations in the cline. In practice n was always even, so there was no colony at the point corresponding to $z=0$; the region $z<0$ was replaced by $i<n/2$ and $z>0$ by $i>n/2+1$.

The parameter n was large enough that the allele frequencies were nearly 1 or 0 at the boundaries. I always chose the boundaries to be approximately ten times the characteristic length and assumed a reflecting boundary condition. Further increases in n did not have any significant effect on the equilibrium solution. The model was iterated until an equilibrium was reached. The results in the first case are shown in Figure 1; the linkage disequilibrium, D, is plotted for different values of r. It could be anticipated from the analysis of the differential equations that the size of the region for which D could be significantly greater than 0 is of the same order of magnitude as the characteristic length. The maximum slope of the cline ($\Delta=x(n/2+1)-x(n/2)$) is given in Table 2 for different values of r. Also given in Table 2 are the results for different amounts of epistasis. These calculations verify the prediction about the relationship between r and s and about the importance of e.

Using the above results, we can speculate about the properties of a system with more linked loci. For simplicity, we can consider the case in which there are k loci of equal effect and assume that the selection coefficients between the loci

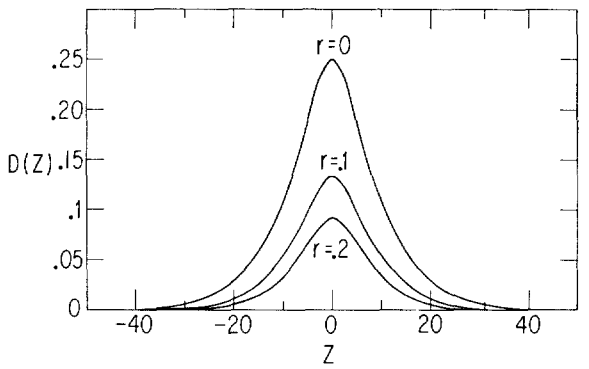


FIGURE 1.—Linkage disequilibrium *vs.* distance for fitness as in Table 1, $s=t=.1$ and $l=3$.

are additive. In the case of complete linkage ($r=0$), by analogy with the two-locus case, the system again reduces to a two-allele model with the selective differences between the two alleles of ks (ignoring epistasis). Therefore, the maximum slope of the cline would be $\sqrt{4ks/3l^2}$ and the associated characteristic length $l\sqrt{3/4ks}$. In the two-locus case, the linkage did not have a great effect on the characteristic length but with more loci, the characteristic length decreases by a factor of $1/\sqrt{k}$. Therefore, there is the possibility, at least, that the characteristic length may be reduced by an order of magnitude. That result would not depend strongly on either the additivity of the selective values or the equivalence of the loci assumed here.

Returning to the two-locus model, when the two loci contribute unequally to the fitnesses, the analytic problem becomes more complex and less revealing. I will limit myself to a presentation and discussion of the numerical results. The selection model is the same as in the previous cases (Table 1), but now s is significantly greater than t. When $r=0$, numerical study indicates that the results from the one-locus model still apply and we can predict that allele frequencies at the two loci will be the same and can be predicted from a one-locus model with selection coefficients, $s+t+e$. Since we have assumed $s>>t$ and e to be of the order of st, the frequencies at the A locus (the more strongly select locus)

