Genet. Res., Camb. (1975), **24**, *pp.* 151–162
With 1 *text-figure*
Printed in Great Britain

The effect of population subdivision on two loci without selection

By MARCUS W. FELDMAN and FREDDY BUGGE CHRISTIANSEN*

Department of Biological Sciences, Stanford University, Stanford, California 94305

(*Received* 14 *May* 1974)

SUMMARY

This paper is devoted to the study of the effects of population subdivision on the evolution of two linked loci. Two simple deterministic models of population subdivision without selection are investigated. One is a finite linear 'stepping stone' model and the other is a finite linear stepping stone chain of populations stretching between two large populations of constant genetic constitution. At equilibrium in the first model the gene frequencies in each population are equal and there is linkage equilibrium in each population. The rate of decay to zero of the linkage disequilibrium functions is the larger of $(1-c)$ and λ_1^2, where λ_1 is the rate of convergence of the gene frequencies to equilibrium and c is the recombination frequency. In the second model at equilibrium there will be a linear cline in gene frequencies connecting the two large constant populations. This cline will be accompanied by a 'cline' of linkage disequilibria. The rate of convergence to this equilibrium cline is independent of the recombination frequency, and, in fact, the gene frequencies and the linkage disequilibria converge to equilibrium at the same rate.

The effect of population subdivision on one locus without selection in diploid organisms was early recognized by Wahlund (1928). When previously isolated random mating populations are mixed, the genotypic proportions show a deficit of heterozygotes at a locus of variable gene frequency if compared to the Hardy–Weinberg proportions with the mean gene frequency. However, only one round of random mating is required to restore Hardy–Weinberg proportions. On the other hand, it has long been recognized that the mixing of isolates could in a similar fashion generate linkage disequilibrium between pairs of loci, although quantitative expression of this has only recently been presented (Sinnock & Singh, 1972; Prout, 1973; Nei & Li, 1973). Unlike the distortion from Hardy–Weinberg proportions this distortion from random combination of the alleles at two loci is not expected to vanish after one generation of random mating (Robbins, 1918). This suggests that the fundamental rules for the dynamics of a two locus polymorphism need further consideration in subdivided populations where the subpopulations are not com-

* Permanent address: Department of Genetics, University of Aarhus, DK8000 Aarhus C, Denmark.

pletely isolated but interchange individuals regularly through recurrent migration. This paper is an attempt at clarification of these rules in simple deterministic models of subdivided populations without selection.

1. WAHLUND'S PRINCIPLE FOR TWO LOCI

Consider two loci with alleles A and a segregating at the first and B and b at the second locus. In the ith subpopulation let the frequencies of the chromosomes AB, Ab, aB and ab be x_{1i}, x_{2i}, x_{3i} and x_{4i} respectively. Write the gene frequency of A as $p_i = x_{1i}+x_{2i}$ and the gene frequency of B as $r_i = x_{1i}+x_{3i}$ with $q_i = 1-p_i$ and $s_i = 1-r_i$. Instead of the chromosome frequencies we may use the system of independent variables p_i, r_i and $D_i = x_{1i}x_{4i} - x_{2i}x_{3i}$ to describe the evolution of the population. D_i is called the linkage disequilibrium and measures, in the ith subpopulation, the departure from random association of the alleles at the two loci.

Now suppose that the whole population is divided into n subunits of relative size e_i, $i = 1, 2, \ldots, n$ with $\Sigma e_i = 1$. Then the average value of the linkage disequilibrium is

$$\bar{D} = \sum_{i=1}^{n} e_i D_i = D_T - \mathrm{cov}\,(p, r), \tag{1}$$

where
$$D_T = \bar{x}_1\bar{x}_4 - \bar{x}_2\bar{x}_3 \quad \text{with} \quad \bar{x}_j = \sum_{i=1}^{n} e_i x_{ji} \tag{2}$$

and
$$\mathrm{cov}\,(p, r) = \sum_{i=1}^{n} e_i(p_i - \bar{p})\,(r_i - \bar{r}) \tag{3}$$

with
$$\bar{p} = \sum_{i=1}^{n} e_i p_i \quad \text{and} \quad \bar{r} = \sum_{i=1}^{n} e_i r_i. \tag{4}$$

