Genet. Res., Camb. (1970), **16**, *pp.* 165–177
Printed in Great Britain

Development of associative overdominance through linkage disequilibrium in finite populations*

BY TOMOKO OHTA AND MOTOO KIMURA
National Institute of Genetics, Mishima, Japan

(*Received* 18 *February* 1970)

SUMMARY

Associative overdominance arises at an intrinsically neutral locus through its non-random association with overdominant loci. In finite populations, even if fitness is additive between loci, non-random association will be created by random genetic drift.

The magnitude of such associative overdominance is roughly proportional to the sum of σ_d^2's between the neutral and the surrounding overdominant loci, where σ_d^2 is the squared standard linkage deviation, defined between any two loci by the relation

$$\sigma_d^2 = E(D^2)/E\{p(1-p)q(1-q)\},$$

in which p and $1-p$ are frequencies of alleles A_1 and A_2 in the first locus, q and $1-q$ are frequencies of alleles B_1 and B_2 in the second locus, and D is the coefficient of linkage disequilibrium. A theory was developed based on diffusion models which enables us to obtain formulae for σ_d^2 under various conditions, and Monte Carlo experiments were performed to check the validity of those formulae.

It was shown that if A_1 and A_2 are strongly overdominant while B_1 and B_2 are selectively neutral, we have approximately

$$\sigma_d^2 = 1/(4N_e c),$$

provided that $4N_e c \gg 1$, where N_e is the effective population size and c is the recombination fraction between the two loci. This approximation formula is also valid between two strongly overdominant as well as weakly overdominant loci, if $4N_e c \gg 1$.

The significance of associative overdominance for the maintenance of genetic variability in natural populations was discussed, and it was shown that $N_e s'$, that is, the product between effective population size and the coefficient of associative overdominance, remains constant with varying N_e, if the total segregational (overdominant) load is kept constant.

The amount of linkage disequilibrium expected due to random drift in experimental populations was also discussed, and it was shown that $\sigma_d^2 = 1/(n-1)$ in the first generation, if it is produced by extracting n chromosomes from a large parental population in which $D = 0$.

* Contribution No. 757 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411, Japan. Aided in part by a Grant-in-Aid from the Ministry of Education, Japan.

1. INTRODUCTION

It has been pointed out by several authors that linkage disequilibrium may create an apparent overdominance at intrinsically non-overdominant loci. In other words, non-random association with overdominant or ordinarily dominant loci may result in an apparent heterozygote advantage (Comstock & Robinson, 1952; Frydenberg, 1963; Chigusa & Mukai, 1964; Maruyama & Kimura, 1968). However, the underlying mechanism for such apparent overdominance has never been clarified.

Recently, Sved (1968); Ohta & Kimura (1969*b*) presented theoretical treatments of this problem by considering non-random association between neutral and overdominant loci due to random drift in finite populations. They showed that the degree of associative overdominance depends on the square of the coefficient of linkage disequilibrium, D^2, which in turn depends on the effective population size and the recombination fraction. Sved used a model in which all gene frequencies are assumed to be held at 50 % by strong overdominance while mutation is so rare as to be negligible. In considering natural populations, however, it may be more appropriate to assume a steady state in which random drift, recurrent mutation and natural selection balance each other.

Ohta & Kimura (1969*b*) developed a more general theory based on diffusion models to obtain the expected value of D^2 at steady state determined by random drift and mutation. They also showed that associative overdominance may appear at an intrinsically neutral locus when it is associated with overdominant or ordinarily dominant loci. Their treatment is valid under linear evolutionary pressures, such as mutation and migration, but, to extrapolate this to include selection, even if the selective change of gene frequencies may be linearized without serious error, should need justification. So, in the present paper, we treat a situation in which a neutral locus is linked with a strongly overdominant locus.

Thus, the present paper is an extension and elaboration of our previous work (Ohta & Kimura, 1969*b*), with special reference to the development of associative overdominance due to linkage disequilibrium. We will first present a theoretical treatment based on diffusion models and then demonstrate its validity using Monte Carlo methods.

Also, the bearing of associative overdominance on the maintenance of genetic variability in natural populations will be discussed.

2. ASSOCIATIVE OVERDOMINANCE

Let us consider two linked loci and assume that a pair of alleles, A_1 and A_2, are segregating (with respective frequencies p and $1-p$) in the first locus, and the other pair, B_1 and B_2 (with frequencies q and $1-q$) in the second locus. No selection is assumed at the B locus and overdominance or ordinary dominance is assumed at the A locus. Let the relative fitnesses of A_1A_1, A_1A_2 and A_2A_2 be respectively $1-s$, $1-hs$ and 1, and let p_1 and p_2 be the relative frequencies of A_1

among B_1-and B_2-carrying chromosomes. Then the mean fitnesses of B_1B_1, B_1B_2 and B_2B_2, for a given set of values of p_1 and p_2, are

$$\left.\begin{aligned} W_{B_1B_1} &= 1-2hsp_1-s(1-2h)p_1^2, \\ W_{B_1B_2} &= 1-hs(p_1+p_2)-s(1-2h)p_1p_2, \\ W_{B_2B_2} &= 1-2hsp_2-s(1-2h)p_2^2. \end{aligned}\right\} \tag{1}$$