TABLE 2

Maximum slope of cline for selection in Table 1 with s=t=.1 *and* $l=3$

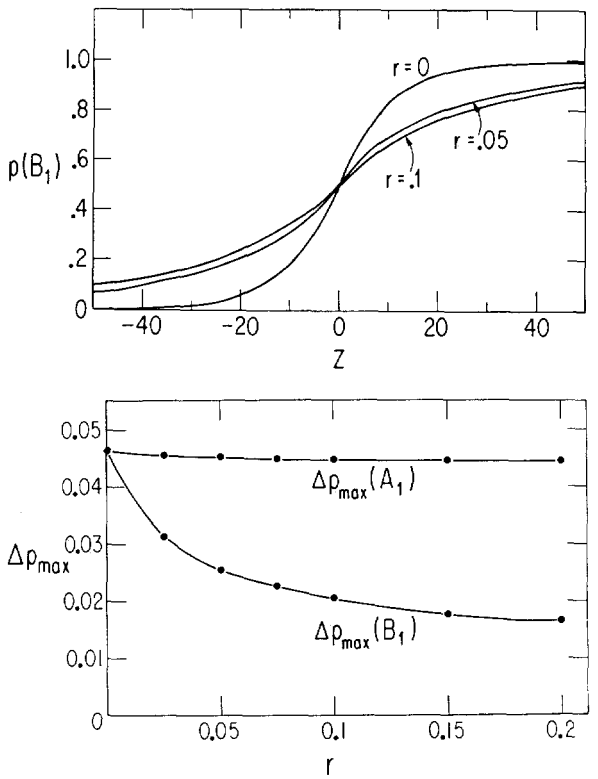
		Δp_{max}		
r	$e=$	0.0	0.01	0.02
0.00		.0624	.0630	.0634
0.05			.0578	
0.10		.0538	.0548	.0558
0.15			.0532	
0.20		.0508	.0520	.0528
0.25			.0511	
0.5		.0478	.0489	.0498

will be approximately the same as in the case without linkage. At the B locus, however, the cline will be greatly steepened because of linkage. The maximum slope of the cline at the B locus will be increased from roughly $\sqrt{4t/3l^2}$ to $\sqrt{4s/3l^2}$. For example, if s=.1 and t=.01, the slope increases from .12/l to .37/l.

The results from the computer iteration of equations (1)–(3) are shown in Figures 2 and 3. In Figure 2, $p(B_1)$ $(=x_1(z)+x_3(z))$ the frequency of the B_1 allele is plotted against z. When r=0, $p(A_1)=p(B_1)$. The maximum values of D ($D(i)$ at $i=n/2$ or $n/2+1$) are given in Table 3. That is another measure of the importance of linkage.

The maximum slope of the clines in the A and B loci are shown in Figure 3 as functions of r. We can distinguish three regions. For $r<t$, the loci are acting effectively as a single unit. For $t<r<s$, linkage is affecting the allele frequencies but the loci are more or less independent. For $r>s$, linkage has little effect, and the loci are almost independent. Again, we find that the loci do not have to be very closely linked for linkage to be important in determining allele frequencies.

The first two cases analyzed using the selection model in Table 1 are those in which the change in environmental conditions is assumed to occur at the same



FIGURES 2 AND 3.—Allele frequencies and maximum slopes of clines produced by selection of the type given by Table 1 with s=1 and t=.01 and l=3.

TABLE 3

Maximum linkage disequilibrium for selection in Table 1 with s=.1 t=.01

r	$D(n/2)$
0.0	0.25
0.025	0.1358
0.05	0.0937
0.075	0.0713
0.1	0.0579
0.15	0.0409
0.2	0.0325

location for both loci. That model is motivated by the problem of the evolution of coadapted gene complexes in response to different patterns of environmental variations. A similar problem is one in which the changes in selection occur at different locations for the two loci. The selection model for this case is given in Table 4. I shall assume that $\gamma_1(z) = \theta(z+z_0)$ and $\gamma_2(z) = \theta(z-z_0)$ where $\theta(z) = -1$ for $z<0$ and $+1$ for $z>0$. The geographic pattern is illustrated in Figure 4.

In this case, when z_0 is large, the clines in frequencies of the two loci are independent of each other and are determined by the selection coefficients at the two loci; linkage has no effect. Only when there is some overlap of the clines can linkage be important. The clines will overlap when $2z_0$ is approximately the same as or less than the larger of the two characteristic lengths associated with the two loci. In such cases, the two effects of linkage that are of interest are the

TABLE 4

Selection matrix when environmental change is different at the two loci

	A_1A_1	A_1A_2	A_2A_2
B_1B_1	$1+s\gamma_1(z)+t\gamma_2(z)+e\gamma_1(z)\gamma_2(z)$	$1+t\gamma_2(z)$	$1+t\gamma_2(z)-s\gamma_1(z)-e\gamma_1(z)\gamma_2(z)$
B_1B_2	$1+s\gamma_1(z)$	1	$1-s\gamma_1(z)$
B_2B_2	$1+s\gamma_1(z)-t\gamma_2(z)-e\gamma_1(z)\gamma_2(z)$	$1-t\gamma_1(z)$	$1-s\gamma_1(z)-t\gamma_2(z)+e\gamma_1(z)\gamma_2(z)$

