

# LINKAGE DISEQUILIBRIUM AT STEADY STATE DETERMINED BY RANDOM GENETIC DRIFT AND RECURRENT MUTATION<sup>1</sup>

TOMOKO OHTA AND MOTOO KIMURA

*National Institute of Genetics, Mishima, Japan*

Received January 28, 1969

**I**N order to understand the genetic architecture of Mendelian populations, we have to know, in addition to gene frequencies, the degree of linkage disequilibrium between loci, namely, the extent to which combinations of genes at linked loci deviate from randomness.

In panmictic populations, two factors are mainly responsible for the production of linkage disequilibrium. They are epistatic interaction in fitness and random sampling of gametes in reproduction.

Although a number of papers have been published since the work of KIMURA (1956), treating the problem of epistasis and linkage in an infinite population, it was only last year that the first systematic treatment of linkage disequilibrium due to random drift was presented by HILL and ROBERTSON (1968). Using the method of moment-generating matrix, they obtained the variance of the linkage disequilibrium coefficient as a function of time and initial gametic frequencies for the case of no recombination, assuming that mutation and selection are absent. In addition, they studied numerically several cases of recombination by multiplying the moment-generating matrix and also by carrying out simulation experiments. They demonstrated that significant linkage disequilibrium may result from random drift under tight linkage and small population number. Their results were extended by OHTA and KIMURA (1969) who obtained, by the method of Kolmogorov backward equation, the formula for the variance of linkage disequilibrium in which recombination is incorporated.

In natural populations, however, it is expected that random drift and recurrent mutation balance each other so that a steady state is reached with respect to linkage disequilibrium. SVED (1968) studied this problem assuming that all gene frequencies are held at 50% by strong overdominance while mutation is so rare as to be negligible. Presumably, his treatment is applicable to a certain transient state but not necessarily to the equilibrium state.

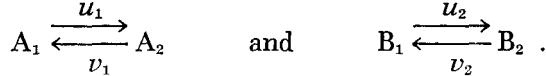
In the present paper we intend to present a theoretical foundation for the treatment of linkage disequilibrium at steady state determined by random drift and mutation (or more generally, linear evolutionary pressure). The problem of linkage disequilibrium will become particularly important when we consider two neighboring nucleotide sites within a cistron for which the recombination fraction may be much smaller than the reciprocal of the population number. It will also

<sup>1</sup> Contribution No. 711 from the National Institute of Genetics, Mishima, Shizuoka-ken, Japan.

be important in relation to the occurrence of pseudo-overdominance due to linkage disequilibrium.

#### BASIC THEORY

Let us consider two linked loci and assume that a pair of alleles,  $A_1$  and  $A_2$ , are segregating in the first locus, and  $B_1$  and  $B_2$  in the second. Let the mutation rates be as follows:



We will denote by  $N_e$  the effective population number and by  $c$  the recombination fraction between the 2 loci. Though we assume no selection, the treatment in the present paper may be extended to cover some cases of selection when selection pressure is approximated by a linear function of gene frequencies.

We shall derive the formula for the variance of linkage disequilibrium at steady state first by using the method of moment-generating matrix as developed by HILL and ROBERTSON (1968) and then, by a more convenient method based on diffusion models.

*Treatment by the method of moment-generating matrix:* Let us denote the frequencies of 4 types of gametes,  $A_1B_1$ ,  $A_2B_1$ ,  $A_1B_2$  and  $A_2B_2$  by  $P_1$ ,  $P_2$ ,  $P_3$  and  $P_4$ , respectively ( $P_1+P_2+P_3+P_4=1$ ). Let  $p=P_1+P_3$ ,  $q=P_1+P_2$  and  $D=P_1P_4-P_2P_3$ , where  $p$  is the frequency of  $A_1$ ,  $q$  is the frequency of  $B_1$ , and  $D$  is the coefficient of linkage disequilibrium. The change of these quantities in one generation consists of changes due to mutation, crossing over and random sampling of gametes. Let  $p_m$ ,  $q_m$  and  $D_m$  be the gene frequencies and linkage disequilibrium after mutation. Then the amounts of change in these quantities, neglecting higher order terms of  $u$  and  $v$ , are

$$\begin{aligned} p_m - p &= v_1 - (u_1 + v_1)p \\ q_m - q &= v_2 - (u_2 + v_2)q \\ D_m - D &= -(u_1 + u_2 + v_1 + v_2)D \end{aligned} \quad (1)$$

