MULTILOCUS BEHAVIOR IN RANDOM ENVIRONMENTS.
II. LINKAGE DISEQUILIBRIUM IN AN ADDITIVE MODEL

JOHN H. GILLESPIE

Department of Biology, G7, University of Pennsylvania,
Philadelphia, Pennsylvania, 19174

Manuscript received November 2, 1976
Revised copy received June 20, 1977

ABSTRACT

The effect of a stochastic environment on an additive, two-locus model of a diploid population is examined. The appropriate diffusion equation is derived and its asymptotic properties are approximated by an Orstein-Uhlenbeck process. The first and second order moments of this approximating process are given. The mean linkage disequilibrium will be nonzero if the alleles at the different loci are correlated. The sign of the mean disequilibrium is determined by the sign of the correlation.

In the previous paper in this series (GILLESPIE and LANGLEY 1976) we examined the consequences of multi-locus selection in a spatially subdivided population in which the fitnesses are assigned additively across loci and alleles. The asymptotic behavior of the model suggested that correlations in the fitnesses of alleles at different loci can be a potent factor in determining the genetic structure of the genome. The principle which emerged from this study was that two alleles at different loci whose fitnesses are negatively correlated across environments have a higher overall fitness due to the reduction in their variance in fitness resulting from the negative correlation. The deviations in behavior of the additive random environment model from the additive constant-fitness model can be completely accounted for by this inverse relationship between the overall fitness of a genotype and its variance in fitness.

In the present paper the Random Levene model will be generalized to allow a finite number of patches, with temporal fluctuations in the environment. Of particular interest will be the stochastic behavior of the linkage disequilibrium, D, which was shown previously to be asymptotically zero in the Random Levene model.

THE DIFFUSION MODEL

The format in this section follows very closely that given in GILLESPIE and LANGLEY (1976), to which the reader is referred for additional motivation.

Consider a diploid species occupying an environment that is subdivided into a number, m, of patches. Let there be two loci that are segregating and let the relative frequencies of the gametes $A_1B_1$, $A_1B_2$, $A_2B_1$, and $A_2B_2$ be $x_1$, $x_2$, $x_3$, and
In any one patch in any particular generation, the fitnesses of the zygotes will be determined by the random vectors

\[
\tilde{U}_{ij} = [U_{A_1(i,j)}, U_{A_2(i,j)}, U_{B_1(i,j)}, U_{B_2(i,j)}]
\]

where \(i\) refers to the \(i^{th}\) generation and \(j\) refers to the \(j^{th}\) patch, according to the following scheme:

\[
\begin{align*}
A_1A_1 & : 1 + U_{A_1} + U_{B_1} & 1 + U_{A_1} + U_{B_1} & 1 + U_{A_1} + U_{B_1} \\
B_1B_1 & : 1 + U_{A_1} + \frac{1}{2}(U_{B_1} + U_{B_2}) & 1 + U_{A_1} + \frac{1}{2}(U_{B_1} + U_{B_2}) & 1 + U_{A_1} + \frac{1}{2}(U_{B_1} + U_{B_2}) \\
B_2B_2 & : 1 + U_{A_1} + U_{B_2} & 1 + U_{A_1} + U_{B_2} & 1 + U_{A_1} + U_{B_2}
\end{align*}
\]

In this zygotic scheme, the \(i\) and \(j\) have been suppressed. Using the assumptions about the mating structure of the population given in Levene (1953) and Gillespie and Langley (1976), we can write

\[
\Delta x_k(i) = \sum_{j=1}^{M} C_j \Delta x_k(i,j) \quad k = 1, 2, 3, 4
\]

where \(C_j\) is the proportional contribution of the \(j^{th}\) patch to the random-mating pool, and

\[
2\overline{W} \Delta x_k(i,j) = x_k H_k(i,j) + l(k)2R W_{zz} D_k \quad k = 1, 2, 3, 4
\]