The identity (1) has been obtained for the case of two populations by Sinnock & Singh (1972) and for the general case by Prout (1973) and Nei & Li (1973). We propose that (1) be called Wahlund's principle for two loci by analogy with the famous principle for one locus (Wahlund, 1928). To see the analogy, let h_i be the frequency of heterozygotes in the ith population. Define the deviation from the Hardy–Weinberg proportions as $\delta_i = \frac{1}{2}h_i - p_i q_i$. The average deviation from Hardy–Weinberg proportions in the population is then

$$\bar{\delta} = \sum_{i=1}^{n} e_i \delta_i = \delta_T + \mathrm{var}\,(p), \tag{5}$$

where
$$\delta_T = \tfrac{1}{2}\bar{h} - \overline{pq} \quad \text{with} \quad \bar{h} = \sum_{i=1}^{n} e_i h_i \tag{6}$$

and
$$\mathrm{var}\,(p) = \sum_{i=1}^{n} e_i(p_i - \bar{p})^2. \tag{7}$$

The qualitative implications of (1) are also analogous to that for one locus; if a population is subdivided and the disequilibrium, D_T, is calculated as though no subdivision existed and found to be zero, there may be a substantial error made in inferring that the average disequilibrium is zero. The magnitude of the error is governed by the covariance in the gene frequencies at the two loci.

The transient properties of the linkage disequilibrium are fundamentally different from the transient properties of deviations from the Hardy–Weinberg proportions (Robbins, 1918). Random mating in a mixed population produces an offspring population in Hardy–Weinberg proportions at each locus, whereas the linkage disequilibrium among offspring becomes

$$D'_T = (1-c)\,\bar{D} + \mathrm{cov}\,(p, r), \tag{8}$$

where c is the recombination fraction. We shall see that the gene frequencies play a key role in the transient properties of the disequilibria in subdivided populations, a role which is basically due to the Wahlund's covariance principle as stated in equations (1) and (8).

2. 'STEPPING-STONE' MODELS FOR TWO LOCI

We have seen that covariance in the gene frequencies of subpopulations can produce linkage disequilibrium in the overall system despite the random association of the loci within each subpopulation. In this section we investigate the long term behaviour of the linkage disequilibrium in some simple models of population subdivision in view of the Wahlund principle.

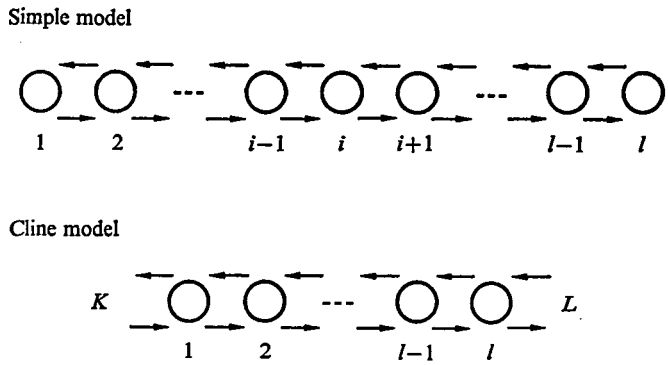

Fig. 1. Stepping-stone models (see text).

We consider the simple finite one-dimensional stepping-stone model of migration used by Kimura & Weiss (1964) and depicted in Fig. 1. There are two autosomal loci with alleles A, a, B and b in frequencies $p_i(t)$, $q_i(t)$, $r_i(t)$ and $s_i(t)$ respectively in the ith subpopulation at generation t. The chromosome frequencies are designated $x_{ji}(t)$ while the linkage disequilibrium in the ith population is $D_i(t)$ at generation t. There are l subpopulations of equal size arrayed along a line and after migration each subpopulation consists of a proportion m of immigrants from the population on the left and a proportion m from the population on the right with the remaining individuals raised on location. Mating is at random within each of the subpopulations and it is assumed that there is no selection.

With this basic structure we consider two qualitatively different models. The first is essentially as described above with no further assumptions. We call this model the *simple stepping-stone* model. The second model is arrived at by assuming that at the left and right hand ends of the array there are two large populations K and L respectively (Fig. 1). In these the gene frequencies of A and B are p_K, r_K and p_L and r_L respectively and the disequilibria D_k and D_L are constant, uninfluenced by the gene flow through the cline. We call this model the *stepping-stone cline* model.

3. THE SIMPLE STEPPING STONE MODEL

The changes in gene frequencies of A between successive generations may be written for the simple model as

$$\mathbf{p}(t) = \mathbf{M}p(t-1), \tag{9}$$

where

$$\mathbf{M} = \begin{bmatrix} 1-m & m & 0 & 0 & \dots & 0 & 0 \\ m & 1-2m & m & 0 & \dots & 0 & 0 \\ 0 & m & 1-2m & m & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \dots \\ 0 & 0 & 0 & 0 & \dots & m & 1-m \end{bmatrix} \quad \text{and} \quad \mathbf{p}(t) = \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ \dots \\ p_l(t) \end{bmatrix} \tag{10}$$