Now, let us consider the following 3 quantities;

$$\begin{aligned} X &= pq(1-p)(1-q) \\ Y &= D(1-2p)(1-2q) \\ Z &= D^2 \end{aligned} \quad (2)$$

We will denote the values of these quantities after mutation by  $X_m$ ,  $Y_m$  and  $Z_m$ . Then, neglecting higher order terms involving  $u$  and  $v$ , the amount of their change by mutation can be shown to be as follows;

$$\begin{aligned} X_m - X &= -2kX + v_1q(1-q) + v_2p(1-p) + (u_1 - v_1)pq(1-q) \\ &\quad + (u_2 - v_2)pq(1-p), \\ Y_m - Y &= -2kY + (u_1 - v_1)D(1-2q) + (u_2 - v_2)D(1-2p), \\ Z_m - Z &= -2kZ, \end{aligned} \quad (3)$$

where

$$k = u_1 + v_1 + u_2 + v_2. \quad (4)$$

We then consider the change of  $X$ ,  $Y$  and  $Z$  by crossing over and sampling.

Under the assumption that the moments do not change from one generation to the next, the following relationship holds for the expected values of  $X$ ,  $Y$  and  $Z$  at steady state;

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} E(X_m) \\ E(Y_m) \\ E(Z_m) \end{bmatrix} = \begin{bmatrix} E(X) \\ E(Y) \\ E(Z) \end{bmatrix} \quad (5)$$

where  $E$  denotes the operator of taking expectation and  $[a_{ij}]$  is the moment-generating matrix of HILL and ROBERTSON (1968), namely

$$\begin{aligned} a_{11} &= (1 - \frac{1}{n})^2, a_{12} = \frac{1}{n} (1 - \frac{1}{n})^2 (1 - c), a_{13} = \frac{2}{n^2} (1 - \frac{1}{n}) (1 - c)^2 \\ a_{21} &= 0, a_{22} = (1 - \frac{1}{n}) (1 - \frac{2}{n})^2 (1 - c), a_{23} = \frac{4}{n} (1 - \frac{1}{n}) (1 - \frac{2}{n}) (1 - c)^2 \\ a_{31} &= \frac{1}{n} (1 - \frac{1}{n}), a_{32} = \frac{1}{n} (1 - \frac{1}{n})^2 (1 - c), a_{33} = (1 - \frac{1}{n}) (\frac{1}{n^2} + (1 - \frac{1}{n})^2) (1 - c)^2 \end{aligned} \quad (6)$$

in which  $n = 2N_e$ .

By solving equation (5), we obtain

$$\begin{aligned} E(X) &= \frac{A}{B} \begin{vmatrix} -a_{11} & a_{21}(1-2k) & a_{31}(1-2k) \\ 0 & a_{22}(1-2k)-1 & a_{32}(1-2k) \\ -a_{13} & a_{23}(1-2k) & a_{33}(1-2k)-1 \end{vmatrix} \\ E(Y) &= \frac{A}{B} \begin{vmatrix} a_{11}(1-2k)-1 & -a_{11} & a_{31}(1-2k) \\ 0 & 0 & a_{32}(1-2k) \\ a_{13}(1-2k) & -a_{13} & a_{33}(1-2k)-1 \end{vmatrix} \\ E(Z) &= \frac{A}{B} \begin{vmatrix} a_{11}(1-2k)-1 & a_{21}(1-2k) & -a_{11} \\ 0 & a_{22}(1-2k)-1 & 0 \\ a_{13}(1-2k) & a_{23}(1-2k) & -a_{13} \end{vmatrix} \end{aligned} \quad (7)$$

where

$$A = E\{v_1 q(1-q) + v_2 p(1-p) + (u_1 - v_1) pq(1-q) + (u_2 - v_2) pq(1-p)\} \quad (8)$$

and

$$B = \begin{vmatrix} a_{11}(1-2k)-1 & a_{21}(1-2k) & a_{31}(1-2k) \\ 0 & a_{22}(1-2k)-1 & a_{32}(1-2k) \\ a_{13}(1-2k) & a_{23}(1-2k) & a_{33}(1-2k)-1 \end{vmatrix} \quad (9)$$

