THE EFFECT OF A SELECTED LOCUS ON LINKED NEUTRAL LOCI

GLENYS THOMSON

Geneiics Laboratory, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, United Kingdom

Manuscript received February 10, 1976
Revised copy received October 15, 1976

ABSTRACT

The effects produced on linked neutral loci as a selected locus evolves towards its equilibrium value are considered. Significant effects on the neutral loci arise if the recombination fraction between the neutral and selected loci is smaller than the order of magnitude of the selective differences at the selected locus. The effect on gene frequencies at the neutral loci, that is, the hitchhiking effect, is determined, as well as the linkage disequilibrium generated by this hitchhiking effect. One of the more important findings is that significant disequilibrium can be generated between two neutral loci by the evolution of a linked selected locus. Consideration is given to the problem of determining how the effect of selection operating in natural populations can be detected, the question of the establishment of inversions in populations, and also to the nonequilibrium properties of populations.

THE original models of population genetics deal chiefly with single loci, so that the genome is regarded as a collection of individual independent loci each undergoing its separate evolution. An increasing amount of attention has recently been given to the dynamics of multiple locus systems. The analytic solutions of some special cases of selection acting on two or three loci have been obtained (Kimura 1956; Lewontin and Kojima 1960; Bodmer and Felsenstein 1967; Ewens 1968; Karlin and Feldman 1970; Feldman, Franklin and Thomson 1974). Karlin and Carmeli (1975) have conducted a numerical study on two-locus selection models with general viabilities. For a review of two-locus selection models see Karlin (1975). Simulations of multi-locus systems involving more loci and simple selection schemes have been made (Lewontin 1964a,b; Franklin and Lewontin 1970).

Selection, of course, acts on the whole genome and not on individual loci. Models developed so far do not take account of this, but instead consider the effect of selection acting on a number of individual loci. The present paper is a preliminary attempt to construct a model to take account of the effect of evolutionary forces on the whole genome. This is done by considering the effect produced on linked neutral loci as a selected locus evolves towards its equilibrium value. One can determine, for particular selection values, over what range of recombination

* Research supported by an 1851 Research Fellowship and a Wellcome Trust Travel Grant.

fractions a significant effect is produced on the nearby loci, which are assumed to be selectively neutral. This measures what fraction of the genome is affected as the selected locus is evolving. Consideration will be given to the effect on gene frequencies at the neutral loci, that is, the hitchhiking of genes as recently discussed by Maynard Smith and Haigh (1974), and to the linkage disequilibrium generated. It will be shown that the hitchhiking effect can be a very significant force.

As well as being of relevance to the question of the evolution of the whole genome, the results obtained are also of importance in dealing with the problem of determining how the effect of selection operating in natural populations can be detected. There has been a great deal of interest in this question in recent years, both from the experimental and theoretical sides. The interest extends from the studies of Hubby and Lewontin (1966); Lewontin and Hubby (1966) in Drosophila, and Harris (1966) in man. These were the first studies to measure the extent of polymorphism in natural populations, using the technique of gel electrophoresis. Both studies showed that roughly 30% of the loci considered were polymorphic. These preliminary studies have since been extended considerably and include a wide variety of species (Lewontin 1974). Although it can be argued that the loci which have been considered are not a representative sample of the genome, it is still generally agreed that natural populations are genetically polymorphic at a large proportion of the loci coding for structural genes. Two alternative schools have developed to explain how this variation is maintained in populations. The selection school claims, simply, that this variation is maintained by some form of balancing selection. The other school, the so-called neutral or neoclassical (see for example Kimura and Ohta (1971)), claims that the variants in natural populations are effectively wild-type and that such variation is selectively neutral (Lewontin 1974). The arguments put forward by each side will not be discussed here.

Instead, the problem will be approached from a different point of view and consideration given to the following question. Under a theory of selection, is there any quantity which we would expect to be observed consistently and whose existence would then indicate the action of selection? An obvious candidate for consideration is linkage disequilibrium, which throughout the following will often be referred to as $D$. (Linkage disequilibrium is the difference between the frequency of a gametic type and the product of the frequencies of the two alleles, so is a measure of nonrandom association of alleles.) The existence of linkage disequilibrium is usually taken as strong evidence that selection is acting. This comes from consideration of the theoretical aspects of the interaction of selection and recombination in the maintenance of linkage disequilibrium. It has only been possible to solve the equilibrium theory of selection on two loci for some very special selection schemes. These are when the effects of the two loci are additive or multiplicative, and a fitness scheme termed the symmetric viability model (Kimura 1956; Lewontin and Kojima 1960; Bodmer and Felsenstein 1967; Ewens 1968; Karlin and Feldman 1970). The results may be summarized as
follows. For these particular selection schemes, the population usually goes to an equilibrium in linkage equilibrium, that is, $D = 0$. In certain cases, when the fitnesses are of the multiplicative or symmetric viability form, the population can go to an equilibrium with $D \neq 0$. The stability criterion for these equilibria is that linkage must be sufficiently tight (the degree of linkage required is a function of the fitness parameters). There is a balance between selection and recombination, and if linkage between the genes is sufficiently tight, then the population will maintain these genes together in a complex. Franklin and Lewontin (1970) conducted numerical simulations of multilocus selection schemes and showed that the selection could maintain complexes of genes together for sufficiently tight linkage. These results led to the belief that if selection is operating, then one would expect to find nonrandom association of closely linked genes, that is $D \neq 0$.

