THE STABILITY OF LINKED SYSTEMS OF LOCI WITH A SMALL POPULATION SIZE[1]

J. A. SVED

Department of Genetics, Stanford Medical Center, Palo Alto, California

Received January 22, 1968

IT is well known that in an infinite population, unselected genes at different loci tend to become associated at random, the speed of approach to random association being determined by the frequency of recombination (ROBBINS 1918). More recently, LEWONTIN and KOJIMA (1960), BODMER and FELSENSTEIN (1967) and others have shown that two linked genes may be held in permanent linkage disequilibrium (i.e., nonrandom association) by epistatic interaction of selective values associated with genes at the two loci. The magnitude of the epistatic interaction must, however, be quite large, or the linkage very close, in order for this condition to be achieved. LEWONTIN (1964) has extended the analysis to five loci, showing that the linkage disequilibrium achieved may be greater than predicted on the basis of the two locus calculations. However, high selective intensities and a high degree of epistatic interaction are once again required. For these reasons, calculations made with large numbers of loci and low selective intensities (e.g. KING 1967; MILKMAN 1967; SVED, REED and BODMER 1967) have usually been made under the assumption of random combinations of genes at all loci.

Little attention has apparently been paid to the question of whether finiteness of population size will affect these conclusions. From one point of view it is important that this possibility be tested. The calculations of KING *et al.* were made to test the assumption that heterosis at a large number of loci is feasible. But the assumption of heterosis at a large number of loci comes not so much as the result of direct experimental evidence but rather as a possible explanation for the large amount of variability found in natural populations (HARRIS 1966; JOHNSON *et al.* 1966; LEWONTIN and HUBBY 1966). Were these populations infinitely large, the need for postulating such a mechanism to ensure the maintenance of variability would not seem so great, since the possibility of chance fixation of genes would not have to be taken into account. Accordingly there is an important reason that the conditions for approach to linkage equilibrium be examined when the population size is finite. In the remainder of this paper, I will attempt to show that a certain amount of linkage disequilibrium is expected independently of any selection in a small sized population, and further that the disequilibrium could have important consequences in discussions about the maintenance of variability by selection.

A model system of loci: It seems appropriate to begin by considering experi-

[1] Supported by Public Health Service Grant GM 10452

mental evidence to try to determine what is a reasonable model system of linked loci, since the questions of what numbers of loci, strengths of linkage and intensities of selection have to be explained are crucial ones. The information is available for constructing a very primitive model of a Drosophila chromosome.

Based on molecular weight determinations, the total number of genes in Drosophila may be of the order of 2×10^5 (see CARLSON 1967). This number is an order of magnitude higher than earlier estimates based on salivary band counts (BRIDGES 1942). We will accept the larger number in this discussion, partly because it may be more reliable, but also because it accentuates the problem of accounting for the observed variability.

LEWONTIN and HUBBY (1966) found that approximately one in three proteins which they studied in *Drosophila pseudoobscura* was polymorphic. Accepting a similar figure for all proteins, this would indicate about 6×10^4 polymorphic loci, or perhaps 2×10^4 such loci per major chromosome of approximately 100 map units. This figure is, of course, a very approximate one. But the qualitative conclusion that there are probably many polymorphic loci, most of them closely linked to others, seems hard to dispute.

In considering the possibility that all these loci are maintained segregating by heterozygote advantage, as will be done throughout most of this paper, we must first consider the problem of what intensities of selection are permissible. SVED, REED and BODMER (1967) have argued on the basis of observed intensities of inbreeding depression that there could not be many more than 1,000 loci each with 1% heterozygote advantage. With 6×10^4 segregating loci it would seem that the average heterozygote advantage could not be much more than 0.015%.

An extreme numerical example: Most of the arguments to be made in this paper will be seen more clearly by use of a simple though admittedly extreme example. We will consider a model chromosome set containing 20 linked loci, each with two alleles at 50% frequency initially. The population size is assumed to be very small, and is taken as 25, or 50 chromosomes. The loci are assumed to be closely linked, so that in view of the small population size and short time span involved we may consider the linkage to be complete. The figures of the previous section suggest that a realistic recombination frequency for the outside two markers might be as low as 0.1%.

The 50 chromosome types may be written as, for example,

CHROMOSOME 1 A b c D E F . . .
CHROMOSOME 2 a B C D e F . . .
. .
CHROMOSOME 50 a b c d E F . . .

Genes at different loci are combined at random, so that there is no systematic linkage disequilibrium between any pair of loci. It is interesting to note that with exactly 50 chromosomes it is not possible to write down a set with exact linkage equilibrium at all pairs of loci. However Professor G. SZEKERES (personal communication) has shown that for 20 loci this could be done with as few as 24 chromosomes, or with any number higher than this and divisible by four.

Conceptually the above model is identical with that of a single locus with initially 50 different alleles. The theory of multiple alleles (KIMURA and CROW 1964; EWENS 1964) predicts that all 50 types cannot be kept segregating for long, and that most will be lost very quickly. In a number of computer runs made with this model, the population was usually reduced to either two or three chromosome types after only 15 generations.

Consider a typical population, which after 15 generations has been reduced to having two chromosome types. By this time not all of the twenty loci will still be segregating; on the average about ten will have become fixed. But the important point is that amongst the loci still segregating there is no longer linkage equilibrium but complete linkage disequilibrium between all pairs.

The population size chosen may have been too small for the heterozygote advantage at the individual loci to have significantly retarded gene fixation, but the extreme linkage disequilibrium now attained may change this. If for example the heterozygote at each locus possesses a 5% advantage, then if there is multiplicative interaction of selective values, the population now contains two chromosome types, or effectively two alleles, with $1 - (.95)^{10}$, or approximately 40% heterozygote advantage. This may be sufficient to ensure that the remaining loci are kept from being fixed for many more generations, despite the small population size. However had the loci been unlinked, the 5% heterozygote advantage would not be sufficient to oppose significantly the trend towards homozygosity at any locus.

The remainder of this paper will be devoted to the analysis of very similar models, but attempting to introduce the effects of recombination and a somewhat higher population size.

Some theoretical arguments: The overall argument will be divided into two stages. The build-up of linkage disequilibrium by chance fluctuation will be considered first. Measures of linkage disequilibrium for more than two loci will also be discussed under this heading. Following this, the question of how this disequilibrium affects the consequences of selection will be considered.

Expectation of linkage disequilibrium: The gametic frequencies in a two locus model may be written down as in Table 1. The measure of linkage disequilibrium commonly used is D, which is equal to $x_1x_4 - x_2x_3$. In terms of D, the four gametic frequencies may then be written in the alternative form given in the table. As previously mentioned, the value of D is expected to tend to zero in an infinite population if there is recombination between the two loci and no selection. The

TABLE 1

Gametic frequencies in a two locus model

Gamete	Frequency	Alternative form
AB	x_1	$p_Ap_B + D$
Ab	x_2	$p_A(1-p_B) - D$
aB	x_3	$(1-p_A)p_B - D$
ab	x_4	$(1-p_A)(1-p_B) + D$

aim of this calculation is to study the expected distribution of D if the population size is finite.