The value of  $A$  is determined by the lower moments and it can be shown that

$$\begin{aligned} A &= v_2 \left( \frac{v_1}{u_1 + v_1} \right) \left( \frac{u_2}{u_2 + v_2} \right) \left( \frac{8N_e u_1}{4N_e(u_1 + v_1) + 1} \right) \\ &+ v_1 \left( \frac{u_1}{u_1 + v_1} \right) \left( \frac{v_2}{u_2 + v_2} \right) \left( \frac{8N_e u_2}{4N_e(u_2 + v_2) + 1} \right). \end{aligned} \quad (10)$$

*Treatment by the diffusion models:* Under the assumption of reversible mutation, the stationary distributions of  $p$  and  $q$  can be obtained separately by using WRIGHT's formula for the gene frequency distribution (WRIGHT 1938), but not their joint distribution, unless recombination between the 2 loci is frequent enough to make their combination independent. However, the moments of distribution involving  $p$ ,  $q$  and  $D$  at steady state can be obtained by the following method based on the diffusion models.

As demonstrated in the APPENDIX for the special cases of one and two independent random variables, we have in general, at stationary state

$$E\{L_B(f)\} = 0 \quad (11)$$

In this equation,  $L_B$  denotes the differential operator such that  $\partial\phi/\partial t = L_B(\phi)$  represents the Kolmogorov backward equation (cf. KIMURA 1964), and  $f$  is a continuous function of independent random variables. For two linked loci, the appropriate Kolmogorov backward equation was given by OHTA and KIMURA (1969). For the present model, by taking account of mutation, but neglecting selection, (11) becomes,

$$\begin{aligned} E \left\{ \frac{p(1-p)}{4} \frac{\partial^2 f}{\partial p^2} + \frac{q(1-q)}{4} \frac{\partial^2 f}{\partial q^2} + \frac{D}{2} \frac{\partial^2 f}{\partial p \partial q} + \frac{D(1-2p)}{2} \frac{\partial^2 f}{\partial p \partial D} \right. \\ \left. + \frac{D(1-2q)}{2} \frac{\partial^2 f}{\partial q \partial D} + \frac{1}{4} (pq(1-p)(1-q) + D(1-2p)(1-2q) - D^2) \frac{\partial^2 f}{\partial D^2} \right. \\ \left. + N_e(v_1 - (u_1 + v_1)p) \frac{\partial f}{\partial p} + N_e(v_2 - (u_2 + v_2)q) \frac{\partial f}{\partial q} \right. \\ \left. - D \left( \frac{1}{2} + N_e c' \right) \frac{\partial f}{\partial D} \right\} = 0 \quad (12) \end{aligned}$$

where  $c' = c + k$ .

Now, in order to obtain the expected values of 3 quantities,  $X = pq(1-p)(1-q)$ ,  $Y = D(1-2p)(1-2q)$  and  $Z = D^2$ , we first let  $f = pq(1-p)(1-q)$  in the above equation. Then,

$$\begin{aligned} E \{ -(1 + 2N_e k)X + (1/2)Y + N_e v_1 q(1-q) + N_e v_2 p(1-p) \\ + N_e(u_1 - v_1)pq(1-q) + N_e(u_2 - v_2)pq(1-p) \} = 0 \quad (13) \end{aligned}$$

Next, let  $f = D(1-2p)(1-2q)$ , then

$$\begin{aligned} E \{ 2Z - (5/2 + N_e c' + N_e k)Y + N_e(u_2 - v_2)D(1-2p) \\ + N_e(u_1 - v_1)D(1-2q) \} = 0 \quad (14) \end{aligned}$$

Finally, let  $f = D^2$ , then

$$E \{ (1/2)X + (1/2)Y - (3/2 + 2N_e c')Z \} = 0 \quad (15)$$

Thus we obtain the following simultaneous equation for the expected values of  $X$ ,  $Y$  and  $Z$ :

$$\begin{aligned} (1 + 2N_e k)E(X) - (1/2)E(Y) &= E \{ N_e v_1 q(1-q) + N_e v_2 p(1-p) \\ &\quad + N_e(u_1 - v_1)pq(1-q) + N_e(u_2 - v_2)pq(1-p) \} \equiv N_e A \\ (5/2 + N_e c' + N_e k)E(Y) - 2E(Z) &= E \{ N_e(u_2 - v_2)D(1-2p) \\ &\quad + N_e(u_1 - v_1)D(1-2q) \} = 0 \\ (1/2)E(X) + (1/2)E(Y) - (3/2 + 2N_e c')E(Z) &= 0 \quad (16) \end{aligned}$$

