

Decay of linkage disequilibrium in a finite island model

HIDENORI TACHIDA

Department of Biology, Faculty of Science, Kyushu University 33, Fukuoka 812, Japan

(Received 22 March 1994 and in revised form 21 June 1994)

Summary

Time-dependent behaviour of linkage disequilibrium when there was initial linkage disequilibrium is studied in a finite island model assuming neutrality. Explicit expressions for linkage disequilibrium parameters are obtained. From these expressions, the initial and the ultimate decay rates of linkage disequilibrium parameters are found to be increased and decreased, respectively, by finiteness of the population when recombination rate, migration rate and inverse of subpopulation size are of comparable order. Thus, linkage disequilibrium created in the past may persist longer in smaller subdivided populations. Also, differentiation of the gametic parameter of linkage disequilibrium among subpopulations is found to diminish quickly compared to the linkage disequilibrium in the whole population. Implications of these results for the interpretation of linkage disequilibria in natural populations are discussed.

1. Introduction

The coefficient of linkage disequilibrium, D , is a measure of non-random association of alleles at different loci. This quantity can be measured in natural populations and sometimes used as a device to detect selection (e.g. Mukai, Mettler & Chigusa, 1971; Miyashita & Langely, 1988). Consider two neutral loci for which the recombination rate between them is r . The coefficient of linkage disequilibrium decays with a rate of $1-r$ in an infinite random mating population. If the population has a finite size, the rate becomes $1-r-1/(2N)$ as shown by Wright (1933). Thus, the finiteness of the population size increases the rate of decay by $1/(2N)$.

Nei & Li (1973) showed that the ultimate rate of decay in a structured population with two infinite subpopulations is the larger of $1-r$ and $1-m(2-m)$, where m is the migration rate between the two subpopulations. Thus, if the migration rate is smaller than the recombination rate, the decay of linkage disequilibrium is much slower than that in the random mating case. Later this result was extended to the case of the stepping stone model (Feldman & Christiansen, 1975).

In the present paper, I investigate the effect of finite population size on the decay rate of the linkage disequilibrium in the finite island model. I found that when r , m and $1/(2N)$ are of comparable order, the

ultimate decay rate is decreased by the finiteness of the population while the initial rate is increased. The ultimate decay rate determines how quickly linkage disequilibrium diminishes in the population in a long run and thus linkage disequilibrium may persist longer in smaller populations.

2. Model

A finite island model with discrete generations is assumed. There are n subpopulations each with equal size N . In each subpopulation, individuals mate randomly and random union of zygotes (Kimura, 1963; Watterson, 1970) is assumed.

(i) Two-locus gene frequencies and linkage disequilibrium parameters

I consider two loci and call them the first and the second locus, respectively. There are two alleles A , a and B , b at the first and second loci, respectively. One-gene frequency $p_A(p_B)$ defined as probability of a sampled gamete having the $A(B)$ allele at the first (second) locus, is constant through time. I define two-gene frequencies as probabilities of sampled gametes having the A allele at the first locus and the B allele at the second locus. First, I consider linkage disequilibrium in the whole population. In the island model setting, there are three ways in which two genes

at two loci are located. Two genes are on the same gamete, on different gametes in the same subpopulation, or on two gametes in different subpopulations. I designate gene frequencies for the respective arrangements as P_{AB} , $P_{A,B}$, and $P_{A|B}$, respectively. Accordingly, two linkage disequilibrium parameters are defined,

$$D_{AB} = P_{AB} - P_{A,B} \quad (\text{gamete})$$

$$D_{A,B} = P_{A,B} - P_{A|B} \quad (\text{subpopulation}). \quad (1)$$

In the following, the time-dependent behaviour of these parameters is investigated when there is initial linkage disequilibrium. The parameters at generation t are expressed by adding subscript t to them.