There have been a number of studies looking for linkage disequilibrium. Most of the associations found in Drosophila have involved inversions (Prakash and Lewontin 1968; Prakash and Lewontin 1971; Mukai, Mettler and Chigusa 1971; Prakash and Merritt 1972; Prakash and Levitan 1973; Mukai, Watanabe and Yamaguchi 1974; Langley, Tobari and Kojima 1974; Loukas and Krimbas 1975). In Drosophila, very little evidence has been found of significant and consistent linkage disequilibrium when inversions are not involved. Charlesworth and Charlesworth (1973a) considered five loci in Drosophila melanogaster, not associated with inversions, and found some consistent, but very weak, amounts of linkage disequilibrium. One of the most extensive searches for linkage disequilibrium is in the human histocompatibility system (HLA). Significant linkage disequilibrium between certain pairs of antigens is consistently observed (Histocompatibility Testing 1972; Bodmer 1973a; Bodmer 1973b; Bodmer and Bodmer 1973; Histocompatibility Testing 1975; Thomson, Bodmer and Bodmer 1976). Probably the best known example of permanent linkage disequilibrium is found in the color and banding polymorphisms observed in colonies of the snail Cepaea nemoralis (Cain and Sheppard 1954). Fossil evidence shows that the associations have been maintained for very long periods (Clarke 1975). Another well-known example of nonrandom association is found in the complex mimicry pattern of certain butterflies, which is thought to be controlled by a very tightly linked complex of genes (Sheppard 1959). Examples of linkage disequilibrium have been demonstrated in the blue mussel Mytilus edulis (Mitton and Koehn 1973) and in a coastal marine fish, Fundulus heteroclitus (Mitton and Koehn 1975).

It should be remembered that not all selection schemes lead to $D \neq 0$. Also, there may be overlap of stability classes so that quite different equilibrium points can be stable for exactly the same selection and recombination values (Franklin and Lewontin 1970; Feldman, Franklin and Thomson 1974; Christiansen and Feldman 1975). Under the same conditions one population could go to an equilibrium with $D \neq 0$ and another to an equilibrium with $D = 0$. Also, it is possible that pairwise disequilibria may be zero and higher order
interactions nonzero (Feldman, Franklin and Thomson 1974). So even within the framework of the restricted selection models which have been considered, there is no consistency. Nevertheless, it would be of importance if we could determine that the main cause of linkage disequilibrium is selection. Although we would not be able to detect all selection, at least occasionally selected loci would manifest the observable linkage disequilibrium.

In this context we must look at all possible ways that disequilibrium can be created. There are a number of possible mechanisms, apart from selection acting on the loci, which can give rise to significant disequilibrium. These include migration and admixture, inbreeding and random drift effects. It will be shown below that another mechanism can be added to this list. Significant linkage disequilibrium can be generated between neutral loci by a closely linked selected locus evolving towards equilibrium or fixation. The relative importance of each of these forces in creating linkage disequilibrium, and hence the usefulness of $D$ as a measure of selection, will be considered in the discussion section.

In the following, consideration will be given to various aspects of the so-called hitchhiking effect. (It has been pointed out by Professor Harry Harris that it would in fact be more appropriate to call this the hijacking effect. However, since the term hitchhiking is now commonly used, this terminology will be retained). We will first consider a two-locus model where one locus is selected and the other neutral. Case A deals with the situation where a new mutant has recently occurred at the selected (heterotic) locus and the neutral locus is assumed to be polymorphic. We consider the effect of the evolution of the selected locus on the neutral locus in terms of changes in gene frequency, linkage disequilibrium and heterozygosity. We also determine, for particular selection values, over what range of recombination fractions a significant effect is produced on the closely linked neutral loci. In case B, we consider the situation where the selected locus is in the process of evolving towards its equilibrium value when a neutral mutant occurs. The effect produced on closely linked neutral loci as a new mutant, which is at a selective advantage, evolves towards fixation is briefly discussed in case C. This case has been considered by Maynard Smith and Haigh (1974). Their approach will be slightly extended to include the effects on linkage disequilibrium and average heterozygosity.

Consideration is then given to a three-locus model with one locus selected and two loci neutral. The most important finding from this model is that the linkage disequilibrium between the two neutral loci is greatly affected by the evolution of the selected locus. Two cases are considered. The ordering of loci is assumed to be $ABC$. In case D, the selected locus is $B$ and the loci $A$ and $C$ are neutral. In case E, the loci $A$ and $B$ are neutral and $C$ is the selected locus.

**Two Locus Model**

**Case A**: The selected (heterotic) gene is the new mutant.

(i) New mutant haplotype is $ab$. Consideration will first be given to a simple
two locus model with two alleles at each locus. At the first locus the alleles are $A$ and $a$ and at the second $B$ and $b$. There are four gametic or chromosome types possible (also referred to as haplotypes by Ceppellini et al. 1967) namely $AB$, $Ab$, $aB$ and $ab$. Their frequencies are denoted by $x_1$, $x_2$, $x_3$, and $x_4$ respectively. The gamete frequencies can be written in the well-known form

\[
x_1 = p_A p_B + D \\
x_2 = p_A p_b - D \\
x_3 = p_a p_B - D \\
x_4 = p_a p_b + D
\]

where $p_A$, $p_B$ are the frequencies of the alleles $A$ and $B$ respectively, etc. and $D = x_1 x_4 - x_2 x_3$ is referred to as the coefficient of linkage disequilibrium and is a measure of the nonrandom association of alleles. It should be remembered that the maximum value that $D$ can take is a function of the allele frequencies. It is easy to see from (1) that the largest positive value $D$ can take is the minimum of $p_A p_B$ or $p_a p_B$ while the largest negative value is the minimum of $p_A p_b$ or $p_a p_b$. If $p_A = p_B = \frac{1}{2}$, then $D$ can be as great as $\pm 0.25$, while if $p_A = 0.05$ and $p_B = 0.2$, $D$ must lie within the limits $-0.01$ to $+0.04$.