Note that this relation is equivalent to equation (5). By solving equation (16), we get

$$E(X) = \frac{N_e A}{1 + 2N_e k} \left\{ 1 + \frac{1}{(1+2N_e k)(3+4N_e c')(2.5+N_e c'+N_e k)-2(1+2N_e k)-1} \right\}$$

$$E(Y) = \frac{2N_e A}{(1+2N_e k)(3+4N_e c')(2.5+N_e c'+N_e k)-2(1+2N_e k)-1} \quad (17)$$

$$E(Z) = \frac{(2.5+N_e c'+N_e k)N_e A}{(1+2N_e k)(3+4N_e c')(2.5+N_e c'+N_e k)-2(1+2N_e k)-1}$$

in which  $A$  is given by (10). As shown in (8),  $A$  contains lower moments such as  $p(1-p)$ ,  $q(1-q)$  etc., and their expected values may be obtained by putting  $f = p(1-p)$ ,  $f = q(1-q)$  etc. in the basic equation (12). We note here that  $E(Z)$  of (7) and (17) gives the variance of linkage disequilibrium at steady state, since  $E\{D\} = 0$  in the present case.

#### DISCUSSION

The above formulae can be applied to the cases in which genetic variability is maintained by recurrent mutation in a finite population. Table 1 gives numerical examples computed by using (7) and (17). There are slight differences between the two sets of values. They are due to the approximation involved in (17) based on diffusion models, but the agreement is satisfactory.

Since  $E(Z)$  depends not only on the magnitude of linkage disequilibrium but also very much on gene frequencies, it may be convenient to consider the quantity,  $\sigma_d^2 = E\{D^2\}/E\{pq(1-p)(1-q)\}$  as a measure of linkage disequilibrium which is independent of gene frequencies. We shall call the square root of this quantity, i.e.,  $\sigma_d = [E\{D^2\}/E\{pq(1-p)(1-q)\}]^{1/2}$ , the *standard linkage deviation*.  $\sigma_d^2$  is slightly different from the square of the correlation of gene frequencies considered by HILL and ROBERTSON (1968) who designated it as  $r^2$  which is equal to  $E\{D^2/pq(1-p)(1-q)\}$ . However, these 2 quantities are very close to each other in their actual values for the cases so far checked as pointed out by OHTA and KIMURA (1969). Thus, the standard linkage deviation may be understood as approximately equal to the correlation of gene frequencies between the 2 linked loci.

From the set of formulas given in (17)  $\sigma_d^2$  is obtained as  $E(Z)/E(X)$ , and, we have

$$\sigma_d^2 = \frac{1}{3+4N_e c' - [2/(2.5+N_e c'+N_e k)]} \quad (18)$$

where  $c' = c + k$ .

Since  $k$  is usually much smaller than  $c$ ,  $\sigma_d^2$  is not much influenced by the mutation rate but depends almost entirely on  $N_e c'$ . Now, let us compare  $\sigma_d^2$  in stationary state with that in transient state. HILL and ROBERTSON (1968) and OHTA and

TABLE 1

Comparison of the values of  $E(X)$ ,  $E(Y)$ ,  $E(Z)$  and  $\sigma_d^2$  obtained by using two different formulae, (7) and (17), assuming  $N_e = 50$