(ii) *Changes of gene frequencies*

Alleles are assumed to be neutral with respect to selection (Kimura, 1968). Mutation is assumed to be negligible for the time span considered here. When zygotes are formed, each gamete is sampled from the same subpopulation as its parent or the whole population with a probability $1 - m$ or m , respectively. Here, I define migration rate m in such a way that a migrant is from the same subpopulation with a probability $1/n$ in order to simplify later expressions. Then, transition equations for P s are,

$$\left. \begin{aligned} P_{ABt+1} &= (1-r)P_{ABt} + rP_{A,Bt} \\ P_{A,Bt+1} &= (1-k_1m_*) \left[\frac{P_{ABt}}{2N} + \left(1 - \frac{1}{2N}\right) P_{A,Bt} \right] \\ &\quad + k_1m_* P_{A|Bt}, \\ P_{A|Bt+1} &= \frac{m_*}{n} \left[\frac{P_{ABt}}{2N} + \left(1 - \frac{1}{2N}\right) P_{A,Bt} \right] \\ &\quad + \left(1 - \frac{m_*}{n}\right) P_{A|Bt}, \end{aligned} \right\} \quad (2)$$

where $k_1 = (n-1)/n$ and $m_* = m(2-m)$. From this expression, the transition of the two linkage disequilibrium parameters, D_{AB} and $D_{A,B}$, is expressed as

$$D_{t+1} = TD_t, \quad (3)$$

where

$$D = (D_{AB}, D_{A,B})'$$

and

$$T = \begin{pmatrix} 1-r - \frac{1-k_1m_*}{2N} & k_1m_* \\ \frac{1-m_*}{2N} & 1-m_* \end{pmatrix}$$

I designate a transpose of a matrix or a vector by attaching a prime.

The eigenvalues of the matrix T are

$$\lambda_1 = \frac{C_1 + \sqrt{C_2}}{2}, \quad \lambda_2 = \frac{C_1 - \sqrt{C_2}}{2}, \quad (4)$$

where

$$C_1 = 2-r - \frac{1}{2N} - m_* + \frac{k_1m_*}{2N}$$

$$C_2 = \left(r + \frac{1}{2N} - m_* - \frac{k_1m_*}{2N}\right)^2 + \frac{4k_1m_*(1-m_*)}{2N}.$$

The transition equation (3) can be solved and the resulting expressions for the linkage disequilibrium parameters, D_{ABt} and $D_{A,Bt}$, at generation t using the eigenvalues are

$$D_{ABt} = \frac{1}{\lambda_1 - \lambda_2} \left[(d_1\lambda_1^t - d_2\lambda_2^t) D_{AB0} + \frac{2Nd_1d_2(\lambda_2^t - \lambda_1^t)}{1-m_*} D_{A,B0} \right], \quad (5)$$

$$D_{A,Bt} = \frac{1}{\lambda_1 - \lambda_2} \left[\frac{(1-m_*)(\lambda_1^t - \lambda_2^t)}{2N} D_{AB0} + (d_1\lambda_2^t - d_2\lambda_1^t) D_{A,B0} \right], \quad (6)$$

where

$$d_i = m_* + \lambda_i - 1.$$

When the size of the subpopulations approaches infinity ($N \rightarrow \infty$), the expressions approach those for infinite populations,

$$D_{ABt} = (1-r)^t D_{AB0} + \frac{k_1m_*[(1-r)^t - (1-m_*)^t]}{(m_*-r)} D_{A,B0}, \quad (7)$$

$$D_{A,Bt} = (1-m_*)^t D_{A,B0}, \quad (8)$$

since $\lim_{N \rightarrow \infty} 2Nd_1d_2/(1-m_*) = -k_1m_*$. Whereas $D_{A,Bt} = 0$ if it is initially zero in infinite populations, it is not zero in finite populations since $D_{A,Bt}$ includes the term due to D_{AB0} .