where \(l(k) = +1\) for \(k = 1, 4\), \(l(k) = -1\) for \(k = 2, 3\) and

\[
\begin{align*}
H_1(i,j) &= q_1(U_{A_1} - U_{A_2}) + q_2(U_{B_1} - U_{B_2}) \\
H_2(i,j) &= q_1(U_{A_1} - U_{A_2}) - p_2(U_{B_1} - U_{B_2}) \\
H_3(i,j) &= -p_1(U_{A_1} - U_{A_2}) + q_2(U_{B_1} - U_{B_2}) \\
H_4(i,j) &= -p_1(U_{A_1} - U_{A_2}) - p_2(U_{B_1} - U_{B_2}) \\
\overline{W} &= 1 + p_1 U_{A_1} + q_1 U_{A_2} + p_2 U_{B_1} + q_2 U_{B_2} \\
p_1 &= 1 - q_1 = x_1 + x_2 \\
p_2 &= 1 - q_2 = x_1 + x_3 \\
D &= x_1 x_3 - x_2 x_3
\end{align*}
\]

This completes the description of the model to be investigated. Analytic progress will come only from obtaining a limiting diffusion model under the assumption of weak selection and tight linkage. The limiting model will then be viewed as an approximation to the original model (3). Let \(\tau\) be a parameter that indexes a sequence of the vectors \(\tilde{U}_{ij}\) such that as \(\tau \to 0\), the magnitude of the elements of the vectors \(\tilde{U}_{ij}\) also approach zero. If \(\tau\) also measures the time...
between generations, we will obtain a limiting diffusion model if:

(a) The set \( \{U_{ij}\}, i = 0, 1, 2, \ldots; j = 1, 2, \ldots M, \) is made up of independent, identically distributed random vectors. In biological terms we are assuming that the environments in successive generations are independent as are those in different patches within a generation. By assuming that the vectors are identically distributed for a fixed \( i \) and \( j = 1, 2, \ldots, M \) we are asserting that the patches are “equivalent patches” in the terminology of Gillespie (1975).

(b) The first order moments of \( \tilde{U}_{ij} \) are

\[
E \tilde{U}_{ij} = [\mu_A \tau, \mu_A \tau, \mu_B \tau, \mu_B \tau]
\]

which obviously approach zero as \( \tau \to 0 \).

(c) The second order moments of \( \tilde{U}_{ij} \) are

\[
\text{VAR} \tilde{U}_{ij} = \begin{bmatrix}
\sigma_A^2 & \sigma_A & \sigma_A & \sigma_A \\
\sigma_A & \sigma_A^2 & \sigma_A & \sigma_A \\
\sigma_A & \sigma_A & \sigma_A^2 & \sigma_A \\
\sigma_A & \sigma_A & \sigma_A & \sigma_A
\end{bmatrix} \tau
\]

Thus the second order moments approach zero at the same rate as the first order moments.

(d) All higher order moments of \( \tilde{U}_{ij} \) are \( O(\tau) \).

(e) The recombination fraction approaches zero at the same rate as the first two moments of \( \tilde{U}_{ij} \). To make this explicit write

\[
R = r \tau .
\]

These five assumptions assure that as \( \tau \to 0 \), the limiting process will be a diffusion process. The drift coefficients for the limiting process may be obtained by using assumption (a) to write

\[
Edx_k = dt \lim_{\tau \to 0} {E[\Delta x_k \tau (i) | \tilde{x}^\tau (i)] \over \tau}
\]

\[
= dt \lim_{\tau \to 0} {\Sigma c_i E[\Delta x_k \tau (i,j) | \tilde{x}^\tau (i)] \over \tau}
\]

\[
= dt \lim_{\tau \to 0} {E[\Delta x_k \tau (i,j) | x^\tau (i)] \over \tau}
\]

where \( x^\tau (i) \) is the vector of gametic frequencies. The gametic frequencies have been superscripted by \( \tau \) to emphasize that corresponding to the sequence \( \tilde{U}_{ij} \) is the sequence of processes \( x^\tau (i) \) which will approach the diffusion limit. The
The conditional expectation in the last expression in (9) is given in Appendix I of Gillespie and Langley (1976) and may be written

\[ Edx_1 = \left[ \frac{1}{2} x_1 (q_1 \phi_A + q_2 \phi_B) - rD \right] dt \]
\[ Edx_2 = \left[ \frac{1}{2} x_2 (q_1 \phi_A - p_2 \phi_B) + rD \right] dt \]
\[ Edx_3 = \left[ \frac{1}{2} x_3 (-p_1 \phi_A + q_2 \phi_B) + rD \right] dt \]
\[ Edx_4 = \left[ \frac{1}{2} x_4 (-p_1 \phi_A - p_2 \phi_B) - rD \right] dt \]