In order to evaluate their expected values, let $p_1 = p+b_1$ and $p_2 = p-b_2$. If we denote by g_1, g_2, g_3 and g_4 the relative frequencies of the four types of chromosomes, A_1B_1, A_1B_2, A_2B_1 and A_2B_2, then

$$g_1 = q(p+b_1), \quad g_2 = (1-q)(p-b_2), \quad g_3 = q(1-p-b_1), \quad g_4 = (1-q)(1-p+b_2).$$

Therefore, we have $b_1 = D/q$ and $b_2 = D/(1-q)$, where $D = g_1g_4-g_2g_3$, $p = g_1+g_2$ and $q = g_1+g_3$. By substituting these relations in the formulae (1), we get the expected amount of associative overdominance at the B locus:

$$\left.\begin{aligned} E\{W_{B_1B_2}-W_{B_1B_1}\} &= E\left\{(hs+s(1-2h)p)\frac{D}{q(1-q)}\right. \\ &\qquad \left.+s(1-2h)\frac{D^2}{q^2(1-q)}\right\} = s(1-2h)E\left\{\frac{D^2}{q^2(1-q)}\right\}, \\ E\{W_{B_1B_2}-W_{B_2B_2}\} &= E\left\{-(hs+s(1-2h)p)\frac{D}{q(1-q)}\right. \\ &\qquad \left.+s(1-2h)\frac{D^2}{q(1-q)^2}\right\} = s(1-2h)E\left\{\frac{D^2}{q(1-q)^2}\right\}. \end{aligned}\right\} \tag{2}$$

Here, E stands for the operator of taking expectations and we assume that $E(D) = 0$. The quantity, $D^2/q^2(1-q)$ or $D^2/q(1-q)^2$ may be considered as a measure of association in gene frequencies between neutral and overdominant loci, through which apparent overdominance is created at the neutral locus. Now, as reported earlier (Ohta & Kimura, 1969*b*), the squared correlation coefficient (r^2) of gene frequencies between two loci is approximately equal to the squared standard linkage deviation, i.e. $\sigma_d^2 = E(D^2)/E\{pq(1-p)(1-q)\}$. In the present paper, we are mainly interested in cases in which q takes an intermediate value rather than considering the expected value. Hence we replace the above expressions by,

$$\left.\begin{aligned} E\{W_{B_1B_2}-W_{B_1B_1}\} &= s(1-2h)\sigma_d^2\left\{\frac{p(1-p)}{q}\right\}, \\ E\{W_{B_1B_2}-W_{B_2B_2}\} &= s(1-2h)\sigma_d^2\left\{\frac{p(1-p)}{1-q}\right\}. \end{aligned}\right\} \tag{3}$$

For the special case of symmetric overdominance (with fitnesses of A_1A_1, A_1A_2 and A_2A_2 of $1-s$, 1 and $1-s$), it can easily be shown that the corresponding expressions are,

$$E\{W_{B_1B_2}-W_{B_1B_1}\} \approx \frac{s}{2}\frac{\sigma_d^2}{(q)} \quad \text{and} \quad E\{W_{B_1B_2}-W_{B_2B_2}\} \approx \frac{s}{2}\frac{\sigma_d^2}{(1-q)}. \tag{4}$$

https://doi.org/10.1017/S0016672300002391 Published online by Cambridge University Press

The validity of these approximations was checked by Monte Carlo experiments, as will be shown later. The above expressions are clearly positive, and also expressions (2) and (3) are positive unless $h \geqslant \frac{1}{2}$. We will show in the following sections that $E(D) = 0$ at steady state, unless epistatic interaction is very strong or the recurrent mutations are of special type creating linkage disequilibrium. For example, if the two loci are multiplicatively overdominant, as shown by Bodmer & Felsenstein (1967), stable linkage disequilibrium will be established by selection only when the recombination fraction between them is less than $s^2/4$. Also, we can show that $E(D)$ is not zero at equilibrium, if the direction of mutation at one locus depends on the kind of alleles at another locus.

When B locus is selectively neutral and A locus is overdominant or ordinarily dominant, we need to estimate associative overdominance at B locus. The gene frequency at B locus may often deviate from its equilibrium value. So, in considering associative overdominance we might substitute a possible or observational value of q.