In the present model we will assume that the $A$ locus is selected (heterosis). The selection scheme is given in (2) with $s_1, s_2 \geq 0$.

\[
\begin{array}{ccc}
AA & Aa & aa \\
1-s_1 & 1 & 1-s_2 \\
\end{array}
\]

The $B$ locus is selectively neutral. The transformation relating the gametic frequencies in the next generation to those in the present is given by (3). Primes denote values in the next generation.

\[
\begin{align*}
\tilde{w}x_1' &= x_1 [1-s_1 p_A] - RD \quad (i) \\
\tilde{w}x_2' &= x_2 [1-s_1 p_A] + RD \quad (ii) \\
\tilde{w}x_3' &= x_3 [1-s_2 p_a] + RD \quad (iii) \\
\tilde{w}x_4' &= x_4 [1-s_2 p_a] - RD \quad (iv)
\end{align*}
\]

$R$ is the recombination fraction between the $A$ and $B$ loci and $\tilde{w}$ is the mean fitness ($\tilde{w} = 1 - s_1 p_A^2 - s_2 p_a^2$). At equilibrium the $A$ gene frequencies are given by

\[
\hat{p}_A = \frac{s_2}{s_1 + s_2}, \quad \hat{p}_a = \frac{s_1}{s_1 + s_2}
\]

Hats are used to denote equilibrium values. The mean fitness at equilibrium is $\hat{w} = 1 - s_1 \hat{p}_A = 1 - s_2 \hat{p}_a$.

Suppose a new mutation $a$ has just recently occurred. The frequency of the $a$ gene will be increasing each generation until equilibrium is reached. If we look
at what is happening at the neutral B locus, we see that the gene frequencies there are also changing each generation.

\[ p_{B}' = p_B - \frac{(s_1 p_A - s_2 p_a)}{w} . D . \]  

(5)

\( p_{B}' = p_B \) if \( D = 0 \) or if the A locus is at equilibrium. Otherwise \( p_B \) will change each generation by an amount \( \Delta p_B (= p_{B}' - p_B) \).

\[ \Delta p_B = \frac{\Delta p_A}{p_A p_a} . D \]  

(6)

(as in Sved 1968). The amount of linkage disequilibrium between the A and B loci is also changing each generation.

\[ D' = \left[ \frac{p_A' p_a'}{p_A p_a} - \frac{R}{w} \right] D \]  

(7)

In the present model \( p_a < \hat{p}_a \) since \( a \) is a new mutant. If \( \hat{p}_a < \frac{1}{2} \) then the term \( \frac{p_A' p_a'}{p_A p_a} \) is greater than unity until equilibrium is reached. If \( \hat{p}_a > \frac{1}{2} \) the term \( \frac{p_A' p_a'}{p_A p_a} \) is greater than unity until \( p_A = p_a = \frac{1}{2} \), after which it is less than unity.

So there are two opposing forces. The first term is initially increasing the disequilibrium value, the second term is decreasing it at a rate greater than under neutrality. [If the A and B loci were both neutral then \( D' = (1 - R)D \) (Jenningings 1917).]

Suppose the new mutant \( a \) occurs in a gamete carrying \( b \), so that the new mutant is of type \( ab \). The frequency of the gametic type \( aB \) initially is zero. The occurrence of the new \( a \) mutation creates linkage disequilibrium, since \( a \) is not randomly associated with the alleles at the B locus. This initial disequilibrium is given by

\[ D(0) = p_a(0) p_B(0). \]  

(8)

This value will be quite small as the initial \( a \) gene frequency \( p_a(0) \) is small.

As the \( a \) gene increases in frequency towards its equilibrium value, the \( b \) gene will also increase in frequency, being pulled along with the \( a \) gene, that is, it hitchhikes along. The value of the linkage disequilibrium between the two loci will also be changing each generation. This pulling along of the gene frequencies together increases the value of the linkage disequilibrium, while recombination tends to reduce it. We want to consider the combined effect of these two forces on the linkage disequilibrium.

Throughout most of the following, a symmetric selection scheme has been considered with \( s_1 = s_2 = s \), say, so that the equilibrium frequency of \( a \) is \( \hat{p}_a = \frac{1}{2} \). We consider the effect produced as the \( a \) gene frequency goes from 0.01 to 0.49.

The reason for not considering very small values of \( p_a \) nor values close to the
equilibrium frequency is that the deterministic model overestimates the amount of time spent in these regions (see pages 61 and 62 of Ewens 1969).

It is easy to see from (7) that $D$ can be expressed as

$$D(n+1) = A_n D(n)$$

$$= B_{n+1} D(0)$$

(9)

where $B_0 = 1$ and $B_n = \frac{\sum_{i=0}^{n-1} A_i}{\mathbb{W}(n)}$.

with

$$A_n = \frac{p_A(n+1)p_a(n+1)}{p_A(n)p_a(n)} - \frac{R}{\mathbb{w}(n)} .$$

It then follows from (6) and (9) that the frequency of the $B$ gene can be written as

$$p_B(n+1) = p_B(n) - \frac{\Delta p_a(n)}{p_A(n)p_a(n)} B_n D(0)$$

(10)

In the case under consideration the new mutant is of type $ab$, in which case $D(0) = p_a(0)p_B(0)$ [see (8)]. It thus follows from (10) that

$$p_b(n+1) = p_b(0) [1 - E_n]$$

(11) (i)

and hence

$$p_b(n+1) = p_b(0) + E_n p_B(0)$$

(11) (ii)

where

$$E_n = \sum_{k=0}^{n} \frac{\Delta p_a(k)}{p_A(k)p_a(k)} B_k p_a(0) .$$

Note that $B_n$ and $E_n$ are independent of the gene frequencies at the $B$ locus. They are dependent only on the initial frequencies at the $A$ locus, the selection operating at the $A$ locus and the recombination fraction $R$.