(10)

where

\[ \phi_A = \Delta \Gamma_A + \sigma_A^2 \left( \frac{1}{2} - p_1 \right) - p_2 (\sigma_{A1} b_1 - \sigma_{A2} b_1) + q_2 (\sigma_{A2} b_2 - \sigma_{A1} b_2) \]
\[ \phi_B = \Delta \Gamma_B + \sigma_B^2 \left( \frac{1}{2} - p_2 \right) - p_1 (\sigma_{A1} b_1 - \sigma_{A2} b_1) + q_2 (\sigma_{A2} b_2 - \sigma_{A1} b_1) \]
\[ \Delta \Gamma_A = (\mu_{A1} - \frac{1}{2} \sigma_{A1}^2) - (\mu_{A2} - \frac{1}{2} \sigma_{A2}^2) \]
\[ \Delta \Gamma_B = (\mu_{B1} - \frac{1}{2} \sigma_{B1}^2) - (\mu_{B2} - \frac{1}{2} \sigma_{B2}^2) \]
\[ \sigma_A^2 = \sigma_{A1}^2 + 2 \sigma_{A1} \sigma_{A2} \]
\[ \sigma_B^2 = \sigma_{B1}^2 + 2 \sigma_{B1} \sigma_{B2} \]
\[ p_1 = (1 - q_1) = x_1 + x_2 \]
\[ p_2 = (1 - q_2) = x_1 + x_3 \]
\[ D = x_4 - x_2 x_3 \]

The diffusion coefficients may be obtained by noting that

\[ E[\Delta x_k^T (i) \Delta x_l^T (i) \mid x(i)] = \sum \Sigma C_{i1} C_{i2} E[\Delta x_k^T (i, j_1) \Delta x_l^T (i, j_2) \mid \mathcal{F}(i)] \]  
\[ = \sum \Sigma C_{j1} E [\Delta x_k^T (i, j_1) \Delta x_l^T (i, j_2) \mid \mathcal{F}^+ (i)] + 0(\tau^2) \]
\[ = (1 - \Pi) E[\Delta x_k^T (i, j_1) \Delta x_l^T (i, j_2) \mid \mathcal{F}^+ (i)] + 0(\tau^2) \]

where

\[ \Pi = \sum_{i \neq j} \Sigma C_{i1} C_{i2} \]  

(13)

The terms of order \( \tau^2 \) in these expressions come from the assumption of independence between patches. This assumption produces terms in the expectations which involve products of first order moments, each of which is proportional to \( \tau \).

Expansion of the numerator of the differences \( \Delta x_k^T (i) \) and \( \Delta x_l^T (i) \) in (12) and use of assumption (d) allows us to write

\[ E[\Delta x_k^T (i) \Delta x_l^T (i)] = \frac{1}{4} (1 - \Pi) x_k^T x_l^T E[H_k^T (i, j_1) H_l^T (i, j_2)] \]  

(14)
If we define the matrix

$$[h_{kl}(\tilde{X})] = \left[ \lim_{\tau \to 0} \frac{H_{i,j}^\tau (i,j)}{\tau} \right]$$

we can immediately obtain the diffusion coefficients:

$$Edx_i dx_j = \frac{1}{4} (1-\Pi) x_i x_j h_{ij} (\tilde{X}) \, dt .$$

For example:

$$Edx_1^2 = \frac{1}{4} (1-\Pi) x_1^2 \left( q_1^2 \sigma_A^2 + 2q_1q_2 \sigma_{AB} + q_2^2 \sigma_B^2 \right) \, dt$$

$$Edx_i dx_3 = \frac{1}{4} (1-\Pi) x_i x_3 \left[ p_i q_i \sigma_A^2 + (q_1 q_2 + p_1 p_2) \sigma_{AB} - p_2 q_2 \sigma_B^2 \right] \, dt$$

where

$$\sigma_{AB} = \sigma_{A_1B_2} + \sigma_{A_2B_2} - \sigma_{A_1B_2} - \sigma_{A_2B_1} .$$