(iii) *Rate of decay*

Rates of decay of the linkage disequilibrium parameters are one minus the eigenvalues of the matrix T , λ_i . Since the system has two eigenvalues, there are two rates of decay, the smaller and the larger. The smaller one ($1 -$ the largest eigenvalue) is called the ultimate rate of decay since the linkage disequilibria decay at this rate in the later phase. I tentatively call the larger one ($1 -$ the smallest eigenvalue) the initial rate of decay. Note that the initial rate defined here does not necessarily reflect the initial decay rate because both eigenvalues contribute to the change of linkage disequilibria in the initial phase. However, as shown later, the difference between the two rates is very large when N is small. In this case, the initial phase is dominated by the smaller eigenvalue unless the initial vector is close to the eigenvector corresponding to the larger eigenvalue and the naming seems relevant. When the subpopulation size N is large, λ_i are written as

$$1-r + O\left(\frac{1}{2N}\right), 1-m_* + O\left(\frac{1}{2N}\right). \quad (9)$$

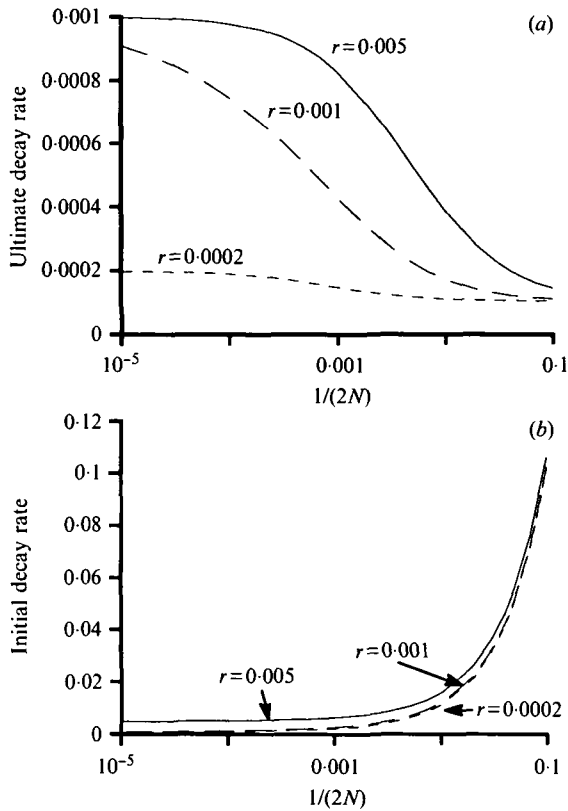


Fig. 1. The ultimate (a) and initial (b) rates of decay of linkage disequilibrium as a function of $1/(2N)$. Other parameters are $m = 0.0005$ and $n = 10$.

When N is infinite, these coincide with those obtained by Nei & Li (1973).

Interesting cases are found when the three parameters, m_* , $1/(2N)$ and r , are comparable and much smaller than one. Note that the diffusion approximation is valid in this range and the following results can be extended to cases with the same relative values among m , $1/(2N)$ and r by an appropriate time scaling (Crow & Kimura, 1970). The ultimate rate of decay of linkage disequilibrium as a function of N is shown in Fig. 1a for several values of m_* , r and n . As the subpopulation size decreases, the ultimate decay rate becomes smaller especially when $1/(2N)$ becomes larger than $m/10$ or $r/10$. In other words, the ultimate decay of linkage disequilibrium becomes slower. For example, when $N = 50$, $r = 0.001$, $n = 10$ and $m_* = 0.0005$, the ultimate decay rate is 0.000169 while it is 0.001 when N is infinite. Thus, the ultimate rate of decay is decreased six-folds in the finite population in this example. The initial rate of decay (1 – the smaller eigenvalue) of linkage disequilibrium is shown in Fig. 1b. The initial rate starts to increase from that in the infinite populations when $1/(2N)$ becomes larger than 0.01 in this example. If m and r are much smaller than $1/(2N)$, the rate is approximately $1/(2N)$. Until $1/(2N)$ becomes much larger than m and r , the rate does not change much as $1/(2N)$ increases. This contrasts with the situation in the ultimate rate. Also the effect of changing r on the initial rate is small.