If the recombination fraction is very small, one can get approximate results for changes in the frequency of the neutral gene $b$ and changes in the linkage disequilibrium $D$. We put $R = 0$, so that linkage is absolute. In this case

$$D(n) = \frac{p_A(n)p_a(n)}{p_A(0)p_a(0)} D(0)$$

[from (7)] so with $D(0)$ as given in (8) it follows that $D$ will increase to the value given below in (12); attained when the $a$ gene is near its equilibrium value.

$$D_{\text{max}} \approx \hat{p}_a \hat{p}_b(0)$$

[for $R$ small] .

(12)

This value of $D$ is the maximum possible for these particular allele frequencies (this is because the gametic type $aB$ cannot be formed since linkage is assumed
to be absolute). If the selection scheme given in (2) is symmetric, with $s_1 = s_2 (= s)$, so $\hat{p}_A = \frac{1}{2}$, then $D$ will increase to $\approx 0.25 \, p_B(0)$ when $R$ is small. From (11) it is easy to show that for small values of $R$, the approximate change in the $B$ gene frequency is as given in (13).

$$p_B(n+1) = p_B(0) \left[1-S_a\right]$$

(13)

where $S_a = \sum_n p_n(n)$ and is the total change in the $a$ gene frequency. So, if $a$ is a new mutant and goes to an equilibrium frequency of $\hat{p}_a = \frac{1}{2}$, then the final $B$ gene frequency will be $p_B(n) \approx \frac{1}{2} \, p_B(0) \, [p_B(n) \approx \frac{1}{2} + \frac{1}{2} \, p_B(0)]$, for small values of $R$. In Table 1 the approximate changes in the frequency of the neutral $b$ gene are given, as well as the linkage disequilibrium $D$ between the neutral and selected locus, when the recombination fraction between the two loci is very small and the $a$ gene frequency at equilibrium is $\frac{1}{2}$.

The effect is very strong when the initial frequency of the neutral gene is small. For example, if initially the neutral gene $b$ is at a frequency of 0.05, it will have increased to a frequency of 0.525 by the time the $a$ gene has reached its equilibrium value. Similarly, the disequilibrium will increase from 0.0095 to 0.238. The effect is not as strong for larger values of $p_B$, but still quite significant changes can be produced. Note that the values in Table 1 are independent of the strength of selection. Since the two loci are considered to be absolutely linked, it is only the initial and final values of the $a$ gene frequency which are needed.

We now go on to consider actual changes produced for particular values of the selection parameters and recombination values. The effect produced is, of course, stronger the smaller the recombination value and the stronger the selection. The values of $E$ and $B$ (as defined in (11) and (9)) when the $a$ gene is near its equilibrium value are determined ($p_B = p_B(0) \, [1 - E]$, $D = B \, D(0)$, $p_a(0) = 0.01$, $\hat{p}_a = \frac{1}{2}$ in the cases under consideration). These are given in Tables 2 and 3 respectively for various values of the selection parameter $s$ and recombination fraction $R$. As stated earlier, these values are independent of the initial gene frequencies at the neutral $B$ locus. Note that the values of $D$ when the $a$ gene is near its

**TABLE 1**

*Approximate changes in frequency of the neutral gene $b$ ($p_b$) and linkage disequilibrium $D$, for small recombination values*

<table>
<thead>
<tr>
<th>Initial $p_b$</th>
<th>Final $p_b$</th>
<th>Initial $D$</th>
<th>Final $D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.525</td>
<td>0.0095</td>
<td>0.238</td>
</tr>
<tr>
<td>0.2</td>
<td>0.6</td>
<td>0.008</td>
<td>0.200</td>
</tr>
<tr>
<td>0.5</td>
<td>0.75</td>
<td>0.005</td>
<td>0.125</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9</td>
<td>0.002</td>
<td>0.050</td>
</tr>
<tr>
<td>0.95</td>
<td>0.975</td>
<td>0.0005</td>
<td>0.013</td>
</tr>
</tbody>
</table>

The new selected mutant $a$ gene occurs in gametic type $ab$, with an initial frequency of 0.01 and a final equilibrium frequency $\hat{p}_a$ of 0.5. For very small recombination values, the effects on $p_B$ and $D$ are independent of the strength of selection and dependent only on the initial neutral gene frequency $p_B(0)$ and the final equilibrium $a$ gene frequency $\hat{p}_a$. See text for further details.
ASPECTS OF THE HITCHHIKING EFFECT

TABLE 2

Values of $E$ [as defined in (11)] when the $a$ gene is near its equilibrium value

<table>
<thead>
<tr>
<th>$x$</th>
<th>0.001</th>
<th>0.005</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.484</td>
<td>0.472</td>
<td>0.458</td>
<td>0.357</td>
<td>0.265</td>
<td>0.090</td>
<td>0.037</td>
</tr>
<tr>
<td>0.2</td>
<td>0.476</td>
<td>0.439</td>
<td>0.398</td>
<td>0.195</td>
<td>0.092</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>0.1</td>
<td>0.465</td>
<td>0.394</td>
<td>0.322</td>
<td>0.087</td>
<td>0.029</td>
<td>0.005</td>
<td>0.002</td>
</tr>
<tr>
<td>0.05</td>
<td>0.444</td>
<td>0.319</td>
<td>0.218</td>
<td>0.028</td>
<td>0.009</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>0.01</td>
<td>0.316</td>
<td>0.083</td>
<td>0.028</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0.005</td>
<td>0.214</td>
<td>0.027</td>
<td>0.009</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\{p_B = p_B(0)(1-E), p_e(0) = 0.01, p_a = 0.5\} The maximum possible value of $E$ in this case is 0.5 ($R = 0$). The selection scheme is as in (2) with $s_1 = s_2 = s$. See text for further details.

Equilibrium values are not necessarily the maximum values attained. The appropriate value of $E$ which determines the maximum $D$ attained is also given in Table 3.