This diffusion model is written in terms of the four gametic frequencies which are constrained to sum to one. Rather than working with this gametic process, it is more instructive to transform this equation into one involving the allele frequencies and the linkage disequilibrium as the state variables. To execute this transformation we require the first two moments of the stochastic differentials \(dp_{1}, dp_{2}, \) and \(dD\). These turn out to be very cumbersome to write down in their full generality, so we will specialize to a symmetric model. For the second order moments let

$$\sigma_{A_1} = \sigma_{A_2} = \sigma_{B_1} = \sigma_{B_2} = \sigma^2$$

$$\sigma_{A_1A_2} = \sigma_{B_1B_2} = \alpha \sigma^2$$

$$\sigma_{A_1B_1} = \sigma_{A_2B_2} = -\sigma_{A_1B_2} = -\sigma_{A_2B_1} = \rho \sigma^2 .$$

For the covariance matrix to be positive-definite under this restriction, the following conditions must be met:

$$|\alpha| < 1$$

$$|\rho| < \frac{1 - \alpha}{2}$$

(see Appendix II, Gillespie and Langley 1976).

In addition, let

$$\Delta \Gamma_A = \Delta \Gamma_B = \Delta \Gamma .$$

The rules for obtaining the moments of the stochastic differentials in a new state space follow immediately from Itô's formula for stochastic differentials (see Friedman 1975, p. 78) and may be stated as follows. Let \( \tilde{X} = (x_1, \ldots, x_n) \)
be a diffusion process and let \( \psi = (\psi_1(X), \ldots, \psi_n(X)) \) be a one-to-one transformation to a new state space. Then

\[
Ed\psi_i = \left[ \sum_{j=1}^{n} \frac{\partial \psi_i}{\partial x_j} E(dx_j) + \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 \psi_i}{\partial x_j \partial x_k} E(dx_j dx_k) \right]
\]

\[
Ed\psi_i d\psi_j = \left[ \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{\partial \psi_i}{\partial x_k} \frac{\partial \psi_j}{\partial x_i} E(dx_j dx_i) \right].
\]

Applying this to \( p_1, p_2, \) and \( D, \) we get, after a very lengthy calculation,

\[
Edp_1 = \left[ \frac{1}{2} p_1 q_1 \phi_A + \frac{1}{2} D \phi_B \right] dt
\]

\[
Edp_2 = \left[ \frac{1}{2} p_2 q_2 \phi_B + \frac{1}{2} D \phi_A \right] dt
\]

\[
EdD = \left[ \frac{1}{2} D \left( (q_1-p_1) \phi_A + (q_2-p_2) \phi_B - 2r \right) - \frac{\epsilon}{2} \sigma^2 \left[ (p_1 q_1 + p_2 q_2) \right.ight.
\]

\[
\left. \left. (1-\alpha) D + (p_1 p_2 q_1 q_2 + D^2) 2p \right] \right] dt
\]

\[
E(dp_1^2) = \frac{\epsilon}{2} \sigma^2 \left[ (q_1^2 + D^2)(1-\alpha) + 4p p_1 q_2 D \right] dt
\]

\[
E(dp_2^2) = \frac{\epsilon}{2} \sigma^2 \left[ (q_2^2 + D^2)(1-\alpha) + 4p p_2 q_1 D \right] dt
\]

\[
E(dp_1 dp_2) = \frac{\epsilon}{2} \sigma^2 \left[ (p_1 q_1 + p_2 q_2) D (1-\alpha) + (p_1 p_2 q_1 q_2 + D^2) 2p \right] dt
\]

\[
E(dp_1 dD) = \frac{\epsilon}{2} \sigma^2 \left\{ (1-\alpha) \left[ p_1 q_2 (q_1 - p_2) D + (q_2 - p_2) D^2 \right] + 2p [p_1 q_2 (q_2 - p_2) D + (q_2 - p_2) D^2] \right\} dt
\]

\[
E(dp_2 dD) = \frac{\epsilon}{2} \sigma^2 \left\{ (1-\alpha) \left[ p_2 q_2 (q_2 - p_2) D + (q_1 - p_1) D^2 \right] + 2p [p_2 q_2 (q_1 - p_1) D + (q_2 - p_2) D^2] \right\} dt
\]

where \( \epsilon = 1-11. \)