and similarly for $\mathbf{r}$ the vector of B gene frequencies. Models like (9) have been studied extensively by Malecot (1950, 1951), Kimura & Weiss (1964) and Bodmer & Cavalli-Sforza (1968). Continuous time and space analogues have been considered by Malecot (1967) and Maruyama (1971). Now if the initial frequencies in the subpopulations are $\mathbf{p}(0)$ then the elementary properties of the matrix $\mathbf{M}$ ensure that the limiting frequency in each subpopulation is

$$p_i = \left\{ \sum_{j=1}^{l} p_j(0) \right\} \Big/ l, \tag{11}$$

i.e. asymptotically there is no variation in gene frequencies. The rate of approach of the system of frequencies to this limiting uniform state is the largest non-unit eigenvalue of $\mathbf{M}$, namely

$$\lambda_1 = 1 - 2m + 2m \cos [\pi/l]. \tag{12}$$

This and other spectral properties of $\mathbf{M}$ may be obtained using arguments analogous to those of Feller (1957, ch. XVI). When the number of populations is large we have

$$\lambda_1 = 1 - m\pi^2/l^2, \tag{13}$$

so the larger the number of populations, the slower is the approach to the limit (11). This value (13) for the rate of approach to (11) is the precise discrete analogue to the continuous time and space result (2–4) of Maruyama (1971), since $2m$ is the variance of the migration distance in (10) and l corresponds to Maruyamas L.

Next we consider the change in the values of the linkage disequilibria for the

simple stepping-stone model. Zygotes are assume to migrate and recombination occurs after migration. Thus after migration and recombination we have, for example,

$$x'_{1i} = (1-2m)\,x_{1i} + m(x_{1,(i-1)} + x_{1,(i+1)}) - c[(1-2m)\,D_i + m(D_{i-1} + D_{i+1})], \tag{14}$$

for $i \neq 1, l$, with appropriate adjustments at $i = 1$ or $i = l$. Here c is the recombination fraction. Similar relations are valid for the frequencies of Ab, aB and ab. From equation (14) we compute the new disequilibria for $i \neq 1$ and l as

$$D'_i = (1-2m)^2 D_i - c[(1-2m)\,D_i + m(D_{i-1} + D_{i+1})] + m^2(D_{i-1} + D_{i+1}) + m(1-2m)\,(\Delta_{i,i-1} + \Delta_{i,i+1}) + m^2\Delta_{i-1,i+1}, \tag{15}$$

where

$$\Delta_{kj} = x_{1k}x_{4j} + x_{1j}x_{4k} - x_{2k}x_{3j} - x_{2j}x_{3k}. \tag{16}$$

Appropriate adjustments must be made for $i = 1$ or $i = l$. The quantities Δ_{kj} are disequilibrium functions bound to the correlation in gene frequencies between populations k and l. In fact we have

$$\Delta_{kj}(t) = (p_k(t) - p_j(t))\,(r_k(t) - r_j(t)) + D_k(t) + D_j(t). \tag{17}$$

Using (17) we may write (15) in the form of equation (8), which in matrix form is as follows

$$\mathbf{D}(t) = (1-c)\,\mathbf{M}\mathbf{D}(t-1) + \mathbf{C}(t-1), \tag{18}$$

where $\mathbf{D}(t)$ is the column vector $(D_1(t), D_2(t), \ldots, D_l(t))'$ and $\mathbf{C}(t)$ is the vector whose ith component is

$$C_i(t) = \mathbf{p}(t)'\,\mathbf{A}_i\,\mathbf{r}(t), \tag{19}$$

where the matrices $\mathbf{A}_i$ are of the form

$$\mathbf{A}_i = \begin{array}{c} \\ \\ \\ \\ \\ i-1 \\ i \\ i+1 \\ \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} & & & i-1 & i & i+1 & \end{array} \\ \begin{bmatrix} 0 \ldots 0 & 0 & 0 & 0 & 0 \ldots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 \ldots 0 & 0 & 0 & 0 & 0 \ldots 0 \\ 0 \ldots 0 & m(1-m) & -m(1-2m) & -m^2 & 0 \ldots 0 \\ 0 \ldots 0 & -m(1-2m) & 2m(1-2m) & -m(1-2m) & 0 \ldots 0 \\ 0 \ldots 0 & -m^2 & -m(1-2m) & m(1-m) & 0 \ldots 0 \\ 0 \ldots 0 & 0 & 0 & 0 & 0 \ldots 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 \ldots 0 & 0 & 0 & 0 & 0 \ldots 0 \end{bmatrix} \end{array} \tag{20a}$$

https://doi.org/10.1017/S0016672300015184 Published online by Cambridge University Press