To determine the actual changes in $p_B$ and $D$ produced for particular initial frequencies at the $B$ locus, the appropriate values of $E$ and $B$, taken from Tables 2 and 3, are substituted in (14) below.

$$p_B = p_B(0)(1-E); \quad p_B = p_B(0) + E p_B(0)$$ \quad (14) (i)

$$D = B.D(0).$$ \quad (14) (ii)

TABLE 3

Values of $B$ when the $a$ gene is near its equilibrium value

<table>
<thead>
<tr>
<th>$x$</th>
<th>0.001</th>
<th>0.005</th>
<th>0.01</th>
<th>0.05</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>24.79</td>
<td>23.02</td>
<td>20.98</td>
<td>9.77</td>
<td>3.54</td>
<td>0.03</td>
<td>0.0</td>
</tr>
<tr>
<td>(15)</td>
<td>24.83(13)</td>
<td>23.52(11)</td>
<td>22.15(10)</td>
<td>14.84(8)</td>
<td>9.55(7)</td>
<td>2.3(4)</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>23.97</td>
<td>19.48</td>
<td>15.01</td>
<td>1.77</td>
<td>0.11</td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>(48)</td>
<td>24.16(37)</td>
<td>21.02(30)</td>
<td>18.08(27)</td>
<td>6.95(19)</td>
<td>2.75(14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>22.69</td>
<td>14.78</td>
<td>8.63</td>
<td>0.11</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>(103)</td>
<td>23.19(73)</td>
<td>17.86(58)</td>
<td>13.52(51)</td>
<td>2.65(30)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>20.32</td>
<td>8.51</td>
<td>2.86</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>(213)</td>
<td>21.54(137)</td>
<td>13.39(105)</td>
<td>8.21(89)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>8.42</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>(1094)</td>
<td>13.29(536)</td>
<td>2.57(317)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.005</td>
<td>2.80</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.0</td>
</tr>
<tr>
<td>(2195)</td>
<td>8.10(913)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The top figure, for each particular case, is the value of $B$ [as defined in (9)] when the $a$ gene is near its equilibrium value [$D = B.D(0)$]. The bottom figure is the value of $B$ which determines the maximum $D$ value attained. The relevant generation number is indicated in parentheses. Also indicated in parentheses is the number of generations taken for the $a$ gene to increase in frequency from 0.01 to 0.49, for each particular selection value $s$, and the half-life of each recombination fraction $R$. (The half-life is the number of generations, $n$, required to reduce the linkage disequilibrium to half its original value if there is no selection acting, so $(1-R)^n = 0.5$).
TABLE 4
Changes in $p_b$ and $D$ when initially $p_b = 0.05$ ($D(0) = 0.0095$) and $R = 0.001$ (half-life 690 generations)

<table>
<thead>
<tr>
<th>$s$</th>
<th>0.2 (48)</th>
<th>0.1 (103)</th>
<th>0.05 (213)</th>
<th>0.01 (1094)</th>
<th>0.005 (2195)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_b$</td>
<td>0.502</td>
<td>0.492</td>
<td>0.472</td>
<td>0.350</td>
<td>0.254</td>
</tr>
<tr>
<td>$D$</td>
<td>0.228</td>
<td>0.216</td>
<td>0.193</td>
<td>0.080</td>
<td>0.027</td>
</tr>
<tr>
<td>$D_{max}$</td>
<td>0.230(37)</td>
<td>0.220(73)</td>
<td>0.205(137)</td>
<td>0.126(536)</td>
<td>0.077(913)</td>
</tr>
</tbody>
</table>

The selection scheme is as in (2) with $s_1 = s_2 = s$. The value in parentheses under the selection parameters is the time taken for the selected $a$ gene to increase in frequency from 0.01 to 0.49. The maximum $D$ value attained is given and the appropriate generation number indicated in parentheses. The maximum final values which could be attained (when $R = 0$) are $p_b = 0.525$ and $D = 0.238$.

[Remember that the initial $a$ gene frequency in this case is $p_a(0) = 0.01$ and $D(0) = 0.01 p_a(0)$].

It is easier to gauge the produced effect by considering some particular numerical examples. In Table 4 are given the actual changes in $p_b$ and $D$ for particular values of the selection parameter, when the recombination fraction between the two loci is $R = 0.001$ and initially $p_b = 0.05$. The effect produced is very strong for selection up to about 5%. For weaker selection, the effect is not as strong, but still quite a significant change is produced and the effect lasts for a longer time.

In Table 5 the effects produced under selection of order 5% ($s = 0.05$) are given for various recombination values, namely $R = 0.001$, 0.005, 0.02 and 0.05, and for various initial frequencies of the neutral $b$ gene ($p_b(0) = 0.05, 0.2, 0.5, 0.8$ and 0.95).

In all the above calculations we have considered cases where the initial $a$ gene frequency is 0.01 and $D(0) = 0.01 p_b(0)$; which is the maximum positive $D$ value possible for these gene frequencies. Most likely, $D$ will not be at its maximum value by the time $p_a$ has reached 0.01, so the above calculations give the maximum possible effect which can be produced. Cases where the $a$ gene fre-

TABLE 5
Changes in $p_b$ and $D$ when $s = 0.05$ (213 generations) for particular values of the recombination fraction $R$

<table>
<thead>
<tr>
<th>Initial $p_b$</th>
<th>Final $p_b$</th>
<th>$R = 0.001$</th>
<th>$R = 0.005$</th>
<th>$R = 0.02$</th>
<th>$R = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>$p_b$</td>
<td>Max (137)</td>
<td>Max (105)</td>
<td>Max (70)</td>
<td>Max (5)</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>$0.005$</td>
<td>0.00095</td>
<td>0.472</td>
<td>0.193</td>
<td>0.205</td>
<td>0.157</td>
</tr>
<tr>
<td>$0.2$</td>
<td>0.0008</td>
<td>0.556</td>
<td>0.163</td>
<td>0.172</td>
<td>0.455</td>
</tr>
<tr>
<td>$0.5$</td>
<td>0.0005</td>
<td>0.722</td>
<td>0.102</td>
<td>0.108</td>
<td>0.659</td>
</tr>
<tr>
<td>$0.8$</td>
<td>0.0002</td>
<td>0.889</td>
<td>0.041</td>
<td>0.043</td>
<td>0.864</td>
</tr>
<tr>
<td>$0.95$</td>
<td>0.0005</td>
<td>0.972</td>
<td>0.010</td>
<td>0.011</td>
<td>0.966</td>
</tr>
</tbody>
</table>