The general problem of finding expectations for D is a difficult one, particularly since it involves studying a three-dimensional process. The problem is however reduced to a one-dimensional one if it is assumed that the gene frequencies p_A and p_B are fixed. Even with this simplification the problem is a difficult one, and in the following calculation it will be assumed that both gene frequencies are one-half. Under these conditions the values of x_1, x_2, x_3 and x_4 are respectively $\frac{1}{4}+D$, $\frac{1}{4}-D$, $\frac{1}{4}-D$ and $\frac{1}{4}+D$. The distribution of D is symmetrical between the limits of $\pm$ 0.25.

It is convenient to study the distribution of $x_2 + x_3$, which will be written as x. In terms of D, this is equal to $\frac{1}{2} - 2D$. Following the usual diffusion theory arguments (Li 1955, *ch.* 23), the distribution of x may be written in the following form

$$\frac{C}{x\,(1-x)} \cdot e^{\,2\int \frac{\Delta x}{\sigma^2_x}\,dx} \tag{1}$$

where Δx is the expected change in frequency of x in one generation, σ^2_x is the variance of the change in gene frequency, and C is a normalizing constant. If there is no selection, the value of Δx depends only on the amount of recombination, and is equal to $2rD$ or $r(\frac{1}{2} - x)$, where r is the recombination frequency. This expectation is not seriously affected if selection is introduced, for example heterozygote advantage at one or both loci, provided that there is no strong epistasis between the genes at the two loci (Lewontin and Kojima 1960). The The value of σ^2_x is assumed to be $x(1-x)/2N$ following the usual binomial laws. This assumes that the recombination frequency is low, and the result will break down to some extent for high recombination frequencies.

Using the above values of Δx and σ^2_x, (1) reduces to

$$\frac{1}{\beta(2Nr,2Nr)} \cdot x^{2Nr-1} \cdot (1-x)^{2Nr-1}.$$

To find the mean value of D, we multiply the above quantity by $\frac{1}{2}\cdot(\frac{1}{2} - x)$ and integrate between the limits of 0 and 1. The result is, as expected, zero. The variance of D, or mean value of D^2, is found by multiplying by $\frac{1}{4}\cdot(\frac{1}{2} - x)^2$ and integrating, giving

$$\frac{1}{16(4Nr+1)}.$$

Some of the approximations in these calculations have been roughly checked by computer simulation. A two locus model was simulated with 20% heterozygote advantage at each locus to ensure that gene frequencies stayed close to 50%, and additive interaction. In one run of 2,000 generations with $N = 50$ and $r = .02$, the average value of D^2 was 0.0138 compared to the expectation of 0.0125. In a second run made with $r = .1$, the observed value was 0.00243, compared to an expected value of 0.00298. Although the populations were run for 2,000 generations, the variance of the estimate of D^2 is probably still quite high, since particularly with the lower recombination frequency the serial correlation

between successive generations was very high. However, the results are sufficient to show that the predictions are not markedly in error.

The restriction of gene frequencies to one-half is a serious limitation. As discussed at some length by LEWONTIN (1964), the statistic D is sensitive to changes in gene frequency. If the frequencies of the genes A and B are 0.3 instead of 0.5 for example, the range of values D can take is only $-.09$ to .21 instead of $-.25$ to .25. If either gene is fixed, the value of D must, of course, be zero. While it is easy to define the range of values of D for various gene frequencies, it is unfortunately not easy to define the expected variance of D. Nor is it obvious what is the effect of assuming that the gene frequencies are fixed and then calculating the expectation for D, rather than deriving the general distribution for D and making this conditional on particular gene frequencies.

More than two loci: The problem here is one of finding a statistic to measure disequilibrium, rather than one of deriving expectations. With $n(n-1)/2$ pairs possible for n loci, there may be too many D values to be readily interpretable.

From the 20 locus example given earlier one property of a system of loci with linkage disequilibrium may be seen. In the first generation, zygotes formed from the 50 chromosomal types will be heterozygous on the average at 10 loci, and rarely heterozygous at say less than 5 or more than 15 loci. On the other hand, when extreme linkage disequilibrium is attained, all individuals are heterozygous at either all, or none of the loci still segregating. Evidently the variance of the number of heterozygous loci per individual, which will be designated as V_H, has increased as the disequilibrium increases. In fact, the expected value of V_H in a random mating population can be written down in a simple way in terms of the values of D (c.f. HILL and ROBERTSON 1966). In particular, if all genes are at 50% frequency, then it is equal to

$$n/4 + 8\,\Sigma D^2 \tag{2}$$

where the summation is over all $n(n-1)/2$ pairs of loci.

This relationship may be demonstrated directly by writing down all terms contributing to both sides of the equation and establishing equivalence between the two. However, for a simpler approach I am indebted to PROFESSOR S. KARLIN who has pointed out that the problem may be reduced to a two locus problem simply by observing that

$$\begin{aligned} V_H &= V\,(\alpha_1 + \alpha_2 + \ldots + \alpha_n) \\ &= \sum_i V\,(\alpha_i) + 2\sum_{i,j} \text{Covariance}\,(\alpha_i, \alpha_j) \end{aligned} \tag{3}$$

where α_i is equal to the number of heterozygotes per locus per individual, i.e. 0 or 1. If the frequency of heterozygotes at the i^{th} locus is $2p_iq_i$, then

$$V(\alpha_i) = 2p_iq_i(1 - 2p_iq_i).$$

When all gene frequencies are equal to one-half, then $V(\alpha_i) = 1/4$ and $\sum_i V(\alpha_i) = n/4$.

To calculate the covariance term we must first write down the zygotic array expected from the gametic array of Table 1. This is given in Table 2, ignoring for the moment the selective values given. Then Covariance (α_A, α_B) may be calculated in the usual way as

TABLE 2

Zygotic array given by gametes of Table 1, showing number of heterozygotes at each locus and selective values assuming heterozygote advantage at B locus and selective neutrality at A