for $i \neq 1$ and l with

$$\mathbf{A}_1 = \begin{bmatrix} m(1-m) & -m(1-m) & 0 \ldots 0 \\ -m(1-m) & m(1-m) & 0 \ldots 0 \\ 0 & 0 & 0 \ldots 0 \\ \cdot & \cdot & \cdot \quad \cdot \\ \cdot & \cdot & \cdot \quad \cdot \\ \cdot & \cdot & \cdot \quad \cdot \\ 0 & 0 & 0 \ldots 0 \end{bmatrix} \tag{20b}$$

and

$$\mathbf{A}_l = \begin{bmatrix} 0 \ldots 0 & 0 & 0 \\ \cdot \quad \cdot & \cdot & \cdot \\ \cdot \quad \cdot & \cdot & \cdot \\ \cdot \quad \cdot & \cdot & \cdot \\ 0 \ldots 0 & 0 & 0 \\ 0 \ldots 0 & m(1-m) & -m(1-m) \\ 0 \ldots 0 & -m(1-m) & m(1-m) \end{bmatrix}. \tag{20c}$$

Through (19) it is clear that we might expect the evolution of D_i to depend on changes in $\mathbf{p}$ and $\mathbf{r}$, the gene frequencies. In fact we may use (9) to compute $C_i(t)$ as

$$C_i(t) = \mathbf{p}(0)' (\mathbf{M}')^t \mathbf{A}_i \mathbf{M}^t \mathbf{r}(0). \tag{21}$$

Now $\mathbf{M}^t$ may be written as $\boldsymbol{\phi}\boldsymbol{\Lambda}^t\boldsymbol{\psi}$, where $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues, $\boldsymbol{\phi}$ is the matrix whose columns are the right eigenvectors of $\mathbf{M}$, and $\boldsymbol{\psi}$ the matrix whose rows are the left eigenvectors of $\mathbf{M}$ and $\boldsymbol{\phi}\boldsymbol{\psi} = \mathbf{I}$, the identity matrix. A right eigenvector of $\mathbf{M}$ for the eigenvalue 1 is obviously $(1, 1, \ldots, 1)$ and the corresponding left eigenvector is $(\pi_1, \pi_2, \ldots, \pi_l)$, the stationary probability distribution, which is the uniform distribution $\pi_i = 1/l$. From the definitions (20) we see that these two vectors are eigenvectors of 0 for all the matrices $\mathbf{A}_i$. Substituting this information into (21) it becomes clear that the matrices $\mathbf{A}_i$ remove the unit eigenvalue of $\mathbf{M}$ from consideration in the iteration of $\mathbf{C}(t)$. The rate determining factor is then the square of the largest non-unit eigenvalue of $\mathbf{M}$ due to the occurrence of $\mathbf{M}'$ and $\mathbf{M}$ in (21). We may rewrite (18) as

$$\begin{aligned} \mathbf{D}(t) &= (1-c)^t \mathbf{M}^t \mathbf{D}(0) + \mathbf{C}(t-1) + (1-c)\,\mathbf{M}\,\mathbf{C}(t-2) + \ldots + (1-c)^{t-1}\mathbf{M}^{t-1}\mathbf{C}(0) \\ &= (1-c)^t \mathbf{M}^t \mathbf{D}(0) + \{\lambda_1^{2(t-1)}\mathbf{I} + (1-c)\,\lambda_1^{2(t-2)}\,\mathbf{M} + \ldots + (1-c)^{t-1}\mathbf{M}^{t-1}\}\boldsymbol{\gamma} \\ &\quad + \text{terms of order } (\lambda_1\lambda_2)^{t-1}, \end{aligned} \tag{22}$$

where λ_1 is the largest non-unit eigenvalue (12) and λ_2 is the third largest eigenvalue of $\mathbf{M}$ and $\boldsymbol{\gamma}$ is a constant vector. Hence

$$\mathbf{D}(t) \approx (1-c)^t \mathbf{M}^t \mathbf{D}(0) + \lambda_1^{2t}\{\mathbf{I} - [(1-c)/\lambda_1^2]\,\mathbf{M}\}^{-1}\{\mathbf{I} - [(1-c)/\lambda_1^2]^t \mathbf{M}^t\}\boldsymbol{\gamma}^*, \tag{23}$$

where $\boldsymbol{\gamma}^*$ is a modified constant vector.