Mutation rate	c	Formula used	$E(X)$	$E(Y)$	$E(Z)$	$\sigma_d^2$	
$u_1 = u_2 = v_1 = v_2 = 1 \times 10^{-4}$	0.0	7	$0.1093 \times 10^{-3}$	$0.3750 \times 10^{-4}$	$0.4832 \times 10^{-4}$	0.4421	
		17	$0.1107 \times 10^{-3}$	$0.3804 \times 10^{-4}$	$0.4831 \times 10^{-4}$	0.4364	
	0.01	7	$0.9816 \times 10^{-4}$	$0.1455 \times 10^{-4}$	$0.2277 \times 10^{-4}$	0.2320	
		17	$0.9958 \times 10^{-4}$	$0.1481 \times 10^{-4}$	$0.2252 \times 10^{-4}$	0.2261	
	0.02	7	$0.9504 \times 10^{-4}$	$0.8145 \times 10^{-5}$	$0.1508 \times 10^{-4}$	0.1587	
		17	$0.9648 \times 10^{-4}$	$0.8367 \times 10^{-5}$	$0.1481 \times 10^{-4}$	0.1535	
	0.05	7	$0.9243 \times 10^{-4}$	$0.2757 \times 10^{-5}$	$0.7677 \times 10^{-5}$	0.0831	
		17	$0.9387 \times 10^{-4}$	$0.2937 \times 10^{-5}$	$0.7401 \times 10^{-5}$	0.0788	
	$u_1 = u_2 = v_1 = v_2 = 1 \times 10^{-3}$	0.0	7	$0.5487 \times 10^{-2}$	$0.1218 \times 10^{-2}$	$0.1797 \times 10^{-2}$	0.3275
			17	$0.5541 \pm 10^{-2}$	$0.1228 \times 10^{-2}$	$0.1781 \times 10^{-2}$	0.3215
		0.01	7	$0.5260 \times 10^{-2}$	$0.5923 \times 10^{-3}$	$0.1039 \times 10^{-2}$	0.1975
			17	$0.5316 \times 10^{-2}$	$0.6000 \times 10^{-3}$	$0.1020 \times 10^{-2}$	0.1919
0.02		7	$0.5176 \times 10^{-2}$	$0.3604 \times 10^{-3}$	$0.7376 \times 10^{-3}$	0.1425	
		17	$0.5233 \times 10^{-2}$	$0.3683 \times 10^{-3}$	$0.7182 \times 10^{-3}$	0.1372	
0.05		7	$0.5095 \times 10^{-2}$	$0.1342 \times 10^{-3}$	$0.4015 \times 10^{-3}$	0.0788	
		17	$0.5152 \times 10^{-2}$	$0.1421 \times 10^{-3}$	$0.3837 \times 10^{-3}$	0.0745	

KIMURA (1969) studied the process of change in linkage disequilibrium due to random drift when initial gamete frequencies are given, and concluded that  $r^2$  and  $\sigma_d^2$  approach to  $1/(4N_e c)$  when  $N_e c$  is large. The present formula (18) shows that  $\sigma_d^2$  in the stationary state also approaches  $1/(4N_e c')$  for a large  $N_e c'$ . On the other hand, when  $N_e c'$  is small,  $\sigma_d^2$  may take different values at these two states. For example, when  $N_e c = 0.5$ ,  $\sigma_d^2$  at the state of steady decay is 0.403 and this is almost twice as large as that of stationary state given in Table 1. In general, however, we may say that the standard linkage deviation as well as the correlation of gene frequencies do not differ very much between these two states.

In natural populations, the recombination fraction  $c$  between neighboring genes or cistrons is probably of the order of  $10^{-3}$  or  $10^{-4}$ . If  $c = 10^{-3}$ ,  $N_e = 10^4$  and  $k = 10^{-5}$ ,  $\sigma_d^2$  is about 0.023 and the correlation of gene frequencies between the two neighboring loci may be about 0.15. Thus, linkage disequilibrium may not be very important for this case. If, on the other hand,  $c = 10^{-4}$ ,  $N_e = 10^3$  and  $k = 10^{-5}$  we have  $\sigma_d = 0.62$  approximately so that linkage disequilibrium becomes more important. When we consider nucleotide sites that are near to each other within a cistron the value of  $N_e c'$  may become quite small and linkage disequilibrium will be very important, even if the sign of  $D$ , not  $D^2$ , may be positive

or negative giving the mean value of  $D = 0$ . A more extensive discussion on this problem, will be given elsewhere.

The greater importance of linkage disequilibrium arises when we consider a region of a chromosome rather than just 2 pairs of genes. Namely, even if  $\sigma_d^2$  between each pair of loci may be small, the total effect of linkage disequilibrium over a given region of a chromosome may not be negligible. SVED (1968) argued that an apparent overdominance might arise due to linkage disequilibrium between a pair of neutral alleles that are linked to overdominant loci. The results obtained in the previous section enable us to treat this problem more fully.