Comstock & Robinson (1952); Chigusa & Mukai (1964) reported the possibility of apparent overdominance for the explanation of their data. Their models are somewhat different from ours in that they assumed selection in all loci. For simplicity's sake, let us assume complete dominance at both A and B loci ($h = 0$). Then, if there is enough negative linkage disequilibrium, the excess of repulsion double heterozygotes will result in an apparent heterosis. This type of pseudo-overdominance may be responsible for many transient polymorphisms in experimental populations as well as for hybrid vigour in many crop plants including the maize. Frydenberg's (1963) interpretation of his experimental result is more similar to our present model. He concluded that the overdominance observed at his marker locus was, at least partly, due to its association with the inversion chromosome, and he termed this phenomenon associative overdominance, although he did not make any quantitative treatment of his model.

We will now proceed to present out basic theory based on diffusion models.

3. BASIC THEORY

The main aim of this section is to derive formulae for σ_d^2, that is, the square of the standard linkage deviation at steady state determined by mutation, selection and random drift, assuming a neutral and an overdominant locus. At the overdominant locus, it is assumed that the selection is so strong that gene frequencies are kept practically constant. If the selection is not strong (with $N_e s$ less than unity), one may use the result already obtained in our previous paper (Ohta & Kimura, 1969*b*). Namely, for a symmetric overdominance at one locus and with symmetric mutation rates at both loci, the square of the standard linkage deviation is

$$\sigma_d^2 = \frac{1}{3 + 4N_e(c + k_m) - 4/(5 + 2N_e(c + 2k_m) + N_e s)}, \tag{5}$$

where c is the recombination fraction between the two loci, s is the heterozygote

advantage over both homozygotes and k_m is the sum of the mutation rates. Therefore, $\sigma_d^2 \approx 1/(4N_e c)$ if $4N_e c$ is large. In the following treatments, we will show that even with a very strong overdominance at one locus the relation between σ_d^2 and $N_e c$ does not much differ from this.

As shown by Ohta & Kimura (1969*b*), if f is a polynomial of random variables describing the stationary distribution, then we have

$$E\{L_B(f)\} = 0, \tag{6}$$

where L_B denotes the differential operator such that if there are n independent random variables, $x_1, x_2, \ldots, x_n$, the equation becomes

$$E\left\{\frac{1}{2}\sum_{i=1}^{n} V_{\delta x_i}\frac{\partial^2 f}{\partial x_i^2} + \sum_{i>j}^{n} W_{\delta x_i \delta x_j}\frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{i=1}^{n} M_{\delta x_i}\frac{\partial f}{\partial x_i}\right\} = 0, \tag{7}$$

where $M_{\delta x_i}$ and $V_{\delta x_i}$ are the mean and the variance of δx_i and $W_{\delta x_i \delta x_j}$ is the covariance between δx_i and δx_j per unit time (generation).

Equation (6) enables us to calculate the moments of the frequency distribution. Let us apply this equation to the treatment of the present problem. We will assume that overdominance at the A locus is so strong that the frequency p of allele A_1 is constant which we denote by $\hat{p}$. At B locus, we assume that a pair of alleles B_1 and B_2 are selectively neutral, and we denote their frequencies respectively by q and $1-q$. Furthermore, let q_1 and q_2 be respectively the frequencies of B_1 among A_1- and A_2-carrying chromosomes. Both q_1 and q_2 are random variables. We will denote by u the mutation rate from B_1 to B_2, and by v the rate in the reverse direction. Let N_e be the 'variance' effective size of the population and c be the recombination fraction between A and B loci. Then the following equation can be obtained at steady state for q_1 and q_2 by taking account of mutation, recombination and random sampling of gametes.

$$\begin{aligned} E\Bigg\{&\frac{q_1(1-q_1)}{4N_e\hat{p}}\frac{\partial^2 f}{\partial q_1^2} + [(1-\hat{p})cq_2 + v - ((1-\hat{p})c + u + v)q_1]\frac{\partial f}{\partial q_1} \\ &+ \frac{q_2(1-q_2)}{4N_e(1-\hat{p})}\frac{\partial^2 f}{\partial q_2^2} + [c\hat{p}q_1 + v - (c\hat{p} + u + v)q_2]\frac{\partial f}{\partial q_2}\Bigg\} = 0. \end{aligned} \tag{8}$$

We now transform the set of independent random variables q_1 and q_2 into that of q and D using the relations,

$$q = \hat{p}q_1 + (1-\hat{p})q_2,$$

and

$$D = \hat{p}(1-\hat{p})(q_1 - q_2).$$

Then, equation (8) becomes,

$$\begin{aligned} E\Bigg\{&\frac{1}{4}\left[q(1-q) - \frac{D^2}{\hat{p}(1-\hat{p})}\right]\frac{\partial^2 f}{\partial q^2} + \frac{1}{2}\left[(1-2q)D + \frac{2\hat{p}-1}{\hat{p}(1-\hat{p})}D^2\right]\frac{\partial^2 f}{\partial q \partial D} \\ &+ \frac{1}{4}\left[\hat{p}(1-\hat{p})q(1-q) + (1-2\hat{p})(1-2q)D - \frac{1-3\hat{p}(1-\hat{p})}{\hat{p}(1-\hat{p})}D^2\right]\frac{\partial^2 f}{\partial D^2} \\ &+ N_e[v - (u+v)q]\frac{\partial f}{\partial q} - N_e(c+u+v)D\frac{\partial f}{\partial D}\Bigg\} = 0. \end{aligned} \tag{9}$$

https://doi.org/10.1017/S0016672300002391 Published online by Cambridge University Press

We should note here that the same result can be obtained by computing directly the means, the variances and the covariance of changes in q and D per generation. Using this equation, we will derive the moments of the distributions of q and D and therefore σ_d^2 at steady state.