We therefore conclude that $\mathbf{D}(t) \to \mathbf{0}$, as $t \to \infty$. Ultimately the loci in all popula-

tions are randomly associated. Now since **M** has an eigenvalue of unity, the rate at which $\mathbf{D}(t)$ vanishes is governed by the larger of $(1-c)$ and λ_1^2 (see (12) or (13)). Thus if the number of subpopulations is large, i.e. λ_1 is close to unity, then the recombination fraction only plays a role in the asymptotic decay of disequilibrium if it is extremely small. For moderate linkage values, and large numbers of subpopulations, only the migration rate governs the asymptotic decay of **D**, through the definition of λ_1, namely (12). This constitutes a substantial generalization of Nei & Li's result (1973) that in the case of two populations the rate of decay of **D** is the larger of $1-c$ and $(1-2m)^2$. This completes our discussion of the simple stepping-stone model.

4. THE STEPPING-STONE CLINE MODEL

It seems likely that in natural populations spatial variations in density may be correlated with variations in the gene frequencies at polymorphic loci. In sparsely populated areas the gene frequencies will to a large extent be influenced by migration from more dense regions. However, the effect of immigration in the denser areas may not be so important. Any differences in genetic composition between isolated dense areas of a population will result in clines through the sparse areas between them. It seems reasonable to assume that such clines at different loci will be parallel in shape (Lewontin & Krakauer, 1973). From equations (1) and (8) above the steady mixing of the populations in the cline area may be expected to generate overall disequilibrium between varying loci in every generation (Prout, 1973).

Now consider the effect of two large constant populations K and L at the left and right respectively of the stepping-stone chain. For the stepping-stone cline model the gene frequencies of A and B in K and L are p_K and p_L respectively. The constant values of the disequilibria are D_K and D_L respectively. In population 1 a proportion m of the individuals is exchanged with population K, and similarly population l exchanges individuals with population L. The recursion system governing the change in A frequencies is then

$$\mathbf{p}(t) = \mathbf{M}_1 \mathbf{p}(t-1) + m\boldsymbol{\eta}, \tag{24}$$

where

$$\mathbf{M}_1 = \begin{bmatrix} 1-2m & m & 0 & \dots & 0 & 0 \\ m & 1-2m & m & \dots & 0 & 0 \\ 0 & m & 1-2m & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1-2m & m \\ 0 & 0 & 0 & \dots & m & 1-2m \end{bmatrix} \tag{25}$$

and

$$\boldsymbol{\eta}' = (p_K, 0, 0, \dots, 0, p_L). \tag{26}$$

We therefore have

$$\mathbf{p}(t) = \mathbf{M}_1^t \mathbf{p}(0) + m[\mathbf{I} - \mathbf{M}_1^t][\mathbf{I} - \mathbf{M}_1]^{-1}\boldsymbol{\eta} \tag{27}$$

since the largest eigenvalue of $\mathbf{M}_1$ is clearly less than unity. Hence

$$\hat{\mathbf{p}} = \lim_{t\to\infty} \mathbf{p}(t) = m[\mathbf{I} - \mathbf{M}_1]^{-1}\boldsymbol{\eta}. \tag{28}$$

This is of course analogous to the well-known equilibrium result when there is stabilizing migration from a constant outside population (Bodmer & Cavalli-Sforza, 1968) or linear selection (Malécot, 1950, 1951). To express the limit (28) we need only note that the inverse of $\mathbf{I}-\mathbf{M}_1$ has the form

$$\frac{1}{m(l+1)}\begin{bmatrix} l & l-1 & \dots & 2 & 1 \\ l-1 & x & \dots & x & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 2 & x & \dots & x & l-1 \\ 1 & 2 & \dots & l-1 & l \end{bmatrix}, \tag{29}$$

where the interior elements, marked x, play no role because of the form of (26). Together (26), (28) and (29) produce the limiting values

$$p_i[(l-i+1)\,p_K + i p_L]/(l+1). \tag{30}$$

The rate of approach of the gene frequencies to the equilibrium cline (30) is the largest eigenvalue of $\mathbf{M}_1$. With a total of l populations (excluding K and L) this is

$$\lambda_1^* = 1-2m+2m\cos(\pi/(l+1)), \tag{31}$$

which is equivalent to the result (12) for the simple stepping-stone model.

The changes in linkage disequilibria are specified by

$$\mathbf{D}(t) = (1-c)\,\mathbf{M}_1\mathbf{D}(t-1) + (1-c)\,m\boldsymbol{\delta} + \mathbf{C}(t-1), \tag{32}$$

where

$$\boldsymbol{\delta}' = (D_K, 0, 0, \dots, 0, D_L) \tag{33}$$

and $\mathbf{C}(t)$ is the vector (19) with changes at positions 1 and l, such that

$$C_1(t) = \begin{bmatrix} p_K \\ p_1 \\ p_2 \end{bmatrix}' \begin{bmatrix} m(1-m) & -m(1-2m) & -m^2 \\ -m(1-2m) & 2m(1-2m) & -m(1-2m) \\ -m^2 & -m(1-2m) & m(1-m) \end{bmatrix} \begin{bmatrix} r_K \\ r_1 \\ r_2 \end{bmatrix} \tag{34}$$

and similarly for $C_l(t)$.