We wish to study the asymptotic properties of this diffusion equation. It does not seem possible, however, to obtain an exact expression for the asymptotic density, or even to establish the criterion for the existence of such a density. Fortunately, the nature of the model suggests immediately a method of approximating the asymptotic density in a manner analogous to that utilized by \textsc{Felsenstein} (1974) in his study of genetic drift in two-locus systems. Notice the \( \epsilon = 1-11 \) is the reciprocal of the “effective number of patches” as defined in
Gillespie (1974). As the effective number of patches increases, $\varepsilon$ approaches zero and the system (23) approaches the ordinary differential equations:

\[
\frac{dp_1}{dt} = \frac{1}{2} p_1 q_1 \phi_A + \frac{1}{2} D \phi_B
\]

\[
\frac{dp_2}{dt} = \frac{1}{2} p_2 q_2 \phi_B + \frac{1}{2} D \phi_A
\]

\[
\frac{dD}{dt} = \frac{1}{2} D \left[ (q_1 - p_1) \phi_A + (q_2 - p_2) \phi_B - 2 r \right]
\]

which yield a stable polypolymorphism at both loci if and only if

\[|\Delta \Gamma| < \varepsilon \left[ (1 - \alpha) + 2 \rho \right],\]

in which case the equilibrium state is

\[
\hat{p}_1 = \hat{p}_2 = \frac{1}{2} + \frac{\Delta \Gamma}{2 \alpha^2 \left[ (1 - \alpha) + 2 \rho \right]} = \hat{p}
\]

\[D = 0.
\]

Gillespie and Langley (1976).

In the next section the linearization of (23) around (26) will be explored.

THE LINEARIZED PROCESS

If we linearize (23) around (26), keeping the constant and linear terms in the drift coefficients and the constant terms only in the diffusion coefficients we obtain a new process whose asymptotic density is known to be a multidimensional normal density. This procedure has been used previously by Felsenstein (1974) and others referred to in Felsenstein’s paper. This new process, centered on the origin, can be written:

\[
Edp_1^* = [-ab p_1^* - 2ac p_2^*] dt
\]

\[
Edp_2^* = [-2ac p_1^* - ab p_2^*] dt
\]

\[
EdD^* = [-ea h c p_1^* - ea h c p_2^* - (r + \varepsilon ab) D^*] dt
\]

\[
Edp_i^{*2} = 1/2 a^2 b dt, i = 1,2.
\]

\[
EdD^{*2} = 2 \varepsilon a^2 b dt
\]

\[
Edp_1^* dp_2^* = \varepsilon a^2 c dt
\]

\[
Edp_i^* dD = 0; i = 1,2.
\]

where

\[
p_i^* = p_i - \hat{p}
\]

\[
D^* = D + \frac{\varepsilon a^2 c}{r + \varepsilon ab}
\]
and

\[ a = \hat{\rho} \hat{q} \quad b = \sigma^2 (1 - \alpha) \]
\[ c = \sigma^2 \rho \quad h = (\hat{q} - \hat{\rho}) . \]

Since the asymptotic mean of the process (27) is \( E(p_1^* p_2^* D^*) = (0, 0, 0) \), we see immediately that the first order moments of the noncentered process are

\[ E\rho_i = \hat{\rho} = \frac{1}{2} + \frac{\Delta \Gamma}{2\sigma^2[(1-\alpha) + 2\rho]} , \quad i = 1, 2 \]
\[ ED = -\varepsilon \hat{\rho} \hat{q} \sigma^2 \rho + O(\varepsilon^2) . \]