(iv) Two-gene frequencies

Two-gene frequencies are computed from the linkage disequilibrium parameters since the following sum is constant through time:

$$C = P_{AB} + 2NrP_{A.B} + 2Nr(n-1)P_{A|B}. \tag{10}$$

Using this constant and linkage disequilibrium parameters, the two-gene frequencies are

$$P_{ABt} = \frac{2Nr}{2Nrn+1} [nD_{ABt} + (n-1)D_{A.Bt} + C], \tag{11}$$

$$P_{A.Bt} = \frac{1}{2Nrn+1} \{D_{ABt} + 2Nr[(n-1)D_{A.Bt} + C]\}, \tag{12}$$

$$P_{A|Bt} = \frac{1}{2Nrn+1} [D_{ABt} - (2Nr+1)D_{A.Bt} + 2NrC]. \tag{13}$$

In the equilibrium state, the population becomes monomorphic and the final two-gene frequencies are all equal.

$$P_{AB\infty} = P_{A.B\infty} = P_{A|B\infty} = \frac{2NrC}{2Nrn+1}. \tag{14}$$

This is the fixation probability of the AB gametes when the initial condition is given by P_{AB0} , $P_{A.B0}$, $P_{A|B0}$ (Kimura, 1963).

(v) Linkage disequilibria in each subpopulation

Only the whole population was considered thus far. However, sometimes the frequencies in each subpopulation are of interest. Let p_A^i (p_B^i) be the probability that a gamete sampled from the i th subpopulation has the allele A (B) at the first (second) locus. Since neutrality of alleles is assumed, the average, $p_A = \sum_i p_A^i/n$, does not change through time. The change of p_A^i is described by

$$p_{A_t}^i = (1-m)p_{A_t}^i + mp_A, \tag{15}$$

whose solution is

$$p_{A_t}^i = (1-m)^t (p_{A_0}^i - p_A) + p_A. \tag{16}$$

For two genes at different loci there are again three types of gene arrangement, i.e. two genes on the same gamete, on different gametes in the same subpopulation and on different gametes in different subpopulations, but now the two-gene frequencies have superscripts to indicate from which subpopulations the two genes are sampled. Thus, P_{AB}^i and $P_{A.B}^i$ are the probabilities that two genes on the same gamete, and two genes on different gametes from the subpopulation i are A and B, respectively. $P_{A|B}^{ij}$ is the probability that a gene at the first locus sampled from subpopulation i and another gene at the second locus sampled from subpopulation j ($i \neq j$) are A and B, respectively. Define $P_{A|B}^i$ as

$$P_{A|B}^i = \frac{1}{2(n-1)} \sum_{j \neq i} (P_{A|B}^{ij} + P_{A|B}^{ji}). \tag{17}$$

Table 1. Initial gene frequencies in the examples

Example	P_{AB0}	$P_{A.B0}$	$P_{A B0}$	P_{AB}^i	$P_{A.B0}^i$	$P_{A B0}^i$
1 ($1 \leq i \leq 5$)	0.5	0.5	2/9	1	1	4/9
($6 \leq i \leq 10$)	—	—	—	0	0	0
2 ($1 \leq i \leq 5$)	3/4	2499/4000	5/9	1	1	13/18
($6 \leq i \leq 10$)	—	—	—	1/2	499/2000	7/18

two-gene frequencies from those in the whole population as

$$F_{AB}^i = P_{AB}^i - P_{AB} \tag{32}$$

$$F_{A.B}^i = P_{A.B}^i - P_{A.B} \tag{33}$$

$$F_{A|B}^i = P_{A|B}^i - P_{A|B} \tag{34}$$

F_{AB}^i and $F_{A.B}^i$ represent the gametic and sub-population differentiations and correspond to ${}_g\mathcal{F}_i$ and ${}_d\mathcal{F}_i$, respectively, of Tachida & Cockerham (1986). Furthermore, the sums of squares of F_{AB}^i and $F_{A.B}^i$ over subpopulations correspond approximately to D'_{IS} and D_{ST} , respectively, in Ohta (1982a). Incidentally, the sum of squares of D_{AB}^i and the square of D_{AB} correspond approximately to D_{IS} and D'_{ST} , respectively, in Ohta (1982a).

The transition equations for F s are given in (22) and thus the largest eigenvalue for F is $1 - m$. Therefore, the magnitudes of F relative to those of D in the later stage depend on the relative values of $1 - m$ to λ_1 , the largest eigenvalue for D . As shown in Fig. 1, if r is smaller than $m/2$ or if N is small, $1 - m$ is smaller than λ_1 . In these cases, F parameters decay faster than D parameters and the F s are expected to be smaller than the D s in the later stage.