The maximum $D$ attained is indicated and the appropriate generation number is given in parentheses. See text for details.
frequency is very small have been considered, and it was found that for small values of \( R \), \( D \) is maintained near its maximum value. As mentioned earlier, when \( n \) is small or near equilibrium, the deterministic model overestimates the amount of time spent in these ranges. There will, therefore, be fewer opportunities for recombination than the deterministic model implies. Hence, the actual changes in any particular case will be fairly close to those indicated above.

We have only considered cases where the selection scheme is symmetric, so that the equilibrium \( a \) gene frequency is 0.5. There is not much value in extensively studying various non-symmetric selection schemes, as the same general qualitative results obtain. It is obvious that the larger the \( a \) gene equilibrium frequency, the greater will be the hitchhiking effect.

From the above calculations there are two main questions that need to be considered. First, for a particular selection scheme, how long does an effect last and second, over what range of recombination values is a significant effect produced. The appropriate figures for various values of the selection parameter \( s \) are given in Table 6. An effect lasts for as long as it takes the \( a \) gene to reach its equilibrium value. This can be for quite a significant time, for example, up to 1000 generations for selection of order 1%. \( R_1 \), given in Table 6, is the range of recombination values up to which a particular selection coefficient produces a significant effect. \( R_1 \) is defined as the largest \( R \) value for which the maximum \( D \) is achieved at a time greater than twice the half-life of that particular \( R \). The \( R_1 \) values given are approximate and a slight underestimate of the true value. The recombination values scanned were \( R = 0.001, 0.003, 0.005, 0.008, 0.01, 0.02, 0.05, 0.1, 0.3 \) and 0.5 and \( R_1 \) in Table 6 is the appropriate value from this range.

The results of Tables 2–6 indicate that the hitchhiking effect is not important for values of \( R > s \). For values of \( R < s \) the hitchhiking effect can be very strong, and is particularly so when \( R < s/2 \). This means that as the selected \( a \) gene evolves towards its equilibrium value it is going to have a very significant effect.

### Table 6

<table>
<thead>
<tr>
<th>( T_1 )</th>
<th>( s )</th>
<th>( R_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.5</td>
<td>0.3 (2)</td>
</tr>
<tr>
<td>48</td>
<td>0.2</td>
<td>0.1 (7)</td>
</tr>
<tr>
<td>103</td>
<td>0.1</td>
<td>0.05 (13)</td>
</tr>
<tr>
<td>213</td>
<td>0.05</td>
<td>0.02 (34)</td>
</tr>
<tr>
<td>1094</td>
<td>0.01</td>
<td>0.008 (86)</td>
</tr>
<tr>
<td>2195</td>
<td>0.005</td>
<td>0.003 (231)</td>
</tr>
</tbody>
</table>

\( T_1 \) is the length of time (in generations) taken for the selected \( a \) gene to increase in frequency from 0.01 to 0.49 and hence indicates how long the hitchhiking effect lasts. The selection scheme is as in (2) with \( s_1 = s_2 = s \). \( R_1 \) is the range of recombination values up to which a significant effect is produced and is defined as follows. \( R_1 \) is the largest \( R \) value for which the maximum \( D \) is achieved at a time greater than twice the half-life of that particular \( R \). The appropriate half-lifes are indicated in parentheses after each \( R_1 \) value. (See Table 3 for definition of half-life).
on a number of closely linked loci. What is this number? For example, how many loci are there in a recombination region of 0.008 (the range over which a locus with selection of order 1% will produce a significant effect). The exact number of functional genes within a recombination fraction of this order is unknown. LEFEVRE (1971) has estimated that in Drosophila melanogaster a recombination fraction of 0.0001 correlates with a length of 3.7 – 3.8 $\times 10^3$ nucleotide pairs. The results of GELBART et al. (1974) on the rosy locus in Drosophila melanogaster support this estimate. GELBART et al. found that the standard map of the rosy locus extends over a recombination fraction of 0.0009 and that the rosy structural element codes for a polypeptide of approximately 1000 amino acids, requiring a coding sequence of some $3 \times 10^3$ nucleotide pairs. Using LEFEVRE's estimate and assuming that a sequence of 1000 nucleotide pairs is enough to specify a typical polypeptide chain, then it follows that in Drosophila melanogaster there is enough DNA in a recombination region of 0.008 to code for about 300 cistrons. It has been suggested that some of this DNA may have a controlling function. However, mutations affecting controlling genes must surely play an important role in evolution, and the hitchhiking effect would also apply to this class of genes. Calculations of BODMER (1972) show that in man there is enough DNA within a recombination fraction of 0.008 to code for about a thousand cistrons. So, even allowing for appreciable redundancy in the DNA it would seem that the hitchhiking effect may well be quite a powerful force in the determination of the evolution of the whole genome. The evolution of a selected locus may significantly influence a large number of closely linked loci.

(ii) Average effect of the selected a mutant on the neutral B locus. When the new a mutant occurs, with probability $p_b(0)$ it will be in a gamete carrying b, so that the new mutant is of type ab. In this case the frequency of the b gene will increase. With probability $p_a(0)$ the new mutant will be of type aB and the allele B will increase in frequency. To determine the overall effect of the selected a mutant on the neutral B locus, we must average over these two alternative events. In this respect we consider the average effect on the neutral locus in terms of gene frequency and heterozygosity.