Notice that the sign of \( ED \) is opposite to that of \( \rho \), and is zero if and only if \( \rho \) is zero. The value of \( D \) is affected by two successive events. The first is selection and recombination within each patch which, in our case, follows the well-known patterns of additive models. The second is the mixing of gametes from different patches, during the free migration phase, which also affects \( D \) through the relationship

\[ D_T = \bar{D} + \text{COV} (p_1, p_2) \]

due to Prout (1973) and Nei and Li (1973). Since events within each patch affect both the disequilibrium and the allele frequencies, we would not expect the two terms in (31) to be independent. To understand the relationships, we will examine, in the original discrete-time model, what happens when \( D = 0 \) for the individuals entering the patches. In this case, in the \( j \)th patch, \( D_j \) will change according to

\[ \Delta D_j = \Delta X_{1j} \Delta X_{2j} - \Delta X_{2j} \Delta X_{3j} = -\Delta p_{1j} \Delta p_{2j} \]
\[ = -p_1 p_2 q_1 q_2 \frac{\left( [U_{A_1} (j) - U_{A_2} (j)] [U_{B_1} (j) - U_{B_2} (j)] \right)}{4 \bar{W} (j)^2} . \]

The change in \( D(j) \) can be positive or negative depending on the sign of the product appearing in the brackets. The expected change within the patch, to the order of approximation \( \tau \), is

\[ E \Delta D = -p_1 p_2 q_1 q_2 \rho \sigma^2 \tau + O(\tau^2) . \]

Thus events within a patch tend to give \( D \), on the average, a sign opposite to that of \( \rho \). The expression (32) is sufficiently transparent to see why this is the case.

Mixing gametes between patches causes the overall \( D \) to change according to (31). In this expression, since we are assuming, initially, that \( D = 0 \), we have

\[ \bar{D} = \sum C_j \Delta D(j) \]

and

\[ E \bar{D} = \sum C_j E \Delta D(j) = -p_1 p_2 q_1 q_2 \rho \sigma^2 \tau + O(\tau^2) . \]
On the other hand, it is easily shown that

$$\text{COV}(p_1, p_2) = \Sigma C_j \Delta p_{1j} \Delta p_{2j} - \Sigma \Sigma C_j C_k \Delta p_{1j} \Delta p_{2k}$$

so that the expectation of $\text{COV}(p_1, p_2)$ is

$$E \text{COV}(p_1, p_2) = (\Sigma C_j - \Sigma C_j^2) E \Delta p_{1j} \Delta p_{2j} + 0(\tau^2)$$

$$= \Pi p_1 p_2 q_1 q_2 \rho^2 \tau + 0(\tau^2)$$

(37)

using the independence between patches. The fact that this expression is of the same form as (35) is traceable to the relationship $\Delta D = -\Delta p_1 \Delta p_2$. The effect of mixing is to change $D$ in the opposite direction of the events within a patch, but quantitatively the mixing process cannot change $D$ enough to reattain $D = 0$. Adding (35) and (37) we get

$$D_\tau = -(1 - \Pi) p_1 p_2 q_1 q_2 \rho^2 \tau + 0(\tau^2)$$

(38)

which, as $\tau \to 0$, gives us the expression for $EdD$ for the case $D = 0$ in the diffusion model

$$EdD = -\epsilon p_1 p_2 q_1 q_2 \rho^2 dt$$

(39)

As the effective number of patches increases, $E \text{COV}(p_1, p_2)$ approaches $EdD$ so that we get, at the limit, a $D$ which stays identically zero.

These arguments hold for $D = 0$, but qualitatively indicate the behavior of the entire system, since $D$ cannot change sign without passing through $D = 0$. The behavior of $D$ when it is not equal to 0 is, of course, exceedingly complex. The effects causing $D$ to differ, on the average, from zero, are due to temporally variable selection and recombination within the patches, and, indeed, the effect is present even for a population occupying a single patch ($\epsilon = 1$). The effect of spatial heterogeneity is to counter the formation of linkage disequilibrium. This behavior is in many ways analogous to that described in studies of spatial variation with restricted migration (Li and Nei 1973; Christiansen and Feldman 1975; Slatkin 1975) but differs in that additive models with no migration (single-patch models) can generate mean linkage disequilibrium at the dynamic steady state.