Genotype	Frequency	α_A	α_B	Selective values
AB/AB	x_1^2	0	0	$1-s$
AB/Ab	$2x_1x_2$	0	1	1
AB/aB	$2x_1x_3$	1	0	$1-s$
AB/ab	$2x_1x_4$	1	1	1
Ab/Ab	x_2^2	0	0	$1-t$
Ab/aB	$2x_2x_3$	1	1	1
Ab/ab	$2x_2x_4$	1	0	$1-t$
aB/aB	x_3^2	0	0	$1-s$
aB/ab	$2x_3x_4$	0	1	1
ab/ab	x_4^2	0	0	$1-t$

$$\Sigma f_{AB} \cdot \alpha_A \alpha_B - (\Sigma f_A \cdot \alpha_A) \cdot (\Sigma f_B \cdot \alpha_B)$$

which is equal to $x^2_1.0.0 + 2x_1x_2.0.1 + 2x_1x_3.1.0 + \ldots$

$$- (x^2_1.0 + 2x_1x_2.0 + 2x_1x_3.1 + \ldots)$$

$$\times (x^2_1.0 + 2x_1x_2.1 + 2x_1x_3.0 + \ldots).$$

After some simplification this expression may be shown to be

$$4D^2 + 8D(\tfrac{1}{2} - p_A)(\tfrac{1}{2} - p_B).$$

Then, substituting in (3), we have

$$V_H = 2\sum_i p_iq_i\,(1 - 2p_iq_i) + 8\sum_{i,j} D^2 + 16\sum_{i,j} D.(\tfrac{1}{2} - p_i)(\tfrac{1}{2} - p_j) \qquad (4)$$

If all gene frequencies are equal to one-half, all terms in D disappear giving formula (2). In general terms in D will be both positive and negative and will tend to cancel out in comparison to terms in D^2. As discussed later, however, there are some circumstances under which terms in D may make an appreciable contribution to formula (4).

The parameter V_H is conceptually useful in giving an overall measure of linkage disequilibrium. It is also operationally useful, as will be seen later, since the amount of labor needed to calculate D for all pairs of loci may be prohibitively great, while for any value of n, V_H can readily be estimated.

Linkage disequilibrium and heterozygote advantage: We may, in general, consider a population with a large number of linked heterotic loci, each of which is approximately at equilibrium. A single unselected locus is also present amongst these at some intermediate frequency. It is immediately evident that the unselected locus is not free to fluctuate in frequency to the same extent as is an unselected locus not similarly surrounded by selected loci. Any fluctuation in frequency of the unselected locus will almost certainly be accompanied by fluctuation of frequency at many other loci. These will tend to return to their equilibria, taking the unselected locus back towards its former frequency.

A two locus model may be used to illustrate the argument in more detail. The same two locus model as depicted in Table 1 will be considered, assuming, in addition, that the genes at the A locus are selectively neutral while those at the B locus show heterozygote advantage, the selective values of the genotypes BB,

Bb and bb being, respectively, $1-s : 1 : 1-t$. The B gene is assumed to be at equilibrium, so that $p_B = t/s+t$.

The joint segregation of the genes A and B over one generation will be studied in two stages. Beginning with the frequencies of the two genes in mature individuals of the first generation (Table 1), we consider for a start frequencies in the zygotes of the next generation. Up to this stage there is assumed to be no selection, so that only chance fluctuations inherent in segregation and the mating system are of importance. The point of particular interest is to determine how much the fluctuation of frequency at the A locus will be accompanied by fluctuation at the B locus.

Selection is assumed to occur in the stage between zygote and mature individual. Selection affects directly only the genotype at the B locus, and since the B gene is assumed to be at equilibrium initially, selection will be such as to oppose the effect of the random fluctuation and drive the frequency of the gene back towards its equilibrium. Since the A and B genes fluctuated in frequency together in the first stage, the net effect of selection will be to drive the A gene back towards its previous value, also in opposition to the effect of random fluctuation. In considering the effect of selection on the frequency at the A locus, it should be noted that since selection affects directly only the B genotype, the relative frequencies of the A genotypes amongst a particular B genotype, say BB, will not be changed by selection. Therefore, at the gametic level, the relative frequencies of, for example, the gametes AB and aB, will be unaltered by selection. However, the *overall* frequency of the A gene will not remain unaltered, due to the non-random association of the A and B genes. The argument in detail follows.

To estimate the probable effect of random fluctuation, we assume the FISHER-WRIGHT model of binomial sampling, or in this case multinomial sampling. In the generation following that depicted in Table 1, the frequencies of the four zygotes before selection, x'_1, x'_2, x'_3 and x'_4, respectively, are thus generated by the multinomial $(x_1+x_2+x_3+x_4)^{2N}$, where N is again the population size. Here we ignore the effects of recombination, but this has already been taken into account to a large extent in the derivation of the expected disequilibrium. As a measure of the amount by which the fluctuations at the A locus are accompanied by fluctuation at the B locus, we calculate the covariance of the frequencies of the genes A and B. This comes to $\mathrm{Cov}\,(x'_1+x'_2, x'_1+x'_3)$

$$= \mathrm{V}(x'_1) + \mathrm{Cov}(x'_1,x'_2) + \mathrm{Cov}(x'_1x'_3) + \mathrm{Cov}(x'_2,x'_3)$$

$$= \frac{1}{2N}\left[\, x_1\,(1-x_1) - x_1x_2 - x_1x_3 - x_2x_3\right]$$

$$= \frac{1}{2N}\left[x_1x_4 - x_2x_3\right] = \frac{D}{2N}\,.$$

Using regression theory, the expected change at the B locus is equal to the change at the A locus $\times \dfrac{\mathrm{Cov(A,B)}}{\mathrm{V}ce\mathrm{(A)}}$. Thus, if the frequency at the A locus changes by an amount ε, the expected change at the B locus is equal to $\varepsilon \cdot \dfrac{D}{p_A\,(1-p_A)}$.

This may, of course, be either positive or negative depending on the sign of D.

We now consider the effect of selection. Let us assume that the effect of selection is to change the frequency of the B gene by an amount, δ, which may be either positive or negative. The frequency before selection is equal to $x'_1 + x'_3$, and after selection to $x'_1 + x'_3 + \delta$. The frequencies of the gametes AB and aB after selection may be written as $x'_1 + \gamma.\delta$ and $x'_3 + (1-\gamma)\delta$. But as previously mentioned, the relative frequencies of the two gametes remain unaltered by selection, so that

$$\frac{x'_1 + \gamma.\delta}{x'_3 + (1-\gamma)\delta} = \frac{x'_1}{x'_3}$$

from which

$$\gamma = \frac{x'_1}{x'_1 + x'_3}.$$

Similarly it is shown that the frequencies of the gametes Ab and ab after selection are, respectively, $x'_2 - \delta.\dfrac{x'_2}{x'_2 + x'_4}$ and $x'_4 - \delta.\dfrac{x'_4}{x'_2 + x'_4}$.

The overall frequency of the A gene after selection is, thus

$$\begin{aligned} &x'_1 + \delta.\frac{x'_1}{x'_1 + x'_3} + x'_2 - \delta.\frac{x'_2}{x'_2 + x'_4} \\ &= x'_1 + x'_2 + \delta.\frac{x'_1 x'_4 - x'_2 x'_3}{(x'_1 + x'_3)(x'_2 + x'_4)} \\ &= p'_A + \delta.\frac{D'}{p'_B(1-p'_B)}, \text{ where } D' = x'_1 x'_4 - x'_2 x'_3. \end{aligned}$$

Further assumptions must now be made to specify the magnitude of δ in terms of the magnitude of the chance fluctuations described previously. As a first approximation we assume that selection acts to bring the frequency of the B gene back to its equilibrium within a single generation. Then $\delta = -\varepsilon.\dfrac{D}{p_A(1-p_A)}$.