From (27) we have that $\mathbf{p}(t)-\hat{\mathbf{p}}$ approaches zero at the geometric rate λ_1^{*t}. We may use this fact in the iteration of $\mathbf{C}(t)$ and write (32) as

$$\mathbf{D}(t) = (1-c)\,\mathbf{M}_1\mathbf{D}(t-1) + m(1-c)\,\boldsymbol{\delta} + \hat{\mathbf{C}} + \lambda_1^{*t-1}\boldsymbol{\gamma} + \text{terms of order } (\lambda_1^{*2t}, \lambda_2^{*t}), \tag{35}$$

where $\hat{\mathbf{C}}$ is the equilibrium $\mathbf{C}$ vector obtained by substituting the equilibrium values $\mathbf{p}$ (from (28)) and $\mathbf{r}$ into (19) and (34). λ_2^* is the second largest eigenvalue of $\mathbf{M}_1$ and $\boldsymbol{\gamma}$ is a constant vector. Hence

$$\begin{aligned}\mathbf{D}(t) = (1-c)^t\mathbf{M}_1^t\mathbf{D}(0) + [\mathbf{I}-(1-c)^t\mathbf{M}_1^t]\,[\mathbf{I}-(1-c)\,\mathbf{M}_1]^{-1}\,[m(1-c)\,\boldsymbol{\delta}+\hat{\mathbf{C}}] \\ + \lambda_1^{*t-1}\{\mathbf{I}-[(1-c)/\lambda_1^*]^t\mathbf{M}_1^t\}\,\{\mathbf{I}-[(1-c)/\lambda_1^*]\,\mathbf{M}_1\}^{-1}\,\boldsymbol{\gamma}^*,\end{aligned} \tag{36}$$

where $\boldsymbol{\gamma}^*$ is a modified constant vector. The limiting value of $\mathbf{D}(t)$ is thus seen to be

$$\hat{\mathbf{D}} = \lim_{t\to\infty}\mathbf{D}(t) = [\mathbf{I}-(1-c)\,\mathbf{M}_1]^{-1}\,[m(1-c)\,\boldsymbol{\delta}+\hat{\mathbf{C}}]. \tag{37}$$

From (36) we note the interesting fact that the rate of approach of the disequilibrium functions $\mathbf{D}(t)$ to the equilibrium value $\mathbf{D}$ is quite different from the

https://doi.org/10.1017/S0016672300015184 Published online by Cambridge University Press

result for the simple stepping-stone model. Since λ_1^* is the largest eigenvalue of $\mathbf{M}_1$, the recombination fraction, c, never plays any role in the ultimate rate of convergence. It may well be influential in the early stages of evolution, but the only parameter involved in the asymptotic rate of approach is m. These remarks depend, of course, on the vector $\boldsymbol{\gamma}$ being non-zero. If $p_K = p_L$ so that the gene frequencies are asymptotically uniform across the cline it is easy to see that the structure of $\mathbf{A}_i$ removes the term in λ_1^{*t} leaving terms in $(\lambda_1^{*2})^t$. Then we must compare $(1-c)\lambda_1^*$ and λ_1^{*2} to determine the rate, the larger being the relevant one.

Another point worthy of note is that even when D_K and D_L are zero there may still exist what may be called a cline of the disequilibria due (through Wahlund's effect as in section 1) to the influence of the differences in gene frequencies between populations. Only if $p_K = p_L$ will this effect vanish.

A direct derivation of the equilibrium linkage disequilibrium values can be made as follows: from (19), (30) and (34) the equilibrium value of C_i can be shown to be

$$\hat{C} = [2m/(l+1)]\,(p_K - p_L)\,(r_K - r_L). \tag{38}$$

From (32) we have

$$\hat{D}_i = \{[1-(1-c)\,(1-2m)]/[(1-c)\,m]\}\,\hat{D}_{i+1} - \hat{D}_{i+2} - \hat{C}/[(1-c)\,m]. \tag{39}$$