In what follows we will discuss the application of formula (11) to the cases where nonepistatic selection is involved in the two linked loci. Let us assume that  $A_1$  and  $A_2$  are selectively neutral, and that  $B_1$  and  $B_2$  are overdominant, such that heterozygote  $B_1B_2$  has selective advantage  $s$  over both homozygotes,  $B_1B_1$  and  $B_2B_2$ . Then the amount of change of  $q$  (frequency of  $B_1$ ) in one generation by selection is

$$\Delta q = sq(1-q)(1-2q).$$

At equilibrium,  $q = 1/2$  and around this equilibrium point  $\Delta q$  may be approximated by  $\Delta q = (s/4)(1-2q)$ . Here we assume that the change of  $p$  by selection pressure at the B locus is negligible. It will be shown in the next paper of this series that this approximation does not introduce a serious error in the following derivation of the standard linkage deviation. We will also assume that the change of  $D$  by selection is negligible. By taking account of selection pressure in this way and letting  $u_1 = v_1 = u_2 = v_2 \equiv u$ , formula (12) becomes,

$$E \left\{ \frac{p(1-p)}{4} \frac{\partial^2 f}{\partial p^2} + \frac{q(1-q)}{4} \frac{\partial^2 f}{\partial q^2} + \frac{D}{2} \frac{\partial^2 f}{\partial p \partial q} + \frac{D(1-2p)}{2} \frac{\partial^2 f}{\partial p \partial D} \right. \\ \left. + \frac{D(1-2q)}{2} \frac{\partial^2 f}{\partial q \partial D} + \frac{1}{4} (pq(1-p)(1-q) + D(1-2p)(1-2q) - D^2) \frac{\partial^2 f}{\partial D^2} \right. \\ \left. + N_e u (1-2p) \frac{\partial f}{\partial p} + N_e \left( \frac{s}{4} + u \right) (1-2q) \frac{\partial f}{\partial q} - D \left( \frac{1}{2} + N_e c' \right) \frac{\partial f}{\partial D} \right\} = 0 \quad (19)$$

This gives

$$\sigma_d^2 = \frac{1}{3 + 4N_e c' - \left[ 2 / \left( 2.5 + N_e c' + 4N_e u + \frac{N_e s}{2} \right) \right]} \quad (20)$$

This is not much different from formula (18) when  $N_e c' \gg 1$ , though the value of  $A$  corresponding to (10) may become considerably different due to higher equilibrium gene frequencies. More generally, for any linear systematic evolutionary pressure involving  $p$  and  $q$ , the formula for  $\sigma_d^2$  can be obtained in the same way and it can be shown that  $\sigma_d^2$  is determined almost entirely by  $N_e c'$  for  $N_e c' \gg 1$ .

We will now proceed to estimate the fitnesses of the homozygotes and the heterozygote at the A (neutral) locus. Let  $q_1$  and  $q_2$  be the relative frequencies of  $B_1$  among  $A_1$ -carrying and  $A_2$ -carrying chromosomes, respectively. Then the

mean fitnesses of  $A_1A_1$ ,  $A_1A_2$  and  $A_2A_2$ , for a given set of values of  $q_1$  and  $q_2$  are as follows;

$$\begin{aligned} W_{A_1A_1} &= 1-s(1-2q_1)-2sq_1^2 \\ W_{A_1A_2} &= 1-s(1-q_1-q_2)-2sq_1q_2 \\ W_{A_2A_2} &= 1-s(1-2q_2)-2sq_2^2. \end{aligned}$$

In order to evaluate their expected values, let  $q_1=q+b_1$  and  $q_2=q-b_2$ . This yields:

$$\begin{aligned} \text{Frequency of } A_1B_1 &= p(q+b_1) \\ A_1B_2 &= p(1-q-b_1) \\ A_2B_1 &= (1-p)(q-b_2) \\ A_2B_2 &= (1-p)(1-q+b_2) \end{aligned}$$

Then we have

$$\begin{aligned} E(W_{A_1A_1}) &= 1-sE(1-2q_1)-2sE(q_1^2) = 1-sE\{1-2(q+b_1)\} \\ &\quad -2sE\{q^2+2b_1+b_1^2\}, \\ E(W_{A_1A_2}) &= 1-sE\{1-2q-b_1-b_2\}-2sE\{q^2-b_1b_2\}, \end{aligned}$$

and

$$E(W_{A_2A_2}) = 1-sE\{1-2(q-b_2)\}-2sE\{q^2-2b_2+b_2^2\}.$$

At the equilibrium  $E(q_1)=E(q_2)=E(q)=1/2$ ,  $E(b_1)=E(b_2)=0$ . Therefore, noting that  $b_1=D/p$  and  $b_2=D/(1-p)$ ,

$$E(W_{A_1A_2})-E(W_{A_1A_1}) = 2sE\{D^2/[p^2(1-p)]\} > 0$$

and

$$E(W_{A_1A_2})-E(W_{A_2A_2}) = 2sE\{D^2/[p(1-p)^2]\} > 0.$$

Namely, pseudo-overdominance appears by linkage disequilibrium. In this case we can evaluate the approximate amount of pseudo-overdominance at the A locus by using the expected value of  $D^2$  or  $\sigma_d^2$  at equilibrium. The magnitude of pseudo-overdominance will be enhanced if many overdominant loci are linked to the A locus, since its magnitude would be roughly proportional to the sum of  $D^2$  values between A and the accompanying overdominant loci.