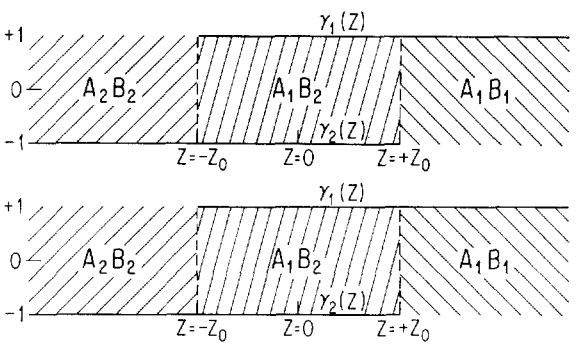
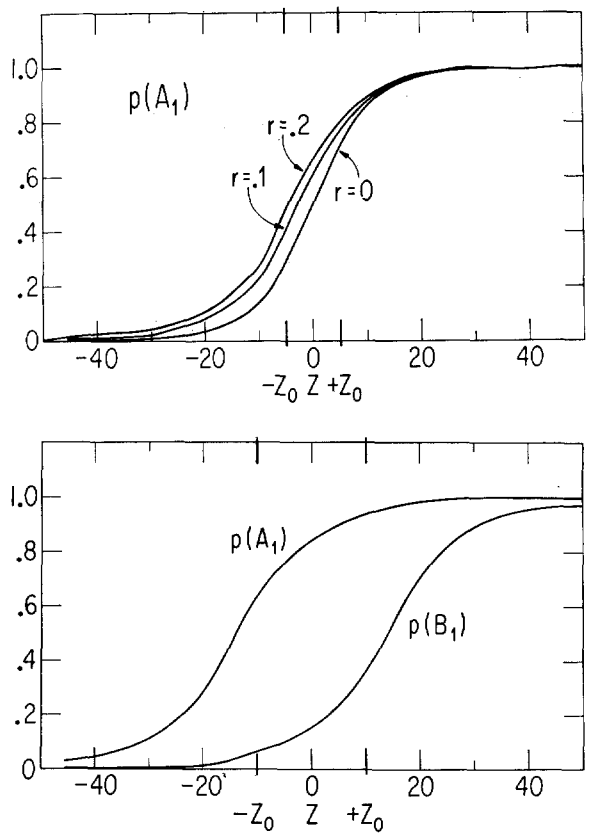


FIGURE 4.—Illustration of the pattern of selection in Table 4.

shifting of the two clines from z_0 and $-z_0$ toward 0 and the steepening of the clines as occurs in the cases analyzed previously. By referring to equation (12), we can see that linkage disequilibrium modifies the effective selection at each locus. When the selection intensities at the two loci change at different places, one result is that the effective selection does not reflect the actual spatial patterns.

We consider first the case with $s=t$. The results from the numerical iteration of the basic equations (1)–(3) are shown in Figure 5 for $z_0=5$ and three values of r. In this case $l_c \simeq l/\sqrt{s} = 9.48$ and we would expect, from the above argument, that linkage would have an effect on the cline. We see that the cline does shift from $-z_0$ toward 0 as r decreases from .2 to 0. The cline is somewhat steeper for smaller values of r but that is not apparent from the figure. From the previous cases analyzed, we would not expect much steepening, even for very small z_0. Figure 5 can be contrasted with the example shown in Figure 6. The only difference is that $z_0=10$ and, in that case, even with $r=0$, the cline is not shifted perceptibly.

There are three parameters of interest in this problem, z_0, s, and r. The main effect of linkage is the shifting of the cline from the location expected if there



FIGURES 5 AND 6.—Gene frequencies in clines produced by selection of the kind in Table 4 with $s=t=.1$ and $l=3$. In Figure 5, $z_0=10$.