The second order moments of the linearized process (27) can be obtained by the following standard argument. Let $\bar{X} = (x_1, x_2, \ldots, x_n)$ be an $n$-dimensional Orstein-Uhlenbeck process centered on the origin with moments

$$Edx_i = [\Sigma a_{ik} x_k] dt$$

$$Edx_i^2 = V_i dt$$

(40)

$$Edx_i dx_j = C_{ij} dt$$

Let

$$b_{ij}(t) = E[x_i(t) x_j(t)]$$

(41)
It is easy to see that

$$\frac{db_{ij}(t)}{dt} = 2 \sum_{k=1}^{n} a_{ik}b_{jk} + V_i$$

$$\frac{d b_{ij}(t)}{dt} = \sum_{k=1}^{n} [a_{ik}b_{jk} + a_{jk}b_{ik}] + C_{ij}.$$  \tag{42}

The asymptotic second order moments must satisfy (42) with the derivatives set equal to zero. When this procedure is followed for (27), we obtain the asymptotic second order moments:

$$\text{VAR}(p_i) = \frac{e \hat{p}\hat{q} \left[ (1-\alpha)^2 - 2\rho^2 \right]}{4 \left[ (1-\alpha)^2 - 4\rho^2 \right]}, \quad i = 1, 2$$

$$\text{COV}(p_1, p_2) = \frac{-e \hat{p}\hat{q} \rho}{4 \left[ (1-\alpha)^2 - 4\rho^2 \right]}$$

$$\text{VAR}(D) = \frac{e \hat{p}\hat{q}^2 \alpha^2 (1-\alpha)}{r} + 0 (\varepsilon^2)$$

$$\text{COV}(p_i, D) = 0 (\varepsilon^2).$$

Of immediate interest, in these second order moments is the fact that the allele frequencies are almost uncorrelated to the linkage disequilibrium. Their covariance is of order $\varepsilon^2$, their correlation of order $\varepsilon$. This is similar to the results of Felsenstein (1974). Unlike drift process, however, we have a significant correlation between the allele frequencies at the different loci. The correlation, like mean disequilibrium, is opposite in sign from the correlation, $\rho$, for reasons analogous to those giving the ED result. Consider, again, the case where $\rho > 0$. Examination of (42) shows that two forces are acting in opposite directions on $\text{COV}(p_1, p^2)$. The term $C_{12}$ causes the covariance to increase, i.e., move toward a value with the same sign as $\rho$. But the terms in the sum work in the opposite direction. For example, if $p_1$ happens to exceed $1/2$, then the “genetic background” of the allele $B_1$ is increased and its overall fitness is decreased. Consequently $B_1$ will tend to decrease due to the effects of the correlation. This effect tends to give $p_1$ and $p_2$ a correlation opposite in sign to $\rho$. To discover which effect predominates, we must resort to an examination of the mathematics which, as we saw, gave the second effect the nod.

This correspondence in sign between $\text{COV}(p_1, p_2)$ and $ED$ suggests that evidence for this model could be obtained by examining suitable data. Although no published data are helpful, CHARLES LANGLEY has just finished an analysis of a large data set from Drosophila melanogaster which FRANK JOHNSON obtained and discovered that there is a tendency for $ED$ and $\text{COV}(p_1, p_2)$ to have the same sign. This analysis will be published shortly.

The results given here further strengthen the view that correlations in fitnesses can be a major force in organizing the genome. In the previous paper (GILLESPIE and LANGLEY 1976), it was shown that the asymptotic allele frequencies are
MULTILOCUS SELECTION

highly interdependent due to the effects of correlations, while the present paper shows that correlations can also cause nonrandom associations of alleles on chromosomes. This is particularly remarkable given the additive nature of the underlying model.

LITERATURE CITED


LEVENE, H., 1953 Genetic equilibrium when more than one niche is available. Amer. Nat. 87: 331-333.


Corresponding editor: W. J. EWENS