Therefore, the frequency of the A gene becomes

$$p'_A - \varepsilon.\frac{D}{p_A(1-p_A)} \cdot \frac{D'}{p'_B(1-p'_B)}$$

If we now make the approximation that $D' = D$ and $p'_B(1-p'_B) = p_B(1-p_B)$ then the frequency of A after selection reduces to

$$p_A + \varepsilon\left[1 - \frac{D^2}{p_A(1-p_A)p_B(1-p_B)}\right].$$

Thus, the magnitude of the chance oscillations at the A locus can be expected to be reduced by a proportion $1 - \dfrac{D^2}{p_A(1-p_A)p_B(1-p_B)}$.

The magnitude of the selective values at the B locus has so far not entered into the argument. The return to equilibrium at the B locus will, however, only be achieved in one round of selection if $s = t = 1$. In general, the smaller the values of s and t, the longer the B locus will take to respond to any fluctuation from its equilibrium. During this time the association between the A and B genes will be affected by recombination, and the stabilizing effect of the B locus on the A locus will be diluted.

It is, of course, very artificial to try to describe the effects of selection in a single generation. Ideally this should be considered jointly with the distribution of D as a continuous process, but the problems involved in such a treatment are formidable. Nevertheless, the above arguments have shown that there will be some stabilizing effect of a heterotic locus on a selectively neutral linked locus, whose magnitude is dependent on the amount of linkage disequilibrium generated by chance segregation. However, no quantitative statement about the long-term magnitude of the effect can be given from these calculations. Nor is it obvious how the segregation of the unselected locus is constrained by more than one linked heterotic locus.

One point which might be made here is that the arguments given so far are really directed at gene frequency stabilizing mechanisms in general rather than just at heterozygote advantage. Evidence for an alternative mechanism of maintaining stable gene frequencies has come from experiments of PETIT (1958) and EHRMAN (1967) on frequency-dependent mating success, and KOJIMA and YARBROUGH (1967) on frequency-dependent viability. This explanation of the maintenance of variability possesses one intuitive advantage over explanations based on heterosis in that the consequences of inbreeding are expected to be very severe under the latter hypothesis but not necessarily under the former.

The apparent selective value: An alternative approach to describe the effect of a heterotic locus on an unselected locus is to examine the apparent selective value at the unselected locus. With the gametes as given in Table 1, and selective values $1-s : 1 : 1-t$ associated with the B locus, the array of zygotes and selective values may be written down as in Table 2. Although the A locus is itself unselected, it will appear to have a selective effect due to its association with the B locus. The relative selective value of the AA genotype is

$$\frac{x^2_1(1-s)+2x_1x_2:1+x^2_2(1-t)}{x^2_1+2x_1x_2+x^2_2}.$$

$$=1-s\cdot\frac{x^2_1}{(x_1+x_2)^2}-t\cdot\frac{x^2_2}{(x_1+x_2)^2}$$

Similarly the selective value associated with the Aa genotype is

$$1-s\cdot\frac{x_1x_3}{(x_1+x_2)(x_3+x_4)}-t\cdot\frac{x_2x_4}{(x_1+x_2)(x_3+x_4)},$$

and with aa

$$1-s\cdot\frac{x^2_3}{(x_3+x_4)^2}-t\cdot\frac{x^2_4}{(x_3+x_4)^2}.$$

With arbitrary values of x_1, x_2, x_3 and x_4 little can be said about the relative magnitudes of these three quantities, which may be labelled as $1-S_{AA}$, $1-S_{Aa}$ and $1-S_{aa}$ respectively. However, if the B locus is at equilibrium, some interesting relationships emerge. The quantity $S_{AA}-S_{Aa}$, for example, is equal to

$$s\left[\frac{x^2_1}{(x_1+x_2)^2}-\frac{x_1x_3}{(x_1+x_2)(x_3+x_4)}\right]+t\left[\frac{x^2_2}{(x_3+x_4)^2}-\frac{x_2x_4}{(x_1+x_2)(x_3+x_4)}\right]$$

$$=\frac{x_1x_4-x_2x_3}{(x_1+x_2)^2(x_3+x_4)}\,(sx_1-tx_2).$$

If the B gene is at equilibrium, then

$$x_1 + x_3 = \frac{t}{s+t}$$

so that

$$(x_1 + x_3)(s+t) = \mathrm{t}$$

and

$$(x_1x_2 + x_2x_3)(s+t) = tx_2 .$$

Similarly

$$(x_1x_2 + x_1x_4)(s+t) = sx_1 ,$$

and subtracting,

$$sx_1 - tx_2 = (x_1x_4 - x_2x_3)(s+t).$$

Thus $S_{AA} - S_{Aa} = \frac{D^2(s+t)}{p^2{}_A \cdot p_a}$, and is, therefore, always positive.

Similarly $S_{aa} - S_{Aa} = \frac{D^2(s+t)}{p^2{}_a \cdot p_A}$.

Therefore, at the A locus the Aa genotype will, temporarily at least, have a selective advantage over both homozygotes. Furthermore, since $(S_{AA} - S_{Aa})/(S_{aa} - S_{Aa})$ is equal to p_a/p_A, the selective values are such as not to change the frequency of the A gene from its present value. A similar argument can be made for all heterotic genes linked to the A locus. Then if, for example, there is an additive interaction of selective values for all of these loci, the effect from each could be summed to give an overall effect on the unselected locus. In systems with large numbers of closely linked loci the overall value could be a large one.

There are obvious similarities between these results and those derived under the earlier approach of this section. Both indicate that the effect of the heterotic locus will be to give a short term stability to the unselected locus. The unselected gene is at what might be described as a pseudo-equilibrium, which possesses some, but not all, of the properties of a classical heterotic equilibrium. For example, the selective values associated with an unselected gene starting at 50% frequency will tend to oppose any fluctuation from 50%. However, the gene frequency could over some period of time decline to 25%, at which point a pseudo-equilibrium might be reached tending to keep the gene frequency from either declining further or from returning to its former value.

At any point during this process, the measurement of selective values would tend to reveal that the unselected gene was being maintained at equilibrium by a quite strong heterozygote advantage. This figure would, however, be somewhat illusory.

COMPUTER SIMULATION

In order to investigate numerically the effects discussed so far with realistic selective values, recombination frequencies and numbers of loci, it seems necessary to turn to computer simulation. The aim of the study made was to repeat the type of simulation described previously with a twenty locus model, but allowing for recombination and as high a number of loci and population size as pos-

sible. The largest number of loci found to be economically feasible (in terms of computer time) was 180. Similarly, although a population size of 1,000 or more would be a realistic, it was not possible to simulate populations larger than 100. The 180 loci were assumed to lie equally spaced along a chromosome segment of 18 map units. This was an attempt to simulate a portion of a chromosome of 100 map units with 1,000 segregating loci. Each locus was started with two alleles at 50% frequency. Some runs were made without selection, and in other runs selective advantages of 1% and 0.5% were assigned to each heterozygote.