Let $f = D$ in (9), then we get $E(D) = 0$. Similarly, we get $E(q) = v/(u+v)$ by putting $f = q$ in (9). Next, if we substitute three functions, D^2, q^2 and qD for f, we get the following simultaneous equations for $E(D^2)$, $E(q^2)$ and $E(qD)$.

$$\left.\begin{aligned} &E\left\{\hat{p}(1-\hat{p})q(1-q)+(1-2\hat{p})(1-2q)D \right.\\ &\qquad\left. -\frac{1-3\hat{p}(1-\hat{p})}{\hat{p}(1-\hat{p})}D^2-4N_e(c+u+v)D^2\right\}=0,\\ &E\left\{q(1-q)-\frac{D^2}{\hat{p}(1-\hat{p})}+4N_e[vq-(u+v)q^2]\right\}=0,\\ &E\left\{(1-2q)D+\frac{2\hat{p}-1}{\hat{p}(1-\hat{p})}D^2+2N_e[vD-(u+v)qD]-2N_e(c+u+v)qD\right\}=0. \end{aligned}\right\} \tag{10}$$

Solving these equations, we obtain the following formula for

$$\sigma_d^2 \equiv E(D^2)/\hat{p}(1-\hat{p})E\{q(1-q)\}.$$

$$\sigma_d^2 = \frac{1}{1+4N_e(c+k_m')+\dfrac{(1-2\hat{p})^2}{\hat{p}(1-\hat{p})}\dfrac{N_e(c+2k_m')}{1+N_e(c+2k_m')}}. \tag{11}$$

In this equation, $k_m' = u+v$, c is the recombination fraction between the two loci, and $\hat{p}$ is the frequency of the overdominant allele A_1 supposed to be kept constant in a population of effective size N_e. For a special case of symmetrical overdominance at A locus, $\hat{p} = 1/2$ and the last term in the denominator vanishes giving

$$\sigma_d^2 = \frac{1}{1+4N_e(c+k_m')}. \tag{12}$$

If we compare this formula with the corresponding formula, equation (5) obtained assuming weak overdominance, we note that for a large value of $N_e c$, they become practically the same. We may also note that the total mutation rate, k_m' in this formula is the sum only for the B locus, whereas k_m in formula (5) is the sum for both A and B loci, since the effect of mutation is neglected at the overdominant locus in the present treatment.

When $N_e c$ is small, and especially at the limit of $N_e(c+k_m) \to 0$, these two formulae give somewhat different values. Namely, at this limit, σ_d^2 in formula (11) or (12) approaches 1, whereas σ_d^2 in formula (5) approaches a value between $1/2{\cdot}2 \sim 1/3$. Also, for such an extremely tight linkage, and for an intermediate intensity of selection such as $N_e s = 2$, the exact evaluation of σ_d^2 appears to be very difficult.

Considering all these results, we may conclude that if $N_e c$ is much larger than unity, we have

$$\sigma_d^2 \approx 1/(4N_e c), \tag{13}$$

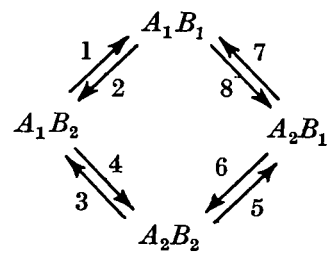
with good approximation. This simple approximation formula should have a wide applicability, because in most of the natural populations and for most of the linked loci, $4N_e c \gg 1$ is expected.

It is interesting to note that this approximation formula is also valid for the case of steady decay (cf. Ohta & Kimura, 1969*a*).

4. MONTE CARLO EXPERIMENTS

Using the IBM 360 computer, Monte Carlo experiments were performed simulating a two-locus genetic system. A simple scheme following Ohta (1968) was used for the experiments, that is, selection and recombination were carried out deterministically and sampling and mutation were performed by generating uniform pseudo-random numbers $X(0{\cdot}0 \sim 1{\cdot}0)$ using subroutine RANDU in FORTRAN IV. Each generation consists of mutation, selection, recombination and sampling. The initial frequencies of four gamete types were read into the computer and the simulation experiments were continued up to 200 generations so that the results represent the equilibrium state.

Let us assign the numbers 1, 2, 3 and 4 to four gamete types, A_1B_1, A_1B_2, A_2B_1 and A_2B_2. We will denote by g_i the frequency of gamete i and by z_{ij} the frequency of the zygote formed by the union of gametes i and j $(i, j = 1, 2, 3, 4)$.