Furthermore, it is possible to show that not only overdominance but also ordinary dominance may lend pseudo-overdominance to the linked neutral loci. Taking the fitnesses of  $B_1B_1$ ,  $B_1B_2$  and  $B_2B_2$  as  $1-s$ ,  $1-sh$  and  $1$ , respectively, and denoting by  $q_1$  and  $q_2$  the frequencies of  $B_1$  among  $A_1$ -carrying and  $A_2$ -carrying chromosomes as before, we have, for a given set of values of  $q_1$  and  $q_2$ ,

$$\begin{aligned} W_{A_1A_1} &= 1-2shq_1-s(1-2h)q_1^2 \\ W_{A_1A_2} &= 1-sh(q_1+q_2)-s(1-2h)q_1q_2, \end{aligned}$$

and

$$W_{A_2A_2} = 1-2shq_2-s(1-2h)q_2^2.$$

Since  $E(q_1)=E(q_2)$ , and,  $E(q_1^2) > E(q_1q_2) < E(q_2^2)$  as before, it is possible to show that  $E(W_{A_1A_1}) < E(W_{A_1A_2}) > E(W_{A_2A_2})$  unless  $h \geq 0.5$ . The magnitude of pseudo-overdominance in such a case may be small, since equilibrium frequencies of  $q_1$  and  $q_2$  are determined by the balance between mutation and selection, and therefore they are low. However, it is important to note that pseudo-overdominance may arise by linkage disequilibrium even if there are no overdominant loci in



the strict sense. This was clearly shown in the Monte Carlo experiments reported by MARUYAMA and KIMURA (1968).

Other important effects of linkage disequilibrium due to random drift include the increase of probability of simultaneous segregation as pointed out by HILL (1968). Further discussion on this and other topics relating to linkage disequilibrium in finite populations will be presented elsewhere.

## SUMMARY

Linkage disequilibrium in finite populations was investigated by the method of moment-generating matrix and also by the method of diffusion equations, assuming that steady state is reached with respect to distribution of chromosome frequencies under recurrent mutation and random sampling of gametes.—It was shown that if we measure the amount of linkage disequilibrium by  $\sigma_d^2 = E\{D^2\} / E\{p(1-p)q(1-p)\}$ , in which  $D$  is the coefficient of linkage disequilibrium and  $p$  and  $q$  are gene frequencies in the two loci, then  $\sigma_d^2 = 1 / [3 + 4N_e c' - 4 / (5 + 2N_e c' + 2N_e k)]$ , where  $N_e$  is the effective size of the population,  $c' = c + k$  in which  $c$  is the recombination fraction, and  $k$  is the sum of mutation rates. The square root of this quantity, i.e.,  $\sigma_d$ , may be termed the *standard linkage deviation*. It is approximately equal to the correlation of gene frequencies between the two loci.  $\sigma_d$  is almost entirely determined by  $N_e c'$  when  $N_e c' \gg 1$ . It was further shown that  $\sigma_d$  is not much influenced by nonepistatic selection.—The importance of linkage disequilibrium in finite populations consists mainly in its influence on the selective values of neighboring genes. Apparent overdominance of intrinsically neutral alleles may arise by the presence of overdominant or ordinarily dominant loci that are tightly linked to it. The magnitude of such pseudo-overdominance is roughly proportional to the sum of  $D^2$  values between the neutral and the remaining loci.