Description of computer programs: The programs were written for the IBM 7090 computer. The binary structure of this machine makes it suitable for this type of simulation, since each word consists of 36 bits, each of which may represent a locus with two alleles. The use of five words thus suffices for simulating a chromosome of 180 loci. The programs were written to minimize wherever possible the manipulation of individual bits.

The population was set up with the two alleles at each locus having frequencies of exactly 50%. Two types of population were run. In the first, genes at each locus were assigned independently of those at all other loci, thereby giving approximate linkage equilibrium at all pairs of loci. In the second, only two chromosome types, each at 50% frequency, were used, thereby starting with complete linkage disequilibrium between all pairs.

In the initial step, one out of the N (50 or 100) monoecious parents was selected at random, and a gamete produced from this. A very simple recombination procedure was used, allowing for either zero or one crossovers per gamete. In the event that there was a crossover it was assigned to a position on the chromosome segment at random. The choice of which of the two chromosome segments went to the gamete was also made at random. The procedure was then repeated to produce a second gamete. In some cases this was done with replacement of the first parent and in other cases without. This appeared to have little effect on the final outcome.

The selective value of the zygote produced was examined at this stage. This was done by adding up the total number of loci at which the individual was heterozygous, and then calculating the selective value assuming a multiplicative interaction of selective values. On the basis of a random number calculation this zygote was then rejected, or alternatively accepted in which case it became the first individual of the next generation. Summing the number of heterozygous loci was the most time-consuming feature of the program, but the use of machine assembly routines for Logical Operations (see FRASER 1957) and Convert Instructions reduced the time necessary for this step.

Some comment on the use of multiplicative interactions, and indeed on the practice of imposing a fixed selective value at each locus, might be made here. This was done purely for convenience, and under the supposition that with relatively small selective values the epistatic component would not be important. Some attempt was made to write a program which utilized truncation selection, which would give a different type of interaction and which possesses the ad-

vantage that fewer zygotes would have to be discarded to achieve a given selective differential. However, difficulties in calculating the exact heterozygote advantage attained at each locus led to the discarding of this method.

The entire procedure described above was repeated until N individuals had been accepted. The new population then replaced the old, and another generation was started. One weakness of using only a single bit to represent each locus was manifested at this stage, when considerable extra calculation was required just to determine gene frequencies at individual loci. Fortunately it was not vital that this censusing be done in every generation, since for most purposes only the long-term trends were of interest. A complete calculation of gene frequencies was usually made only once in every twenty generations. As a compromise the frequencies at five loci spread over the chromosome were calculated in each generation. Frequencies of the three genotypes at these loci were in fact calculated before and after selection, thereby allowing calculation of the selective coefficients at each locus.

Time of fixation: Some thirteen runs were made altogether, a number up to 1,000 generations, but some only up to 100 generations. The most useful index of the instability of the system, the number of loci fixed, is summarized in Table 3. Also included in this table for comparison are two sets of runs made with one locus populations. Only one hundred of these populations were in fact run, and the numbers in the table have been adjusted to give results comparable with the 180 locus populations.

The most striking comparisons are those of the initial generations. In the single locus populations, both selected and unselected, approximately 50% of the loci

TABLE 3

Number of loci fixed by particular time. N = *population size,* s = *selective coefficient,* D = *linkage disequilibrium start,* E = *linkage equilibrium start*

			Time in generations					
N	*s*	*D* or *E*	20	40	100	200	500	1000
100	.01	*D* (1)	..	..	..	..	11	53
		(2)	..	..	..	..	21	..
		E (1)	..	..	5	28	67	98
		(2)	..	2	8	27	60	..
50	.01	*D* (1)	..	..	8	40	116	167
		(2)	..	..	..	1	47	114
		E (1)	..	3	37	61	121	..
		(2)	.	4	28	..	..	..
50	.005	*D*	..	..	1	53	126	..
		E	1	2	36	85	169	..
50	0	*D* (1)	1	21	67	..	..	..
		(2)	1	18	94	..	..	..
		(3)	9	23	122	..	..	..
50	.01	Single	..	22	101	166	176	180
50	0	Locus	..	14	83	148	176	180

had fixed by 100 generations. However, fixation by this time was relatively rare in the multiply heterotic systems.

Although there are quite large discrepancies between replicate runs, nevertheless it is clear that the runs started with complete linkage disequilibrium are more stable than those started with linkage equilibrium. Both of these starting conditions are evidently unrealistic. A realistic starting condition would be somewhere intermediate.

One hundred is a very small population size; it seems doubtful whether there would be many examples of populations surviving over any length of time with such small numbers. In any case it is clear that 1% is not sufficient to ensure the stability over any length of time of a linked system of heterotic genes under such conditions. This is particularly true when the possibility of asymmetrical selective values is taken into account. ROBERTSON (1962) has shown for a single locus that the probability of fixation increases sharply for asymmetrical selective values. While the present computer program has been modified to treat asymmetrical selective values which are the same at each locus, few results have been derived for this case yet. But there seems every reason to believe that fixation would be more rapid in such cases.

It should be noted that a completely stable system is not to be expected with a finite population size. The only reasonable criterion for stability is that the rate of loss of polymorphisms be sufficiently small to be compensated by the production of new polymorphisms. Some information on possible rates of production of new polymorphisms is available (see MUKAI, YOSHIKAWA and SANO 1966), but it seems unlikely that the rate could be sufficient to balance the losses observed in even the most stable of the computer populations.

The magnitude of linkage disequilibrium: As previously mentioned it was not possible to calculate in every generation the gene frequencies at every locus, much less the amount of linkage disequilibrium between all pairs of loci. For 180 loci there are $\frac{1}{2}(180 \times 179)$ or 16,110 pairs of loci, each of which for a population of size 50 involves classifying 100 gene pairs. However, it proved feasible to make the entire calculation in a few cases, and the results of one such calculation are displayed in Figure 1. In this figure the mean value of D^2 for loci separated by a given map distance has been plotted. The left-most point represents the average value of D^2 for the 179 pairs of loci separated by 0.1 map units, while the right-most point represents the value of D^2 for the outermost pair of loci separated by 17.9 map units.

The population in which these calculations were made had been started 40 generations previously with linkage equilibrium. As mentioned previously, even assigning genes at random to individual loci does not produce exact linkage equilibrium. It is readily shown that a mean value of D^2 of $1/32N$, or .0006 in this case, is expected for each pair of loci in generation 0, and this value is indicated in the figure. Clearly linkage disequilibrium has been built up, even between loci separated by nearly 18 map units, and since no selection was applied in this run, the build-up of disequilibrium must be purely as a result of random fluctuation.