Also we will number eight directions of mutations according to the above diagram. The process of mutation is as follows: We generate a sequence of eight random numbers. Then, one mutation of type 1 is induced among the gametes with A_1B_2 if the first random number is less than $m_1 g_2$. Similarly, one mutation of type 2 is induced if the second random number is less than $m_2 g_1$ and so on. Here, $m_1 \sim m_8$ are constants representing the mutation rates. Next, selection was exerted on zygotes using the equations,

$$z'_{ij} = z_{ij} + \Delta z_{ij},$$

and

$$\Delta z_{ij} = \frac{z_{ij}(w_{ij} - \bar{w})}{\bar{w}},$$

https://doi.org/10.1017/S0016672300002391 Published online by Cambridge University Press

where w_{ij} is the fitness (in selective values) of individuals with genotype ij and $\overline{w}$ is the average selective value of individuals in the population. The sampling of zygotes was made by generating pseudo-random numbers N times each generation. Finally, the recombination was carried out deterministically and the frequencies of four gamete types to form the next generation were determined.

Table 1. *Results of Monte Carlo experiments to check the formula* (12)

(Each experimental value is the average of 1200 generations starting with gametic frequencies of 1/4 for all four types. Throughout the experiments the effective population number (N_e) was assumed to be 50 and k, the sum of mutation rates was $k = 0{\cdot}01$ so that $N_e k = 0{\cdot}5$.

		Monte Carlo			
		$N_e s = 4$		$N_e s = 20$	
c	Theoretical σ_d^2	σ_d^2	r^2	σ_d^2	r^2
0	0·333	0·362	0·257	0·258	0·134
0·005	0·250	0·183	0·123	0·205	0·173
0·01	0·200	0·128	0·080	0·168	0·133
0·02	0·143	0·113	0·074	0·134	0·108
0·03	0·111	0·076	0·048	0·082	0·065
0·04	0·091	0·100	0·077	0·080	0·060
0·05	0·077	0·074	0·060	0·073	0·054
0·06	0·067	0·074	0·048	0·064	0·043
0·07	0·059	0·051	0·040	0·041	0·037
0·08	0·053	0·033	0·025	0·042	0·037
0·09	0·048	0·042	0·028	0·041	0·035
0·1	0·043	0·038	0·035	0·033	0·031

Both cases of large and intermediate selection coefficients were tried. The main purpose of the experiments was to check the validity of formula (12). The symmetric overdominance ($s_1 = s_2 \equiv s$) was assumed at A locus and no selection was assumed at B locus. In one set of experiments, we assumed $N_e s = 4$ and in another, $N_e s = 20$. The population size was 50 and mutation rates were equal in all directions with $N_e k_m = 0{\cdot}5$. The experiments were carried out for various levels of recombination ranging from $c = 0$ to $c = 0{\cdot}1$. Each experiment consisting of 1200 generations started with gene frequencies of 1/2 at both loci and without linkage disequilibrium. In Table 1 theoretical and experimental values of σ_d^2 are presented together with corresponding values of r^2 obtained from the experiments. In computing σ_d^2 from the experiments, we took the ratio between the mean of D^2 and that of $pq(1-p)(1-q)$ each averaged over all 1200 generations. On the other hand, r^2 was obtained by taking the average of the ratios of these two statistics over 1200 generations.

In order to show the level of accuracy of formulae (3) and (4), we have produced Table 2 in which $\langle D^2/q^2(1-q)\rangle$ and $\langle D^2/q(1-q)^2\rangle$ are compared respectively with $\sigma_d^2\langle p(1-p)\rangle/\langle q\rangle$ and $\sigma_d^2\langle p(1-p)\rangle/\langle 1-q\rangle$, where $\langle\ \rangle$ denotes the average obtained from the experiments. Also, experimental values are used for σ_d^2. In the present case k_m (sum of mutation rates at A and B loci) is substituted for k'_m (sum of

https://doi.org/10.1017/S0016672300002391 Published online by Cambridge University Press

mutation rates at only B locus) in the formula, since mutation is not negligible at the overdominant locus in our experiments. As seen from the tables, the agreement between the theoretical predictions and the experimental results is satisfactory. However, the approximation formula (12) seems to overestimate slightly the true value. The reason for this appears to be that the gene frequency at the overdominant locus is not strictly fixed but slightly fluctuating.

Table 2. *Experimental check on the approximation involved in formula* (3)

(The data are derived from the same experiments which were performed to construct Table 1. In the table, the symbol $\langle\ \rangle$ denotes the average obtained from the experiments. For details, see text.)