(vii) Examples

A few examples will illustrate the behaviour of the gametic parameter of linkage disequilibrium and the differentiation of two-gene frequencies. In the following examples, parameters assumed are $n = 10$, $N = 500$, $m = 0.001$, $r = 0.001$. Thus, the eigenvalues are

$$\lambda_1 = 0.999341, \quad \lambda_2 = 0.996661, \quad \eta_1 = 0.998525$$

and $\eta_2 = 0.996479$.

In the first example, the frequencies of the AB gametes are one in the first five subpopulations and zero in the remaining five subpopulations initially. The initial frequencies are shown in Table 1. The gametic parameter of linkage disequilibrium within a subpopulation is

$$D_{ABi}^i = 0.186479\lambda_1^i - 0.186479\lambda_2^i.$$

This is plotted in Fig. 2. In this case, there is no differentiation of linkage disequilibrium. The linkage disequilibrium is initially zero, first increases and then decreases to zero again. As this example shows, if the

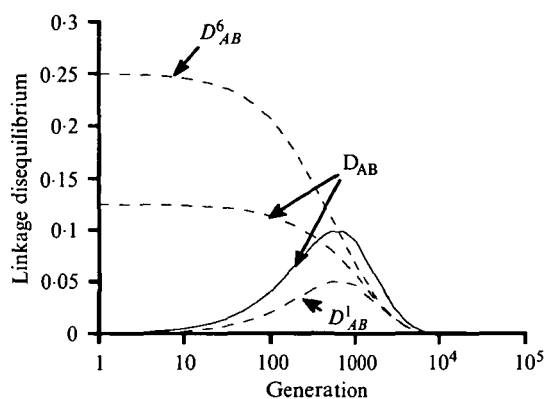


Fig. 2. Change of the gametic parameters of linkage disequilibrium. Two initial gene frequencies (example 1 (—) and 2, (---), see the text) are used. In the example 2, three parameters, D_{AB} , D_{AB}^i representing those of subpopulation 1–5 and D_{AB}^6 representing those of subpopulation 6–10 are different and they are shown. Parameters are $m = 0.001$, $r = 0.001$, $1/(2N) = 0.001$, $n = 10$.

initial values of the linkage disequilibrium parameters are zero, the absolute values of the coefficients of the two terms due to the two eigenvalues are the same but the signs are different. The increase of the disequilibrium is caused by the different rates of decrease in the two terms due to the eigenvalues. Thus, how big the gametic disequilibrium becomes depends on the relative magnitude of the two eigenvalues. As shown in Fig. 1, the finiteness of the subpopulation size broadens the difference of the two eigenvalues. Thus, as the size becomes smaller, the maximum gametic disequilibrium achieved increases, especially when the size becomes very small ($N < 50$).

In the second example, the frequencies of the AB gametes are one in the first five subpopulations and one-half in the remaining five subpopulations initially. The rest of the gametes in the last five subpopulations do not have the A or B gene. As in the first example, initial frequencies are tabulated in Table 1. The gametic parameter of linkage disequilibrium in the i th subpopulation is

$$D_{ABi}^i = -0.0319765\eta_1^i - 0.0932735\eta_2^i + 0.109096\lambda_1^i + 0.0161544\lambda_2^i \quad (1 \leq i \leq 5),$$

$$D_{ABi}^i = 0.0319765\eta_1^i + 0.0932735\eta_2^i + 0.109096\lambda_1^i + 0.0161544\lambda_2^i \quad (6 \leq i \leq 10).$$

These are plotted in Fig. 2 with D_{ABi} . Initially, the gametic linkage disequilibrium is zero in the first five