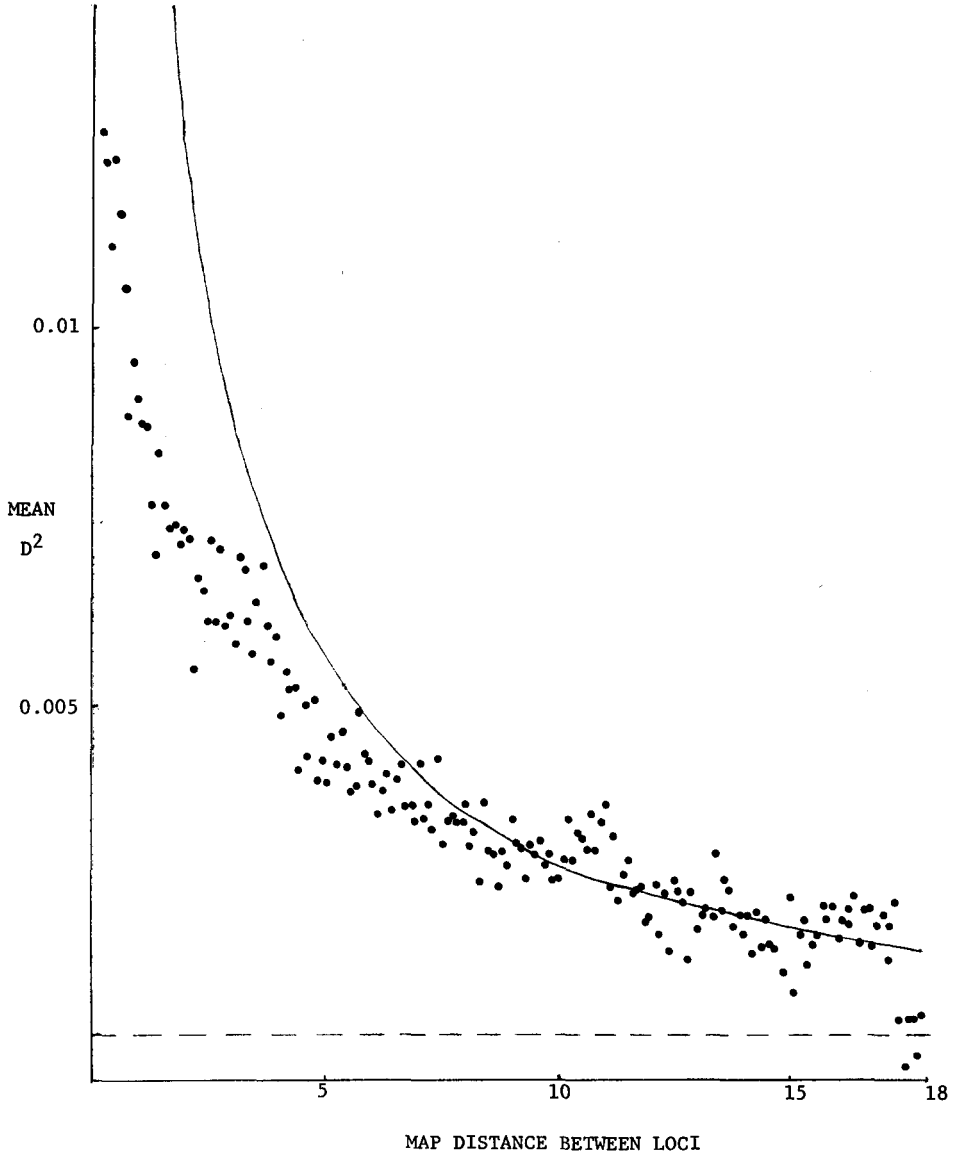

FIGURE 1.—Relationship between linkage disequilibrium and map length in population of size 50 after 40 generations without selection. The curved line gives the expected equilibrium value for D^2, while the broken line gives the expected starting value.

The expected steady state value of D^2 derived earlier, $1/16(4Nr+1)$, has also been plotted in Figure 1. For loci separated by ten or more map units this value seems to have been reached. Clearly, it has not been for the most closely linked loci. Forty generations is evidently not sufficient for the rather high predicted disequilibrium to be attained. Unfortunately it was not possible to study the build-up of disequilibrium any further, since after 40 generations the effects

TABLE 4

Summary of linkage disequilibrium in unselected population of size 50. See text for explanation of terms

Generation	$8 \Sigma D^2$	$16 \Sigma D(\frac{1}{2}-p_A)(\frac{1}{2}-p_B)$	V_H
20	553.8	29.0	565.6
40	660.4	53.8	693.2
60	212.8	216.7	445.0
80	75.3	23.9	115.1
100	42.4	19.8	66.5

of fluctuation of gene frequencies away from 50% began to become apparent, and the values of D^2 declined rapidly. This trend is seen in Table 4, where the values of $8 \Sigma D^2$ over all pairs of loci are given in column 2 at 20, 40, 60, 80 and 100 generations.

These calculations enable us to check on the accuracy of formula (2). Columns 2 and 3 give the values of $8 \Sigma D^2$ and $16 \Sigma D(\frac{1}{2} - p_A)(\frac{1}{2} - p_B)$. The value of V_H given in column 4, which is actually the average over a few generations surrounding the stated generation number, is in most cases close to the value of $8 \Sigma D^2$. The principal exception is at 60 generations, where the term in D makes an appreciable contribution to V_H. What appears to be happening is that when gene frequencies are changing rapidly within a certain range there is a tendency for $(\frac{1}{2} - p_A)(\frac{1}{2} - p_B)$ to be positively correlated with D, and the value of V_H does not drop immediately to the same extent as does $8 \Sigma D^2$. Fortunately there is not very much interest in knowing the exact rate of decrease of the D values, and formula (2) appears to be quite accurate in cases where the values of D are not falling rapidly. In particular, where complete calculations have been made in populations with heterozygote advantage, the contribution of the D term has not been more than 5–10% of the total. In general, therefore, the statistic V_H should usually give a good measure of the overall amount of linkage disequilibrium.

Values of V_H estimated from the first 100 generations of a number of runs starting with linkage equilibrium are given in Table 5. Two population sizes were run, with and without selection, and the values given in the table are averages from three replicate runs made with each combination. The value of V_H fluctuated severely from one generation to the next. This does not necessarily mean that the actual D values fluctuated severely, but is more likely due to the lack of precision in estimating the infinite population variance using a population of small size.

Some trends are evident from the table. The overall disequilibrium is, as expected, higher in the smaller populations than in the larger. The build-up of disequilibrium also appears to be slightly faster in the unselected than the selected populations, which is perhaps a manifestation of the slower rates of change in the more stable selected populations. The disequilibrium begins to drop off rapidly as the gene frequencies drift from 50% in the unselected populations.