$N_e s$	c	$\left\langle \frac{D^2}{q^2(1-q)} \right\rangle$	$\sigma_d^2 \frac{\langle p(1-p)\rangle}{\langle q\rangle}$	$\left\langle \frac{D^2}{q(1-q)^2} \right\rangle$	$\sigma_d^2 \frac{\langle p(1-p)\rangle}{\langle 1-q\rangle}$
4	0·0	0·139	0·125	0·178	0·224
	0·005	0·099	0·096	0·076	0·069
	0·01	0·090	0·079	0·059	0·043
	0·02	0·058	0·041	0·088	0·065
	0·03	0·026	0·019	0·075	0·053
	0·04	0·058	0·044	0·056	0·051
	0·05	0·056	0·040	0·043	0·030
	0·06	0·061	0·046	0·026	0·023
	0·07	0·013	0·014	0·067	0·055
	0·08	0·047	0·021	0·024	0·011
	0·09	0·045	0·024	0·026	0·016
	0·1	0·035	0·017	0·020	0·014
20	0·0	0·058	0·077	0·165	0·341
	0·005	0·101	0·117	0·098	0·090
	0·01	0·073	0·063	0·116	0·114
	0·02	0·090	0·088	0·055	0·051
	0·03	0·072	0·064	0·029	0·029
	0·04	0·081	0·065	0·032	0·028
	0·05	0·047	0·033	0·058	0·038
	0·06	0·040	0·026	0·063	0·039
	0·07	0·059	0·032	0·019	0·014
	0·08	0·019	0·016	0·049	0·034
	0·09	0·031	0·021	0·038	0·019
	0·1	0·014	0·012	0·039	0·024

5. DISCUSSION

The main aim of the present paper is to estimate, using our formulae for the square of the standard linkage deviation (σ_d^2), the approximate magnitude of associative overdominance created by linkage disequilibrium. Certainly, it is based on several approximations, but as demonstrated by the Monte Carlo experiments, these formulae give us a sufficiently accurate estimate for σ_d^2. One problem in our procedure of estimating associative overdominance is that we substitute the ratio of the expectations for the expectation of the ratio, cf. formulae (2) and (3). To see the magnitude of errors involved, we compared $\sigma_d^2 = E(D^2)/E\{p(1-p)q(1-q)\}$

https://doi.org/10.1017/S0016672300002391 Published online by Cambridge University Press

which is the ratio of the expectations, with $r^2 = E\{D^2/pq(1-p)(1-q)\}$ which is the expectation of the ratio. The latter is the square of the correlation of p and q in the usual sense. In Table 1, values of σ_d^2 and r^2 are compared. The values of r^2 are slightly smaller than the corresponding values of σ_d^2 mainly due to deviation of gene frequency (q) from 1/2 at the neutral locus. The reason for $r^2 < \sigma_d^2$ may most easily be understood by considering the situation in which q happens to become extremely small. Then r^2 must approach 0, but σ_d^2 will not be much influenced by such a deviation. Thus, excluding such extreme situations, we may conclude that, for large $N_e c$, the approximation $\sigma_d^2 \approx 1/\{4N_e(c+k_m)\}$ is valid and useful to evaluate associative overdominance.

It is quite interesting that essentially the same formula for σ_d^2 holds also at the transient state in which the variability is steadily decaying, as long as $N_e c$ is large. Hill & Robertson (1968) found that the square of the correlation of gene frequencies (r^2) at steady decay becomes approximately $1/(4N_e c)$ when $N_e c$ is large. Ohta & Kimura (1969*a*) also showed that under such a situation σ_d^2 becomes approximately $1/(4N_e c)$. On the other hand, when $N_e c$ is small, there are some differences in σ_d^2 between the stationary state and the steadily decaying state. Also, the effects of selection and mutation become important.

Let us investigate how much associative overdominance will be developed if a neutral locus is linked with a number of overdominant loci on the same chromosome. Our model is as follows. We assume that there are n_1 overdominant loci on the left and n_2 overdominant loci on the right of the neutral locus in such a way that the recombination fraction between the neutral locus and the ith overdominant locus, either on the left or on the right, is ic_0. Let s be the selection coefficient against either homozygote in each overdominant locus. We then ask how much associative overdominance will develop at the neutral locus (B), where alleles B_1 and B_2 are segregating with respective frequencies of q and $1-q$. We assume that linkage among overdominant loci is loose in comparison with the selection intensity, so that these overdominant loci do not constitute 'super genes' and that they are more or less randomly combined. However, a small amount of linkage disequilibrium will be created among them by random genetic drift. If we assume strict additivity in fitness among the overdominant loci, i.e. additive overdominance, and if σ_d^2 between the neutral and each overdominant locus is adequately given by our formula (12), the associative overdominance at the neutral locus simply becomes the sum of the effects of all linked overdominant loci.