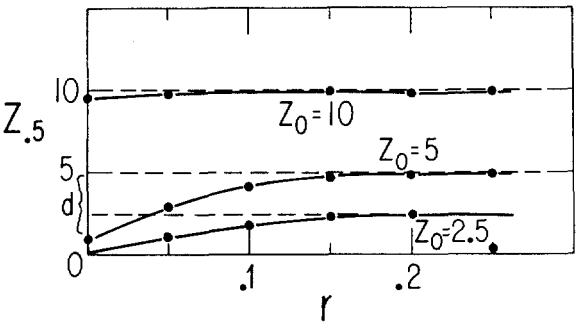


FIGURE 7.—Location of midpoint of the cline plotted against r for different values of z selection as in Table 4 with $s=t=.1$ and $l=3$.

were no linkage, $\pm z_0$, to the actual location with linkage. For these purposes, we can define the location of the clines as the point at which $p=\frac{1}{2}$. We can best understand the behavior of the system by plotting $z_{.5}$ (the point at which $p(B_1)=\frac{1}{2}$) against r for different values of z_0, as *is* shown in Figure 7. A measure of the importance of linkage is $d=z_0-z_{.5}$. We can see that for a fixed value of r, d is a maximum when z_0 is roughly $l_c/2$. For smaller values of z_0, linkage disequilibrium is larger but the distance that the cline can be shifted is smaller. For larger values of z_0, the D is smaller and the cline is not shifted as much.

The final case considered is the same as above but with $s>>t$. As before in the case of unequal selection coefficients, there is little change in the more strongly selected locus. We would expect that the cline in the more weakly selected locus be both shifted and made more steep as a result of the linkage. Two cases are shown in Figure 8. In part a, $z_0=20$ in which $l_c=8.2$ and 26 for the A and B loci respectively, so there is little overlap in the clines and linkage has no effect, even when $r=0$. In part b, $z_0=10$, which is intermediate between the two characteristic lengths. Linkage shifts the cline in the B locus but does not steepen it significantly. For smaller values of z_0 ($z_0<5$), the results are almost the same as in the case with $z_0=0$, which was treated above.

We can again plot the maximum effect of linkage on the clines by $z_{.5}$ again s for different values of z_0. In Figure 9, we see that the maximum effect of linkage occurs when z_0 is roughly one-half of the larger of the two characteristic lengths associated with the two loci. In other words, the shift in the cline of the more weakly selected locus depends mainly on its selection coefficient.

DISCUSSION

We have found that, with gene flow and selection on linked loci, when the recombination fraction, r, is of the same order of magnitude as the selection coefficients for the loci, then linkage will be important in determining the gene frequencies. In such cases, there will be a substantial linkage disequilibrium between the two loci in the interior part of the cline.

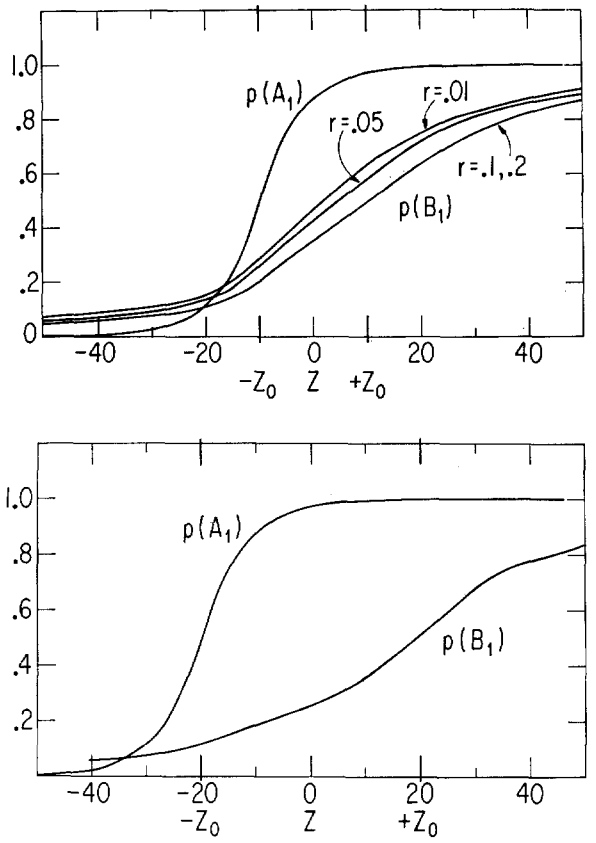


FIGURE 8.—Gene frequencies when selection is determined by Table 4 with s=.1, t=.01 and l=3.