Now set
$$\Delta_i = \hat{D}_i + \hat{C}/[2m(1-\alpha)\,(1-c)] = \hat{D}_i - \hat{C}/c \tag{40}$$

with
$$\alpha = [1-(1-c)\,(1-2m)]/[2(1-c)\,m]. \tag{41}$$

Then substituting Δ_i into the recursion system (39) produces a homogeneous system of equations

$$\Delta_i = 2\alpha\Delta_{i+1} - \Delta_{i+2}. \tag{42}$$

Using the fact that the characteristic roots of this equation system have the product unity, we may write the solutions in the form

$$\Delta_i = [A_{l-i+1}\Delta_K + A_i\Delta_L]/A_{(l+1)}, \tag{43}$$

where
$$A_j = \alpha^j\{[1+\sqrt{(1-\alpha^{-2})}]^j - [1-\sqrt{(1-\alpha^{-2})}]^j\}. \tag{44}$$

Since the relation between Δ_i and D_i is given by (40), (43) is an explicit expression for D_i. Observe that if both D_K and D_L are zero, then

$$\hat{D}_i = \left[1 - \frac{A_{l-i+1} + A_i}{A_{l+1}}\right]\frac{\hat{C}}{c}. \tag{45}$$

Thus $\hat{D}_i$ increases from both ends of the cline and attains its maximum in the middle. In this case, if l is not too small and α is not too close to one, i.e. c is not too small with respect to the migration rate, then in the middle of the cline we have

$$\hat{D}_i \approx \hat{C}/c \tag{46}$$

as a fairly good approximation. Table 1 shows the magnitude of the terms in (45) that are neglected in the approximation (46). Table 2 shows some of the ranges of the parameters for which (46) is valid.

The linkage disequilibria D_i are defined as the linkage disequilibria among uniting gametes in the subpopulations. Equation (46) shows that in the centre of the cline the equilibrium values $\hat{D}_i$ are proportional to the equilibrium correlation

Table 1. *Values of A_i/A_{2i} between 0·1 and 0·01 for various values of i and α*

(The signs + and − indicate values above 0·1 and below 0·01 respectively.)

	α				
	1·1	1·25	1·5	2	5
i					
2	+	+	+	0·7	0·01
3	+	+	0·06	0·02	—
4	+	0·06	0·03	0·01	—
5	+	0·03	0·01	—	—
6	0·06	0·02	—	—	—
7	0·04	0·01	—	—	—
8	0·03	—	—	—	—
9	0·02	—	—	—	—
10	0·01	—	—	—	—

Table 2. *Values of α for various values of m and c*

	m				
	0·01	0·05	0·1	0·2	0·5
c					
0·01	1·51	1·10	1·05	1·03	1·01
0·05	3·6	1·53	1·26	1·13	1·05
0·1	6·6	2·1	1·56	1·28	1·11
0·2	14	3·5	2·3	1·62	1·25
0·5	15	11	6·0	3·5	2·0

in gene frequencies among immigrants. The linkage disequilibrium among adults after migration should be (2) which from (1) is

$$\hat{D}_{Ti} = (1-2m)\hat{D}_i + m(\hat{D}_{i-1} + \hat{D}_{i+1}) + \hat{C} \approx \hat{C}(1+c)/c \tag{47}$$

with the approximation in (46).

5. DISCUSSION

The occurrence of linkage disequilibrium in a population may be due to the action of almost any of the usual evolutionary forces considered in population genetics. It can result from small population sampling effects (Hill & Robertson, 1968; Sved, 1971) from epistatic selection on tightly linked loci (Lewontin & Kojima, 1960; Bodmer & Felsenstein, 1967; Karlin & Feldman, 1970) or from departures from random mating such as have been investigated here. Discrimination among these on the basis of disequilibrium data from natural populations promises to be as difficult a problem as the discrimination based on allele frequency data has been. A number of results have been obtained on the joint effects of drift, linkage and migration (Feldman, Stam & Wagner in preparation). Ohta (1973) has

incorporated a special form of selection into the island model of Wright (1943) thereby allowing all of the above effects to be studied at once. The problem then becomes highly non-linear, of course, and the results are based on diffusion approximations. At the moment, however, it appears that these theories still are insufficient for the elucidation of an inferential methodology.

In this connexion we should compare the result obtained in this note, that a regular gradient of linkage disequilibrium can be generated by non-uniformity in the gene frequencies, with the fact that selection in a homogeneous environment even with complete uniformity in the gene frequencies, can generate variation in the linkage disequilibrium among subpopulations (Christiansen & Feldman, in preparation). This variation in the linkage disequilibrium may in fact be as expected in small isolated populations. Therefore, neither regularity nor variation can be assigned to a specific evolutionary force.