## APPENDIX

*Basic equations for deriving the moments of the gene frequency distribution at stationary state*

I. *Single variable:* Let  $x$  be the gene frequency and let  $\phi(x)$  be the density function of the stationary distribution. Consider a function  $f(x)$  and let

$$E_{\phi}(f(x)) = \int_0^1 f(x)\phi(x)dx$$

be its expectation with respect to this distribution. If we denote by  $\delta x$  the amount of change in  $x$  for one generation such that  $x' = x + \delta x$  is the frequency in the next generation, than at stationary state

$$E\{f(x + \delta x) - f(x)\} = 0, \quad (A1)$$

where  $E \equiv E_{\phi} E_{\delta}$  in which  $E_{\delta}$  designates an operator of taking expectation with respect to  $\delta x$ . Expanding  $f(x + \delta x)$  in terms of  $\delta x$  and neglecting terms involving  $(\delta x)^3$  and higher powers of  $\delta x$ , we have, from (A1),

$$E_{\phi} E_{\delta} \left\{ \delta x f'(x) + \frac{(\delta x)^2}{2!} f''(x) \right\} = 0$$

or

$$E_{\phi} \left\{ E_{\delta} (\delta x) f'(x) + \frac{E_{\delta} \{(\delta x)^2\}}{2} f''(x) \right\} = 0,$$

Substituting the mean  $M_{\delta x}$  and the variance  $V_{\delta x}$  of the change of  $x$  per generation, respectively, for  $E_{\delta}(\delta x)$  and  $E_{\delta}\{(\delta x)^2\}$  we obtain

$$E_{\phi} \left\{ \frac{1}{2} V_{\delta x} \frac{d^2 f(x)}{dx^2} + M_{\delta x} \frac{df(x)}{dx} \right\} = 0. \quad (\text{A2})$$

To simplify the expression we may omit subscript  $\phi$  in  $E_{\phi}$  and write the above equation as

$$E \left\{ \frac{1}{2} V_{\delta x} f'' + M_{\delta x} f' \right\} = 0. \quad (\text{A2}')$$

II. *Two and more variables*: Since the principle can be explained most easily for two variables, we will here consider the case of two independent random variables  $x_1$  and  $x_2$ . Let  $\phi(x_1, x_2)$  be the stationary distribution and let  $f(x_1, x_2)$  be a given function of  $x_1$  and  $x_2$ . As in the single variable case, if  $\delta x_1$  and  $\delta x_2$  are respectively the amounts of change in  $x_1$  and  $x_2$  for one generation, then we have

$$E\{f(x_1 + \delta x_1, x_2 + \delta x_2) - f(x_1, x_2)\} = 0. \quad (\text{A3})$$

Using the same procedure as was used to derive (A2') from (A1), we obtain

$$E \left\{ \frac{1}{2} V_{\delta x_1} \frac{\partial^2 f}{\partial x_1^2} + W_{\delta x_1 \delta x_2} \frac{\partial^2 f}{\partial x_1 \partial x_2} + \frac{1}{2} V_{\delta x_2} \frac{\partial^2 f}{\partial x_2^2} + M_{\delta x_1} \frac{\partial f}{\partial x_1} + M_{\delta x_2} \frac{\partial f}{\partial x_2} \right\} = 0, \quad (\text{A4})$$

where  $W_{\delta x_1 \delta x_2}$  is the covariance between  $\delta x_1$  and  $\delta x_2$ , and  $M_{\delta x_i}$  and  $V_{\delta x_i}$  ( $i=1,2$ ) are the mean and the variance of  $x_i$  per generation.

#### LITERATURE CITED

- HILL, W. G., 1968 Population dynamics of linked genes in finite populations. Proc. 12th Intern. Congr. Genet. **2**: 146-147.
- HILL, W. G. and A. ROBERTSON, 1968 Linkage disequilibrium in finite populations. Theoretical and Applied Genetics (Der Züchter) **38**: 226-231.
- KIMURA, M., 1956 A model of a genetic system which leads to closer linkage by natural selection. Evolution **10**: 278-287. — 1964 Diffusion models in population genetics. J. Appl. Probab. **1**: 177-232.
- MARUYAMA, T. and M. KIMURA, 1968 Development of temporary overdominance associated with neutral alleles. Proc. 12th Intern. Congr. Genet. **1**: 229.
- OHTA, T. and M. KIMURA, 1969 Linkage disequilibrium due to random genetic drift. Genet. Res. **13**: 47-55.
- SVED, J. A., 1968 The stability of linked systems of loci with a small population size. Genetics **59**: 543-563.
- WRIGHT, S., 1938 The distribution of gene frequencies under irreversible mutation. Proc. Natl. Acad. Sci. U.S. **24**: 253-259.