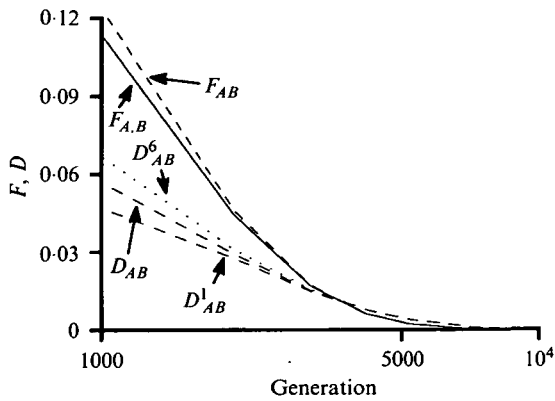


Fig. 3. Change of D and F_s in the later stage. The parameters and the initial condition are the same as those of the example 2 in Fig. 2. The absolute values of F_s are the same among subpopulations and they are plotted without a subscript to indicate the subpopulation.

subpopulations and 0.2505 in the last five subpopulations. There is differentiation of the linkage disequilibrium initially. But it diminishes as time passes and the linkage disequilibrium within each subpopulation becomes equal. The final rate of approach to zero is λ_1 .

Finally, the relative magnitudes of F_s and D_s are compared in the later stage (1000th to 10000th generations) in the second example (Fig. 3). Although F_s are larger until up to 4000 generations, the relationship is reversed after that. In this example, $1 - m = 0.999$ is smaller than $\lambda_1 = 0.999341$ and this is why F_s become smaller than D_s ultimately.

3. Discussion

Linkage disequilibria are generated by various causes such as admixture of populations, bottlenecks and selection. If the markers in consideration are neutral, linkage equilibrium is attained in time. How quickly linkage equilibrium is attained depends upon the recombination rate and the mating structure of the population. In the present paper, I investigated the time-dependent behaviour of linkage disequilibrium in a finite island model. I found that the ultimate decay of linkage disequilibrium is slowed compared to that in an infinite population and the initial rate is quickened when r , m and $1/(2N)$ of comparable order. This implies that in a population where linkage disequilibrium was suspected to be generated in the past and these parameters are of comparable order, it is likely that linkage disequilibrium would be observed in samples taken later. The linkage disequilibrium will persist for generations of the order of the inverse of these parameters. Since structures of natural populations are changing in a long time scale, only likely cause of generating linkage disequilibrium is admixture of subpopulations which had been isolated for a long time. In such situations, linkage disequilibrium which is a remnant of the initial linkage

disequilibrium caused by the past population admixture may be still persisting now. Nei & Li (1973) previously pointed out that the decay of linkage disequilibrium is retarded in a geographically structured population but the present study shows that this is more so when the population size is finite. The acceleration in the initial phase and the retardation in the later phase due to the finite size is sometimes significant as seen in Fig. 1. This is in contrast to the situation in a random mating population where the decay of linkage disequilibrium is just accelerated by the finiteness of the population size (Wright 1933).

As noted above, significant retardation in the later phase of the decay is found when r , m and $1/(2N)$ are of comparable order. For allozyme markers, recombination rates are usually larger than 0.01. Thus, we expect to see the effect of finite population size on the decay of linkage disequilibrium if the subpopulation size is of the order of one hundred or smaller. Retardation in the later phase would be observed if the generation of the linkage disequilibrium occurred one hundred generations ago or so. Such situations might be abundant in recent human evolution, especially in aboriginal populations (Smouse, Neel & Liu, 1986). Even in non-aboriginal populations, linkage disequilibria are sometimes found for two loci with this magnitude of recombination or less. For example, Hastbacka *et al.* (1992) observed significant linkage disequilibrium between a disease locus and a polymorphic marker in Finland. The authors consider that the Finland population was founded about 2000 years ago and has been isolated with little migration. The population structure was thought to have only a small influence on the decay of linkage disequilibrium in this case since both the disease allele and the marker allele associated with it are rare. However, if loci are polymorphic, the retardation of the decay considered in this paper may occur in populations with such a structure.