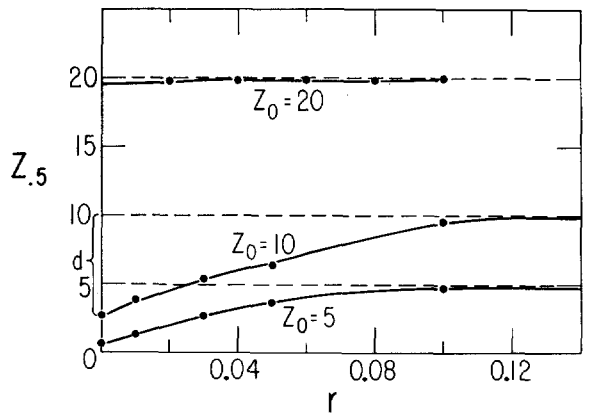


FIGURE 9.—Location of midpoint of clines in Figure 8. Similar to Figure 7, but with s=.1, t=.01 and l=3.

Considering first the effect on the clines in gene frequencies, linkage allows a population to track the changes in the environment more closely than in the unlinked case when the selection intensities change at the same point for both of the loci. This effect is already apparent in the two-locus case and can be even more important with more than two linked loci. This suggests the possibility of selection for a modifier which reduces crossing over. In contrast, when the selection intensities change at different places for different loci, linkage between the loci reduces the ability of the population to track changes in selection. In that case the locations of the clines would not necessarily correspond to the environmental changes. It is possible, then, that there could be selection in favor of a modifier allele which increases the recombination fraction between the loci.

With other models of selection besides the ones used above, particularly those which allow for one locus to modify the dominance relationship at the other locus, very steep clines can result even in a two-locus system. CLARKE (1966) first investigated this problem and recently ENDLER (manuscript in preparation) extended CLARKE's model and found evidence for that kind of selection acting in laboratory populations of *Drosophila melanogaster*.

The shifting of a cline in frequencies at a locus away from the location expected on the basis of its selection pattern could be important in some populations. BISHOP (1972) reports on a very detailed theoretical and experimental analysis of a cline in frequencies of melanic forms of the moth *Biston betularia*. BISHOP found that the observed cline was from 10 to 20 km from the location predicted with a computer model using estimated values for the relevant parameters. Even when factors such as heterosis and nonsymmetric dispersal were included in the model, the center of the predicted cline was at least 8 km from the observed cline. However, the shapes of the predicted and observed clines were similar.

The linkage of the melanic gene to a strongly selected gene or genetic complex is a possible explanation for BISHOP's (1972) findings. BISHOP demonstrated that the selection for or against the melanic form was quite strong, on the order of 30% or more against the less favored form (BISHOP 1972, Table 12). Although he could not estimate the average migration distance per generation, he did find that at least 25% of the males traveled more than 2 km per generation. Thus 2 km is a reasonable estimate for l. Figure 7 shows the displacement of the center of a cline with $s=t=.1$ and $l=3$. We can see that with tight linkage the displacement of the cline can be greater than the average dispersal distance. It is possible that with heterosis and asymmetric dispersal, the expected displacement due to linkage could be even larger, although a more detailed model would be necessary to make any quantitative predictions. The point here is that displacements of the magnitude found by BISHOP (1972) could be caused by linkage and that linkage would not necessarily change the shape of the observed cline from that expected in the absence of linkage.

In a panmictic population with a symmetric selection model assumed, the recombination fraction, r must be less than the deviation from additive fitnesses between the loci for linkage disequilibrium to be present at equilibrium

(LEWONTIN and KOJIMA 1960). If the fitnesses are roughly multiplicative between the loci, then r must be less than the square of the selection coefficient. However, with gene flow, r can be equal to or even greater than the selection coefficients and there still can be a substantial amount of linkage disequilibrium present. Similar conclusions regarding the importance of gene flow in producing linkage disequilibrium have been reached by NEI and LI (1973), by PROUT (1973), by LI and NEI (1974) and by FELDMAN and CHRISTIANSEN (1975).