TABLE 5

Variance of the number of heterozygous loci per individual (V_H)

Generation	Population size 50		Population size 100	
	No selection	$s = .01$	No selection	$s = .01$
1	136.8	66.2	95.4	91.7
2	234.6	119.3	112.4	135.5
3	443.5	346.2	153.5	95.6
4	476.3	251.6	148.1	110.6
5	525.8	158.9	126.0	149.1
6–10	511.9	415.4	255.0	234.5
11–15	700.8	415.0	280.2	288.9
16–20	762.2	455.9	346.7	284.1
21–25	646.5	519.5	424.7	302.8
26–30	647.7	559.1	402.7	372.9
31–40	586.8	544.3	439.3	346.3
41–50	573.2	517.3	396.9	358.2
51–60	502.2	575.1	398.1	361.3
61–70	425.4	603.8	362.2	397.8
71–80	330.4	550.3	315.5	398.6
81–90	248.4	520.1	300.6	391.9
91–100	243.5	535.7	261.2	385.6

As discussed above, the drop is probably faster than indicated by the fall in V_H values. The fall was also noticeable in the selected populations, and continued slowly but steadily after 100 generations.

Estimation of selective values: Some check of the arguments given earlier about the apparent heterozygosity is possible from the computer results. For one run of 500 generation with 1% heterozygote advantage and a population size of 100, results are given in Table 6. In this table, selective coefficients calculated using genotype frequencies before and after selection, are given for the five loci spread over the chromosome segment for which complete gene frequency data were obtained. Treating each locus separately, selective values have been pooled for each generation where the gene frequency lay in a particular range. Line 1, column 3, for example, gives the relative selective coefficient against the rarer homozygote as 19.4%, averaged over the 28 generations during which the gene

TABLE 6

Percentage selective coefficients s *and* t *against homozygotes, at various gene frequencies, based on* n *observations*

Locus	Gene frequency 0.2–0.3			0.45–0.55			0.7–0.8		
	n	*s*	*t*	*n*	*s*	*t*	*n*	*s*	*t*
1	28	19.4	5.9	148	7.7	8.5	14	2.0	3.6
2	120	14.8	5.9	100	9.8	10.4	6	7.3	12.3
3	0	..	..	38	9.6	13.1	112	7.3	18.1
4	146	8.6	4.2	21	7.4	—0.5	2	—5.9	21.3
5	13	20.3	9.8	75	9.1	10.1	62	5.0	4.7

frequency at the first locus lay in the range 0.2 — 0.3, i.e. 0.25 ± 0.05. In all but one case, the selective coefficient against the rare homozygote is higher than against the more frequent homozygote. The overall selective coefficients against the rare and frequent homozygotes are, respectively, 12.6% and 5.6%. One percent must be deducted from each of these to take account of the selection at the loci in question. The resulting values are nearly in the ratio 3 : 1 expected for an equilibrium situation. The selective coefficients against the two homozygotes when the genes are at central frequencies are in most cases approximately equal, and the mean selection coefficient is 9.0%.

The estimated selection coefficients show considerably more variation than would be expected from constant selective values at a particular locus. Evidently the large variation in association between different genotypes is responsible for this. The information in Table 6 is based on nearly one hundred thousand observations, with complete tabulation of frequencies before and after selection. Even ignoring the fact that the selection coefficient may be a misleading parameter when estimated over a long period of time, it is seen that prohibitively large samples are needed to recognize the overall pattern of heterozygote advantage.

DISCUSSION

The fact that the selective values holding a particular locus heterozygous may not be wholly attributable to genes at that locus has been clearly recognized by a number of authors (e.g. HEXTER 1955; FRYDENBERG 1963; POLIVANOV 1964; MERRELL 1965; WALLACE 1966). In addition, as WALLACE emphasizes, when there is epistasis it may not be possible conceptually to draw a line between the selective forces attributable to genes at a locus and at surrounding loci. To some extent, the arguments of this paper merely provide a quantitative statement of these views. But what has perhaps not been sufficiently emphasized previously is that epistasis is not necessary for heterozygote advantages to reinforce each other, and that *any* linkage disequilibrium will contribute towards this reinforcement. It is for this reason that even the relatively small amounts of linkage disequilibrium generated by chance may be of importance, particularly when there are many linked loci each with heterozygote advantage.

An interesting point is the inverse relationship between population size and magnitude of linkage disequilibrium. As the population size decreases, thereby increasing the probability of chance fixation, this is compensated to some extent by the increased disequilibrium. At one extreme, when the population size is infinite, the reinforcement discussed above disappears. Of course, it seems doubtful whether in this case it is necessary to postulate heterozygote advantage, or even a reinforcing of this advantage, to account for the maintenance of variability. At the other extreme, the effect is highest in populations of very small size. The smallest population possible is of size one, which implies inbreeding through self-fertilization. The effect is well documented in this case, because it is obvious that under such conditions loci become homozygous in blocks and not independently of each other (FISHER 1949; FRANKLIN unpublished).

The effect could be of importance in laboratory populations, which are usually of limited size. From one point of view this is unfortunate, since it makes it difficult to attribute selective values to genes at a particular locus, even if all possible precautions have been taken to ensure linkage equilibrium. If heterosis is demonstrated for a particular marker in a laboratory experiment, then as a preliminary hypothesis the marker should really be regarded as an indicator rather than as a cause of heterosis.

If the above description is at all accurate, it seems hard to avoid the implication that in populations of small size the major factor responsible for keeping any locus polymorphic is not the selective values of the genes at that locus but the selective values at surrounding loci. The stability is a property of the system as a whole, and it would perhaps be best to consider the stability of a chromosome region rather than of a locus. Some confirmation of this view was inadvertantly obtained when it was found that in some of the computer programs one of the 36 loci represented by each word was mistakenly ignored in the calculation of selective values, making it selectively neutral. Yet there was no obvious tendency to fix this locus before any of the other 35. Thus it would scarcely seem to matter if a number, or even a large proportion of loci in a multiply heterotic system are unselected. If the estimate of 20,000 segregating loci per Drosophila chromosome is correct, the computer results do not altogether rule out the possibility that only one in twenty of these loci could be selected and, nevertheless a closely linked system of heterotic loci be produced. LEWONTIN and HUBBY (1966) have discussed the possibility that the formation of hybrid enzymes might be a general mechanism for producing heterozygote superiority. While hybrid enzymes may only be formed for a proportion of enzymes, nevertheless this might be sufficient to ensure a relative degree of stability for selectively neutral or near-neutral segregating genes anywhere in the genome.

From a long term point of view, this account seems rather unsatisfactory. It seems scarcely feasible that such capricious selective forces could account for the apparent long term stability of such polymorphisms as ABO. However, it might be speculated that the combination of selective values affecting genes at the locus in question plus the effect of outside loci might be a potent force for stability. Short term fluctuations, which could be of particular importance at times of reduced population size, are reduced by the latter effect, while the former acts continuously and in a systematic manner to bring the gene to intermediate frequencies.