With probability $p_b(0)$ the new mutant is of type ab in which case $D(0) = p_a(0)p_b(0)$. This is the case just considered. The final frequency of the neutral b gene is given by

$$p_b = p_b(0) + E p_a(0)$$

$(15)$

where $E$ is as given in $(11)$. (The summation for $E$ is continued until the a gene is near its equilibrium value. Some representative values are given in Table 2).

With probability $p_a(0)$ the new mutant is of type aB in which case $D(0) = -p_a(0)p_b(0)$, and it follows from $(10)$ that the final frequency of b is given by

$$p_b = p_a(0) \{1 - E\}$$

$(16)$

where $E$ is as for $(15)$. 
It follows from (15) and (16) that the average \( b \) gene frequency after the \( a \) gene has reached its equilibrium frequency is

\[
p_b(0) \left[ p_b(0) + E p_n(0) \right] + p_n(0) p_b(0) \left[ 1 - E \right] = p_b(0).
\]

That is, the average \( b \) gene frequency, after the evolution of the selected locus, is equal to its initial frequency. This result is to be expected from the fact that the \( B \) locus is neutral.

From (15) and (16) it is easy to show that the average heterozygosity at the \( B \) locus after the evolution of \( a \), denoted by \( H \), is as given in (17)

\[
H = 2 \left( 1 - E^2 \right) p_B(0) p_b(0).
\]

So the overall heterozygosity is always reduced to \((1 - E^2)\) times its original value (that is, it is reduced by a fraction \( E^2 \) of its original value). It is, in fact, possible to show that for the general case, with \( n \) alleles at the neutral \( B \) locus, the overall heterozygosity is always reduced to \((1 - E^2)H(0)\), where \( H(0) \) is the original heterozygosity and \( E \) is as above. Note that the percentage reduction in heterozygosity is independent of the initial allele frequencies at the \( B \) locus. It is dependent only on the selection operating at the \( A \) locus and the recombination fraction \( R \).

If \( R = 0 \), so that the loci are absolutely linked, then \( E = S_a \), where \( S_a = \sum_n \Delta p_a(n) \) and is as defined in (13), namely the total change in the \( a \) gene frequency. So if \( a \) is a new mutant, as in the case under consideration, and the equilibrium \( a \) gene frequency is 0.5, then on the average the heterozygosity at the neutral \( B \) locus is reduced by 25% of its original value. If the equilibrium \( a \) gene frequency is \( \hat{p}_a = 0.9 \), then the heterozygosity is reduced by 81%. Thus, for very tight linkage, the hitchhiking effect on the average reduces the heterozygosity at nearby neutral loci by a considerable amount. For looser linkage the effect dies off.

The values of \( E \) for various selection parameters and recombination values are given in Table 2, so the appropriate percentage reduction in heterozygosity, \( E^2 \% \), can be calculated in any case. For example, when \( s = 0.1, R = 0.001 \) the average reduction in heterozygosity is 21.6%. Note that if the equilibrium \( a \) gene frequency \( \hat{p}_a < 1/2 \), the reduction in heterozygosity will be less than that calculated for the case \( \hat{p}_a = 1/2 \), while if \( \hat{p}_a > 1/2 \) the reduction in heterozygosity will be much greater.

On the average, the overall heterozygosity will always be reduced. However, considering the average effect masks the fact that, in any particular case, the heterozygosity may be increased or decreased and the actual change may be very large. As an illustration consider the case where \( s = 0.05, R = 0.001, p_b(0) = 0.05 \) and hence the original heterozygosity \( H(0) = 0.095 \). If the new mutant is of type \( ab \) (which will be the case with probability 0.05), the heterozygosity after the evolution of \( a \) is greatly increased, namely to 0.498. When the new mutant is of type \( aB \) (probability .95), the final heterozygosity is reduced to 0.054. The com-
bination of these two results gives on the average a reduction in heterozygosity of 19.8%.

The decay of heterozygosity due to the hitchhiking effect must be compared with the rate at which heterozygosity decays due to random drift in a finite population. Kimura (1964) has shown that in the absence of selection and mutation the heterozygosity after \( n \) generations is given by

\[
H_n = H_0 \exp \left( -\frac{n}{2N_e} \right).
\]

where \( H_0 \) is the initial heterozygosity and \( N_e \) is the effective population number. The values of \( n \) which need to be considered are the lengths of time taken for the selected gene to increase in frequency to near its equilibrium value. So the appropriate values of \( n \) are 15, 48, 103, 213, 1094 and 2195 generations for selection values of \( s = 0.5, 0.2, 0.1, 0.05, 0.01 \) and 0.005 respectively.

For large selective values, the effect of hitchhiking outweighs that of random drift, even for fairly small population numbers. For example, consider the case where \( s = 0.05 \). The reduction in heterozygosity due to hitchhiking is 19.8% for \( R = 0.001 \) and 10.2% for \( R = 0.005 \). The reduction in heterozygosity due to random drift over 213 generations (the time it takes for the selected gene to increase in frequency to near its equilibrium value) is only 1% if \( N_e = 10^4 \). The reduction is less for larger population sizes. For small selective values the hitchhiking effect will still be the predominant force if the population size is sufficiently large, for example, if \( N_e > 10^2 \) when \( s = 0.005 \).

Thus, as a selected locus evolves towards equilibrium, it will have a significant effect on the level of heterozygosity at closely linked loci. However, it is not possible at this stage to determine the importance of the hitchhiking effect as a mechanism for reducing the total amount of heterozygosity in a population. For this we would need to know what fraction of the genome, at any time, consists of selected loci in the process of evolving towards equilibrium. Absolutely no estimate of this figure can be given at the present time.