When dealing with a single locus the hypothesis that a deficiency of heterozygotes is due to subdivision and geographical variation may be tested by comparing an observed deficit with the variance in gene frequency (e.g. Christiansen *et al.* 1973 or Sick, 1965). One might expect that the covariance in formula (1) could be used in an analogous way when comparing an observed linkage disequilibrium value to the same hypothesis. However, the results here show that considerable variation in the disequilibrium may occur even when there is uniformity in the correlation and the prediction of the magnitude of the linkage disequilibria requires at least an estimate of c. Further, it might also be possible to study disequilibria in a manner similar to the Wahlund's variances used by Lewontin & Krakauer (1973) but with the covariances. One source of the difficulty in doing this arises from the fact that with one locus we have a simple estimate of the inbreeding coefficient, which in certain migration contexts is very useful (e.g. Kidd & Cavalli-Sforza, 1974), whereas with two loci there are many 'inbreeding coefficients'.

From the early work of Robbins (1918) we know that linkage disequilibrium vanishes at a geometric rate determined by the recombination fraction in a panmictic population. The effect of subdivision on linked genes in our work may be interpreted as a secondary restriction on recombination that causes a delay in the convergence to linkage equilibrium for large recombination fractions (the simple stepping-stone model). On the other hand, with external stabilizing forces as in the stepping-stone cline model the long-term rate of change of the linkage disequilibrium in large populations is apparently uninfluenced by the degree of linkage with migration as the only rate determining factor. The equilibrium values do of course depend on the extent of linkage in the cline model.

The authors thank Professor W. F. Bodmer for his valuable comments on the manuscript.

This research supported in part by a grant from National Science Foundation, U.S.A., a grant from the Danish Natural Science Research Council, and by USPHS grant NIH 10452–09–11.

REFERENCES

Bodmer, W. F. & Cavalli-Sforza, L. L. (1968). A migration matrix model for the study of random genetic drift. *Genetics* **59**, 565–592.

Bodmer, W. F. & Felsenstein, J. (1967). Linkage and selection: Theoretical analysis of the deterministic two locus random mating model. *Genetics* **57**, 237–265.

Christiansen, F. B., Frydenberg, O. & Simonsen, V. (1973) Genetics of Zoarces populations. IV. Selection component analysis of an esterase polymorphism using population samples including mother–offspring combinations. *Hereditas* **73**, 291–304.

Feller, W. (1957). *An introduction to probability theory and its applications*, 2nd ed. New York: John Wiley.

Hill, W. G. & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226–231.

Karlin, S. & Feldman, M. W. (1970). Linkage and selection: Two locus symmetric viability model. *Theoretical Population Biology* **1**, 39–71.

Kidd, K. & Cavalli-Sforza, L. L. (1974). The role of genetic drift in the differentiation of Icelandic and Norwegian cattle. *Evolution* (to appear).

Kimura, M. & Weiss, G. H. (1964). The steppibg stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561–576.

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

Lewontin, R. C. & Krakuaer, J. (1973). Distribution of gene frequency as a test of the neutrality of polymorphisms. *Genetics* **74**, 175–195.

Malécot, G. (1950). Quelques schémas probabilistés sur la variabilité des populations naturelles. *Annales Université de Lyon*, Science Section A **13**, 36–60,

Malécot, G. (1951). Un traitement stochastique des problèmes linéares (mutation, linkage et migration en génétique de populations). *Annales Université de Lyon*, Science Section A **14**, 79–117.

Malécot, G. (1967). Identical loci and relationship. *Proceedings of the 5th Berkeley Symposium in Mathematical Statistics and Probability* IV; 317–332.

Maruyama, T. (1971). The rate of decrease of heterozygosity in a population occupying a circular or linear habitat. *Genetics* **67**, 437–454.

Nei, M. & Li, W. (1973). Linkage disequilibrium in subdivided populations. *Genetics* **75**, 213–219.

Ohta, T. (1973). Effect of linkage on behaviour of mutant genes in finite populations. *Theoretical Population Biology* **4**, 145–162.

Prout, T. (1973). Appendix to: 'Population genetics of marine pelecypods. III. Epistasis between functionally related isoenzymes *Ulytius edulus*, by J. B. Mitten and R. C. Koehn. *Genetics* **73**, 487–496.

Robbins, R. B. (1918). Some applications of mathematics to breeding problems: II. *Genetics* **3**, 73–92.

Sick, K. (1965). Haemoglobin polymorphism of cod in the Baltic and Danish Belt Sea. *Hereditas* **54**, 19–48.

Sinnock, P. & Sing, C. F. (1972). Analysis of multilocus genetic systems in Tecimsek Michigan. II. Consideration of the correlation between non-alleles in gametes. *American Journal of Human Genetics* **24**, 393–415.

Sved, J. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology* **2**, 125–141.

Wahlund, S. (1928). Zusammensetzung von Populationen und Korrelations-escheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* **11**, 65–106.

Wright, S. (1943). Isolation by distance. *Genetics* **28**, 114–138.