For a multiplicative overdominance, we may take logarithms of the individual fitnesses in the following formulation. Let us suppose that q happens to take an intermediate value not very far from 1/2. That is, the frequency of B_1 or B_2 has increased in the population by random frequency drift. Then, from equation (3), if we take $q = 1/2$, the coefficient of associative overdominance s' becomes approximately,

$$s' = E\{W_{B_1B_2} - W_{B_1B_1}\} = E\{W_{B_1B_2} - W_{B_2B_2}\} = s\left\{\sum_{i=1}^{n_1}\sigma_d^2(i) + \sum_{i=1}^{n_2}\sigma_d^2(i)\right\},$$

where $\sigma_d^2(i)$ is the standard linkage deviation between the neutral locus and the ith overdominant locus, either on the right or on the left. Then, using

$$\sigma_d^2(i) \approx 1/(4N_e c_0 i),$$

we have

$$s' = \frac{s}{4N_e c_0}\{2\gamma + \log n_1 + \log n_2\}, \tag{14}$$

where $\gamma = 0{\cdot}577$ is Euler's constant. For example, if the effective population size is 1000 and if there are 100 overdominant loci on the chromosome such that $n_1 = n_2 = 50$ covering roughly the map length of 100 units ($c_0 = 0{\cdot}01$), the neutral locus being located just in the middle, s' becomes about $0{\cdot}22s$. If the population is ten times as large, s' is $0{\cdot}022s$. An important point to note here is that $N_e s'$ remains constant with varying N_e. If we change the number of overdominant loci such that the total effect on the given segment does not change (for example, $2n$ loci each with homozygous disadvantage $s/2$ instead of n loci each with s), then the amount of associative overdominance s' changes relatively little (for example, by the factor $\log 2n/\log n$). For a large value of N_e such as 10^5 or 10^6, the value of s' may be quite small, but it does retard fixation at the neutral locus, since $N_e s'$ remains constant.

So far, we have investigated the amount of pseudo-overdominance developed at a neutral locus through its non-random association with overdominant loci in a finite population.

In discussing the effect of such associative overdominance on the amount of heterozygosity at the neutral locus, we must be careful not to overlook the effect of mutation at that locus. Namely, at the neutral locus B, if we disregard the effect of associative overdominance, it can be shown that the expected amount of heterozygosity is

$$H_B = E\{2q(1-q)\} = \frac{8N_e u\bar{q}}{1+4N_e(u+v)}, \tag{15}$$

where $\bar{q} = v/(u+v)$. On the other hand, if we include the effect of associative overdominance from the A locus, it can be shown using equations (10) that the expected heterozygosity is

$$H_{B(A)} = \frac{8N_e u\bar{q}}{1+4N_e(u+v)-\sigma_d^2} = \frac{H_B}{1-\sigma_d^2/\{1+N_e(u+v)\}}, \tag{16}$$

where σ_d^2 is given by equation (11).

If $4N_e(u+v)$ is much larger than unity, H_B approaches $2\bar{q}(1-\bar{q})$ by mutation, and there may be little room left for pseudo-overdominance to enhance heterozygosity. However, if $4N_e(u+v)$ is small, we have

$$H_{B(A)} \approx (1+\sigma_d^2)H_B,$$

provided that $\sigma_d^2 \approx 1/(4N_e c)$ is small. Thus, the heterozygosity at the B locus is enhanced by the fraction $1/(4N_e c)$ by associative overdominance caused by A locus.

When a large number of overdominant loci are linked to the neutral locus, their effect on enhancing heterozygosity at the neutral locus would probably be pro-

portional to $\exp(\Sigma\sigma_d^2)$, as long as each σ_d^2's is small. A more detailed study on this subject will be left to our future reports. However, we should mention here that when both A and B loci are kept polymorphic by strong overdominance, linkage disequilibrium is created between these two overdominant loci just as between a neutral and an overdominant loci. Namely, the approximation equation (13) also holds for them if c is the recombination fraction between them. This can be shown as follows. Let $\hat{p}$ and $\hat{q}$ be respectively the frequencies of A_1 and B_1 in these two loci, and suppose that both $\hat{p}$ and $\hat{q}$ are kept constant by strong overdominance. At the neighbourhood of $D = 0$, the sampling variance of D is approximately

$$V_{\delta D} = \hat{p}(1-\hat{p})\hat{q}(1-\hat{q})/(2N_e),$$

because the problem is analogous to that of sampling in 2×2 contingency table in statistics, and we can show that

$$\chi_1^2 = 2N_e D^2/[\hat{p}(1-\hat{p})\hat{q}(1-\hat{q})],$$

follows approximately Chi-square distribution with one degree of freedom, so that $E(\chi_1^2) = 1$. Thus, noting that $M_{\delta D} = -cD$, by recombination, the equation corresponding to (7) becomes

$$E\left\{\frac{1}{4N_e}\hat{p}(1-\hat{p})\,\hat{q}(1-\hat{q})\,\frac{\partial^2 f}{\partial D^2} - cD\frac{\partial f}{\partial D}\right\} = 0.$$

By setting successively $f = D$ and $f = D^2$ in this equation, we obtain $E(D) = 0$ and $E(D^2) = 4N_e c\hat{p}(1-\hat{p})\hat{q}(1-\hat{q})$. Therefore,

$$\sigma_d^2 = E(D^2)/[\hat{p}(1-\hat{p})\hat{q}(1-\hat{q})] = 1/(4N_e c),$$

as was to be shown. This formula should be valid when σ_d^2 is small.