For DNA polymorphisms, the recombination rate is very small among sites. For example, one centimorgan corresponds to about 570 kb on the average in *Drosophila* (p. 113 of Fincham 1983). Thus, the recombination rate between sites 1 kb apart is about 1.7×10^{-5} . For such small recombination rates, we expect to observe larger linkage disequilibria in larger populations with sizes of the order of 10^5 and small migration rates of the order of 10^{-5} . Indeed, Langley, Montgomery & Quattlebaum (1982) and Miyashita & Langley (1988) found significant linkage disequilibria among sites several kilobases apart and these might be remnants of the initial linkage disequilibria generated more than 10^5 generations ago.

Recently, Miyashita, Aguade & Langley (1993) and Schaeffer & Miller (1993) conducted large scale surveys of linkage disequilibria in *Drosophila* populations and found large linkage disequilibria concentrated in restricted regions of genes. They suggest that selection rather than the population structure is responsible for

the large linkage disequilibria based on the following observations and Ohta (1982a)'s suggestion of discriminating effects of selection from those of the population structure by comparing her statistics, D_{IS} and D_{ST} : 1. The heterogeneity of D_{AB} among subpopulations was not found. 2. The differentiation of two-gene frequencies, D'_{IS} or D'_{ST} , is smaller than the linkage disequilibrium parameter, D_{IS} or D_{ST} . However, the heterogeneity (differentiation) of D_{AB} among subpopulations diminishes quickly for small r , m , $1/(2N)$. Furthermore, if r or N is small, F which approximately corresponds to D'_{IS} or D'_{ST} decays faster than D which approximately corresponds to D_{IS} or D_{ST} as shown in the present study in the non-equilibrium state (see Fig. 1). Although the population structure hypothesis does not explain why the large linkage disequilibria are concentrated in restricted regions of the genes, the hypothesis cannot be rejected based just on the two observations mentioned above if the *Drosophila* populations are not in the equilibrium state. Of course, since we need to assume the non-equilibrium state and also some specific initial configuration to explain the pattern, the population structure hypothesis may not be a parsimonious one. However, a recent study suggests that the population of *Drosophila melanogaster* which is the species analysed by Miyashita, Aguade & Langley (1993) is not in the equilibrium state (Begun & Aquadro, 1993) and the first assumption is satisfied in this species.

Now population data of simple repeated sequences such as microsatellites and minisatellites are accumulating rapidly (e.g. Valdes, Slatkin & Freimer, 1993; Shriver *et al.* 1993). Thus far, only the minisatellite data collected for DNA fingerprinting have been analysed with regard to disequilibrium between loci (Weir, 1992). Since the loci analysed there are not linked or loosely linked, the population structure will not affect the decay of linkage disequilibrium in these loci. However, since microsatellites and minisatellites are found in various positions in genomes, linkage disequilibrium between loci which are separated by various map units can be measured. If we choose a pair of such loci with the recombination rate comparable to the migration rate and the inverse of the size of the population, we may observe linkage disequilibrium which is a remnant of the past history of the population.

In the present formulation, the initial gene frequencies are given. Sometimes founder populations are established by sampling gametes from a base population. In such cases, the initial parameters of linkage disequilibrium are random variables and the expectation of the latter can be computed (Cockerham & Weir, 1973). Since linkage disequilibrium parameters at generation t are expressed as linear combinations of initial linkage disequilibrium parameters as shown in (5), (6) and (30), the expectations of the parameters at t are computed by putting the expected initial values into these equations. If the

initial sampling is made randomly, linkage disequilibrium parameters are non-zero only when there is linkage disequilibrium in the base population.

In the present paper, I investigated the behaviour of linkage disequilibrium parameters. These parameters are estimated from samples and the estimators have variances. In finite populations, the variances of these estimators are non-zero even if the parameters themselves are zero (Ohta & Kimura, 1969; Hill & Robertson 1968; Weir & Cockerham, 1969). The variance becomes larger in geographically structured populations in the mutation-drift equilibrium state where the linkage disequilibrium parameters are zero (Ohta, 1982a, b; Tachida & Cockerham, 1986). Thus, large variances of estimators of the linkage disequilibrium parameters are expected in the present situation. In theory, I can compute these variances deriving transition equations for gene frequencies as was done in the present paper. But the number of gene frequencies necessary for the transition equations to be closed becomes very large and this approach seems impractical. Different approaches such as computer simulations are required to evaluate these variances and this would be a next step toward understanding the decay of linkage disequilibrium in structured populations.