That raises an interesting problem regarding the observations on linkage disequilibrium in natural populations. In many studies, significant linkage disequilibrium has not been found, even between very closely linked loci, except when the loci are on an inversion (CHARLESWORTH and CHARLESWORTH 1973). There are some exceptions (e.g., ROBERTS and BAKER 1973) and the data are hardly complete at this time. Also, there are technical problems in estimating small values of D. Still, if spatial variation in selective pressures is important and the loci which are studied by the current electrophoretic techniques are affected by the many spatial patterns we observe in nature, then one is forced to ask why more linkage disequilibrium is not found. The problem is compounded by FELSENSTEIN's calculations (1965) which imply that in many cases, a substantial amount of linkage disequilibrium can be generated by fluctuating selection in a panmictic population, even though there would be no linkage disequilibrium if equilibrium is reached. Although the calculations have not been performed, it is certain that a combination of spatial and temporal variation in selection will produce some linkage disequilibrium under almost any reasonable set of assumptions. If studies on natural populations continue to indicate that linkage disequilibrium between closely linked loci is often not large, then we will be forced to conclude that either most selection coefficients are constant in space and time or that a majority of the loci observed in electrophoretic studies are, in fact, selectively neutral, as has been proposed by KIMURA (1968) on the basis of substitution rates of alleles.

I wish to thank JOHN ENDLER and JOSEPH FELSENSTEIN for valuable discussions of this subject, and J. F. CROW, T. NAGYLAKI, and the referees for helpful comments on an earlier draft of this paper.

LITERATURE CITED

BISHOP, J. A., 1972 An experimental study of the cline of industrial melanism in *Biston betularia* (L.) between urban Liverpool and rural North Wales. J. Animal Ecol. **41**: 209–243.

CHARLESWORTH, B. and D. CHARLESWORTH, 1973 A study of linkage disequilibrium of *Drosophila melanogaster*. Genetics **73**: 351–359.

CLARKE, B. C., 1966 The evolution of morpho-ratio clines. Am. Naturalist **100**: 389–402.

ENDLER, J. A., 1973 Gene flow and population differentiation. Science **179**: 243–250.

FELDMAN, M. W. and F. B. CHRISTIANSEN, 1975 The effect of population subdivision on two loci without selection. Genet. Res. **24**: 151–162.

FELSENSTEIN, J., 1965 The effect of linkage on directional selection. Genetics **52**:349–363.

KARLIN, S. and J. MCGREGOR, 1974 Towards a general theory of the evolution of modifier genes. Theoret. Pop. Biol. **5**: 59–103.

KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624–626.

KIMURA, M. and G. H. WEISS, 1964 The stepping stone model of population structure and the decrease in genetic correlation with distance. Genetics **49:** 561–576.

LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change.* Columbia University Press, New York.

LEWONTIN, R. C. and K. KOJIMA, 1960 The evolutionary dynamics of complex polymorphisms. Evolution **14:** 458–472.

LI, W. H. and M. NEI, 1974 Stable linkage disequilibrium without epistasis in subdivided populations. Theoret. Pop. Biol. **6:** 173–183.

MALÉCOT, G., 1948 *Les mathematique de l'heredite.* Maisson et Cie, Paris.

NEI, M. and W. S. LI, 1973 Linkage disequilibrium in subdivided populations. Genetics **75:** 213–219.

PROUT, T., 1973 Appendix to "Population genetics of marine Pelecypods. III. Epistatis between functionally related isoenzymes of *Mytilus edulis*" by J. B. MITTON and R. K. KOEHN. Genetics **73:** 493–496.

ROBERTS, R. M. and W. K. BAKER, 1973 Frequency distribution and linkage disequilibrium of active and null estrase isozymes in natural populations of *Drosophila montana.* Am. Naturalist **107:** 709–726.

SLATKIN, M., 1973 Gene flow and selection in a cline. Genetics **75:** 733–756.

Corresponding editor: J. FELSENSTEIN