A closely related question concerns the manner in which a gene might reach high frequency without being itself selectively advantageous. That this might happen at times is clearly suggested by the results of WALLACE (1966). The *sepia* gene of *D. melanogaster*, which was found at low frequency in one population, was introduced into three other genetic backgrounds, and appeared to rise to high and characteristically different frequencies in each case. This type of increase in frequency is closely analogous to the hitch-hiking effect postulated by KOJIMA and SCHAFFER (1967).

LEWONTIN (1964) has previously predicted the build-up of linkage disequilibrium, but for rather different reasons than considered in the present paper. The populations considered by LEWONTIN were large, and the linkage disequilibrium arose as a result of the interaction of selective values at different loci. Although a similar type of interaction (multiplicative) has been used in the present arguments, this does not appear to be the cause of the disequilibrium here. The two types of disequilibrium are of a rather different type. The present paper is focussed on small populations and quite short term changes, and the disequilibrium is of a highly fluctuating and unpredictable type. The type envisaged by LEWONTIN appears to be stable over much longer periods of time and to be most likely in large populations. Both types could well be of importance in different situations.

Artificial selection experiments: The arguments put forward in the paper are compatible with an observation often made in experiments on artificial selection of quantitative characters. A plateau is reached, and some regression may occur on relaxation of selection. After a period of relaxation, renewed selection may lead to a response which causes the previous plateau to be exceeded. This sequence of events is very similar to what might be predicted on the basis of the two locus model discussed previously. LERNER (1954) argued for the existence of gene frequency stabilizing mechanisms from these observations. Taking into account also data from inbreeding experiments, he concluded that there might be heterozygote advantage at the loci controlling the trait under selection. This argument may be strengthened if not only the genes controlling the character but also other linked genes are being held at equilibrium by selection. If genes controlling the character are located at many places throughout the genome, the initial effect of selection might be to change gene frequencies at a large percentage of all loci. This could account for the severity with which natural selection at times opposes artificial selection. These arguments cannot, however, be adduced as a proof of the importance of linkage, since as has often been pointed out the consequences of physiological linkage may be similar to those of genetic linkage.

I am indebted to Drs. W. F. BODMER and R. C. LEWONTIN for a number of suggestions. I would also like to express my appreciation to Mrs. JULIANA HWANG for assistance with the computing.

SUMMARY

When there are a number of closely linked loci and the population size is small, the tendency for recombination to cause nonallelic genes to become associated at random (linkage equilibrium) will be opposed by chance fluctuations tending to produce linkage disequilibrium. This disequilibrium is unpredictable and usually not of great importance for any pair of loci, but summed over a large number of pairs it has some recognizable patterns and important consequences. If there is heterozygote advantage, or any gene frequency stabilizing mechanism, at these loci, then any disequilibrium will lead to the selective values at different loci reinforcing each other. The net result is that in a small popula-

tion the selective values at surrounding loci could be considerably more important for the stability of a particular polymorphism than the selective values at the locus itself.

LITERATURE CITED

BODMER, W. F., and J. FELSENSTEIN, 1967 Linkage and selection: theoretical analysis of the deterministic two locus random mating model. Genetics **57**: 237–265.

BRIDGES, P. N., 1942 A new map of the salivary gland 2L-chromosome of *Drosophila melanogaster*. J. Heredity **33**: 403–408.

CARLSON, E. A., 1967 Corrigenda. Genetics **57**: 970.

EHRMAN, L. R., 1967 Further studies on gene frequency and mating success in Drosophila. Am. Naturalist **101**: 415–424.

EWENS, W. J., 1964 The maintenance of alleles by mutation. Genetics **50**: 891–898.

FISHER, R. A., 1949 *The theory of inbreeding*. Oliver and Boyd. Edinburgh.

FRASER, A. S., 1957 Simultation of genetic systems by automatic digital computers. I. Introduction. Australian J. Biol. Sci. **10**: 484–491.

FRYDENBERG, O., 1963 Population studies of a lethal mutant in *Drosophila melanogaster*. I. Behaviour in populations with discrete generations. Hereditas **48**: 89–116.

HARRIS, H., 1966 Enzyme polymorphisms in man. Proc. Roy. Soc. London B **164**: 298–310.

HEXTER, W. M., 1955 A population analysis of heterozygote frequencies in Drosophila. Genetics **40**: 444–459.

HILL, W. G., and A. ROBERTSON, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8**: 269–294.

JOHNSON, F. M., C. G. KANAPI, R. H. RICHARDSON, M. R. WHEELER, and W. S. STONE, 1966 An analysis of polymorphisms among isozyme loci in dark and light *Drosophila ananassae* strains from America and Western Samoa. Proc. Natl. Acad. Sci. U.S. **56**: 119–125.

KIMURA, M., and J. F. CROW, 1964 The number of alleles that can be maintained in finite populations. Genetics **49**: 725–738.

KING, J. L., 1967 Continuously distributed factors affecting fitness. Genetics **55**: 483–492.

KOJIMA, K., and H. E. SCHAFFER, 1967 Survival processes of linked mutant genes. Evolution **21**: 518–531.

KOJIMA, K., and K. M. YARBROUGH, 1967 Frequency-dependent selection at the esterase 6 locus in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S. **57**: 645–649.

LERNER, I. M., 1954 *Genetic Homeostasis*. Oliver and Boyd. Edinburgh.

LEWONTIN, R. C., 1964 The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49**: 49–67.

LEWONTIN, R. C., and J. L. HUBBY, 1966 A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics **54**: 595–609.

LEWONTIN, R. C., and K. KOJIMA, 1960 The evolutionary dynamics of complex polymorphisms. Evolution **14**: 458–472.

LI, C. C., 1955 *Population Genetics*. University of Chicago Press.

MERRELL, D. J., 1965 Competition involving dominant mutants in experimental populations of *Drosophila melanogaster*. Genetics **52**: 165–189.

MILKMAN, R. D., 1967 Heterosis as a major cause of heterozygosity. Genetics **55**: 493–495.

MUKAI, T., I. YOSHIKAWA, and K. SANO, 1966 The genetic structure of natural populations of *Drosophila melanogaster*. IV. Heterozygous effects of radiation-induced mutations on viability in various genetic backgrounds. Genetics **53**: 513–527.

PETIT, C., 1958 La déterminisme génétique et psycho-physiologique de la compétition sexuelle chez *Drosophila melanogaster*. Bull. Biol. France Belg. **42**: 248–329.

POLIVANOV, S., 1964 Selection in experimental populations of *Drosophila melanogaster* with different genetic backgrounds. Genetics **50**: 81–100.

ROBBINS, R. B., 1918 Some applications of mathematics to breeding problems. III. Genetics **3**: 375–389.

ROBERTSON, A., 1962 Selection for heterozygotes in small populations. Genetics **47**: 1291–1300.

SVED, J. A., T. E. REED, and W. F. BODMER, 1967 The number of balanced polymorphisms that can be maintained in a natural population. Genetics **55**: 469–481.

WALLACE, B., 1966 The fate of *sepia* in small populations of *Drosophila melanogaster*. Genetica **37**: 29–36.