I thank Dr C. C. Cockerham and three anonymous reviewers for constructive comments on an earlier draft of the manuscript. This research was partially supported by grant-in-aid from the Ministry of Education, Science and Culture of Japan and a grant from the Society for the Promotion of Genetics.

References

- Begun, D. J. & Aquadro, C. F. (1993). African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**, 548–550.
- Cockerham, C. C. & Weir, B. S. (1973). Descent measure for two loci with some applications. *Theoretical Population Biology* **4**, 300–330.
- Crow, J. F. & Kimura, M. (1970). *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- Feldman, M. W. & Christiansen, F. B. (1975). The effect of population subdivision on two loci without selection. *Genetical Research* **24**, 151–162.
- Fincham, J. R. S. (1983). *Genetics*. Jones and Bartlett Publishers, Boston.
- Hastbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A. & Lander, E. (1992). Linkage disequilibrium mapping in isolated founder populations: disastrophic dysplasia in Finland. *Nature Genetics* **2**, 204–211.
- Hill, W. G. & Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226–231.
- Kimura, M. (1963). A probability method for treating inbreeding systems especially with linked genes. *Biometrics* **19**, 1–17.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* **217**, 624–626.
- Langley, C. H., Montgomery, E. A. & Quattlebaum, W. F. (1982). Restriction map variation in the Adh region of *Drosophila*. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 5631–5635.

- Miyashita, N. & Langley, C. H. (1988). Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. *Genetics* **120**, 199–212.
- Miyashita, N. T., Aguade, M. & Langley, C. H. (1993). Linkage disequilibrium in the *white* region of *Drosophila melanogaster*. *Genetical Research* **62**, 101–109.
- Mukai, T., Mettler, L. E. & Chigusa, S. I. (1971). Linkage disequilibrium in a local population of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, U.S.A.* **68**, 1065–1069.
- Nei, M. & Li, W.-H. (1973). Linkage disequilibrium in subdivided populations. *Genetics* **75**, 213–219.
- Ohta, T. (1982*a*). Linkage disequilibrium with the island model. *Genetics* **101**, 139–155.
- Ohta, T. (1982*b*). Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 1940–1944.
- Ohta, T. & Kimura, M. (1969). Linkage disequilibrium due to random genetic drift. *Genetical Research* **13**, 47–55.
- Schaeffer, S. W. & Miller, E. L. (1993). Estimates of linkage disequilibrium and the recombination parameter determined from segregating nucleotide sites in the alcohol dehydrogenase region of *Drosophila pseudoobscura*. *Genetics* **135**, 541–552.
- Shriver, M. D., Jin, L., Chakraborty, R. & Boerwinkle, E. (1993). VNTR allele frequency distributions under stepwise mutation model: A computer simulation approach. *Genetics* **134**, 983–993.
- Smouse, P. E., Neel, J. V. & Liu, W. (1983). Multiple-locus departures from panmictic equilibrium within and between village gene pools of amerindian tribes at different stages of agglomeration. *Genetics* **104**, 133–153.
- Tachida, H. & Cockerham, C. C. (1986). Analysis of linkage disequilibrium in an island model. *Theoretical Population Biology* **29**, 161–197.
- Watterson, G. A. (1970). The effect of linkage in a finite random-mating population. *Theoretical Population Biology* **1**, 72–87.
- Valdes, A. M., Slatkin, M. & Freimer, N. B. (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**, 737–749.
- Weir, B. S. (1992). Independence of VNTR alleles defined as fixed bins. *Genetics* **130**, 873–887.
- Weir, B. S. & Cockerham, C. C. (1969). Group inbreeding with two linked loci. *Genetics* **63**, 711–742.
- Weir, B. S. & Cockerham, C. C. (1974). Behavior of pairs of loci in finite monoecious populations. *Theoretical Population Biology* **6**, 323–354.
- Wright, S. (1933). Inbreeding and recombination. *Proceedings of the National Academy of Sciences, U.S.A.* **19**, 420–433.