Finally we intend to discuss problems of linkage disequilibrium in experimental populations. There are many reports on the experimental measure of fitness values with respect to isozyme alleles or other marker genes. Very often, however, the results merely reflect the effects of a group of surrounding genes and hence the effect of individual alleles on fitness is very difficult to measure. In experimental populations, initial linkage disequilibrium may be produced by sampling a relatively small number of chromosomes from a large parental population. As shown by Hill & Robertson (1968), if n chromosomes are sampled to form the first generation of experimental populations, the values of $X = E\{pq(1-p)(1-q)\}$, $Y = E\{D(1-2p)(1-2q)\}$ and $Z = \{D^2\}$ in the first generation is given by

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = \begin{bmatrix} \left(1-\frac{1}{n}\right)^2 & \frac{1}{n}\left(1-\frac{1}{n}\right)^2(1-c) & \frac{2}{n^2}\left(1-\frac{1}{n}\right)(1-c)^2 \\ 0 & \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)^2(1-c) & \frac{4}{n}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)(1-c)^2 \\ \frac{1}{n}\left(1-\frac{1}{n}\right) & \frac{1}{n}\left(1-\frac{1}{n}\right)^2(1-c) & \left(1-\frac{1}{n}\right)\left[\frac{1}{n^2}+\left(1-\frac{1}{n}\right)^2\right](1-c)^2 \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}, \tag{17}$$

where X_0, Y_0 and Z_0 are the corresponding values in the parental population.

https://doi.org/10.1017/S0016672300002391 Published online by Cambridge University Press

Assuming that the linkage disequilibrium in the parental population is negligible so that approximately $Y_0 = Z_0 = 0$, then the squared standard linkage deviation in the first generation of the experimental populations is

$$\sigma^2_{d,1} = \frac{E(Z_1)}{E(X_1)} = \frac{1}{n-1}. \tag{18}$$

Usually, the chromosomes thus sampled are rapidly multiplied from n to n' in the succeeding generations and then they are used for measuring the fitness of the marker gene. The following is a very rough estimate of linkage disequilibrium in such experimental populations. Applying the results of Ohta & Kimura (1969*a*), σ^2_d in the t-th generation may be given as follows,

$$\sigma^2_{d,t} = \frac{\sum_{i=1}^{3} C_{J_i}\left[\frac{X_1}{2(1+\lambda_i)} + \frac{1}{4}(3+4R+2\lambda_i)Y_1 + Z_1\right] e^{2\lambda_i t/n'}}{\sum_{i=1}^{3} C_H \left[\frac{X_1}{2(1+\lambda_i)} + \frac{1}{4}(3+4R+2\lambda_i)Y_1 + Z_1\right] e^{2\lambda_i t/n'}}, \tag{19}$$

where λ_i's are the first three eigenvalues of the Kolmogorov forward equation involved and C_{H_i}'s and C_{J_i}'s are the functions of λ_i's. These parameters are given in Ohta & Kimura (1969*a*). The value of $\sigma^2_{d,t}$ rapidly approaches to $1/2n'c$ for sufficiently large value of $n'c$. By using the formula (3), one can estimate the average degree of associative overdominance. Of course this method gives the overall average due to non-epistatic genes.

If there are strong epistatic effects between loci so that super-genes are formed, $E(D)$ is not zero and the result becomes much more complex.

REFERENCES

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

Chigusa, S. & Mukai, T. (1964). Linkage disequilibrium and heterosis in experimental populations of *Drosophila melanogaster* with particular reference to the *sepia* gene. *Japanese Journal of Genetics* **39**, 289–305.

Comstock, R. E. & Robinson, H. F. (1952). Estimation of average dominance of genes. In *Heterosis*, pp. 494–516. Ames: Iowa State College Press.

Frydenberg, O. (1963). Population studies of a lethal mutant in *Drosophila melanogaster*. I. Behaviour in populations with discrete generations. *Hereditas* **50**, 89–116.

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

Maruyama, T. & Kimura, M. (1968). Development of temporary overdominance associated with neutral alleles. *Proceedings of XII International Congress of Genetics, Tokyo*, vol. 1, p. 229.

Ohta, T. (1968). Effect of initial linkage disequilibrium and epistasis on fixation probability in a small population, with two segregating loci. *Theoretical and Applied Genetics* **38**, 243–248.

Ohta, T. & Kimura, M. (1969*a*). Linkage disequilibrium due to random genetic drift. *Genetical Research* **13**, 47–55.

Ohta, T. & Kimura, M. (1969*b*). Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. *Genetics* **63**, 229–238.

Sved, J. A. (1968). The stability of linked systems of loci with a small population size. *Genetics* **59**, 543–563.