From the above results one could possibly make theoretical predictions regarding the level of heterozygosity expected in populations experiencing strong selection pressures as compared to those undergoing weak selection. However, at the moment there would be little profit in this. It is doubtful if such a prediction could ever be of practical use. Not only would detailed knowledge of the kinds of selection pressures involved, the number of loci being selected etc. be needed, but also an extremely large sample of loci would need to be considered to detect significant differences in levels of heterozygosity.

(iii) Pseudo-selection acting on the neutral B locus. It has been pointed out by several authors that an apparent selective force can be created at intrinsically neutral loci by nonrandom association, that is, linkage disequilibrium, with selected loci. The apparent selection takes the form of heterozygote advantage, termed “associative overdominance” by Frydenberg (1963). Sved (1968) and Ohta and Kimura (1969, 1970, 1971) have considered the theoretical aspects of nonrandom association between neutral and overdominant loci due to random
 drift in finite populations. Ohta and Kimura have dealt extensively with the
case where the selected A locus is at equilibrium and the expected value of D,
denoted E(D), is zero, but due to the finite size of the population E(D^2) ≠ 0.
Thus, they are dealing solely with the associative overdominance caused by ran-
dom drift effects.

The approach to be taken here is basically quite different. We will be dealing
with the situation where the selected locus is evolving towards its equilibrium
value. As already shown, this process will create linkage disequilibrium be-
tween the neutral and selected locus. So we are concerned with situations where D ≠ 0.
This nonrandom association of the loci creates an apparent selective force at the
neutral locus, which may be termed pseudo-selection. We wish to study (deter-
ministically) the pseudo-selection created at the neutral locus as the selected
locus evolves towards equilibrium.

The simplest way to determine the pseudo-selection is to consider the selection
acting on the total zygotic array for the two locus system. This is illustrated
below in (19).

\[
\begin{align*}
AA & \quad (1-s_1) \ [x_1^2 + 2x_1x_2 + x_2^2] \\
Aa & \quad [2x_1x_3 + 2x_1x_4 + 2x_2x_3 + 2x_3x_4] \\
aa & \quad (1-s_2) \ [x_3^2 + 2x_3x_4 + x_4^2]
\end{align*}
\]  

(19)

All the quantities in (19) must be divided by the mean fitness \( \bar{w} \). It is easy to see
that the frequency of the genotype BB in the next generation, denoted by \( f_{BB}' \), is given by

\[ f_{BB}' = \frac{[(1-s_1) x_1^2 + 2x_1x_3 + (1-s_2) x_3^2]}{\bar{w}} \]

and it turns out, using (5), that this can be written in the form

\[ f_{BB}' = (p_{B}')^2 - Z \]  

where

\[ Z = D^2 \left\{ \frac{(s_1 + s_2)}{\bar{w}} + \frac{(s_1p_A - s_2p_A)^2}{\bar{w}^2} \right\} \]  

Similarly,

\[ f_{BB}' = 2p_{B}' p_{b}' + 2Z, \quad f_{bb}' = (p_{b}')^2 - Z . \]

Thus providing \( D ≠ 0 \), there is an excess of heterozygotes and deficiency of both
homozygotes resulting in an apparent overdominance at the neutral B locus.
When the A locus is at equilibrium the second term in (20) (ii) will vanish.
We define parameters $t_1(n)$ and $t_2(n)$ such that the pseudo-selection acting on the neutral $B$ locus can be written as in (21) below

\[
\begin{align*}
BB & \quad 1 - t_1(n) \\
Bb & \quad 1 \\
bb & \quad 1 - t_2(n)
\end{align*}
\]

$t_1$ and $t_2$ depend, of course, on the generation number $n$. For very small recombination values and small values of $p_b(0)$ the pseudo-selection acting on the neutral locus will be approximately equal to the selection acting on the $A$ locus. For looser linkage the pseudo-selection of course decreases. As illustration, consider the case where the selection at the $A$ locus is of order 5% and the initial $b$ gene frequency $p_b(0) = 0.05$. Some representative values of the pseudo-selection parameters, when $R = 0.001$, are as follows. At generation 40 $t_1 = 0.03, t_2 = 0.007$, at generation 120 $t_1 = 0.04, t_2 = 0.04$ and at generation 213 (when the $a$ gene frequency is near its equilibrium value) $t_1 = 0.03, t_2 = 0.03$. If $R = 0.01$ then at say, generation 120 $t_1 = 0.01, t_2 = 0.01$. The pseudo-selection is dependent on the initial $b$ gene frequency, being less for larger initial values $p_b(0)$. For instance, in the above example, if $p_b(0) = 0.8$ and $R = 0.001$, then at generation 80 $t_1 = 0.01, t_2 = -0.01$ compared with $t_1 = 0.04, t_2 = 0.03$ for $p_b(0) = 0.05$. The results of this section clearly illustrate the fact that it is generally very difficult, if not impossible, to distinguish the selective effects of a given detectable gene locus from those that may be due to closely linked loci.

**Case B: The neutral $b$ gene is the new mutant**

In this section we consider the case where the selected $A$ locus is in the process of evolving towards its equilibrium value when a neutral $b$ mutant occurs. A hitchhiking effect is again observed. However, in this case the effect is not nearly as strong as for the previous situation where the new mutant is the selected gene. Moreover, as pointed out by HAIGH and MAYNARD SMITH (1976), it seems that in natural populations the case where the new mutant is selected is probably the more frequent event. The reason for this is that in large populations most of the neutral polymorphism will be contributed by alleles which remain in the population for very long periods (the number of generations until fixation of a neutral mutant has a mean of approximately $4N_e$ (KIMURA and OHTA 1969) where $N_e$ is the effective population size), whereas selected alleles will reach equilibrium rather more quickly.

The equations governing the changes in the allele frequencies at the $B$ locus and the linkage disequilibrium when the neutral $b$ gene is the new mutant are exactly as before, being given by (5) [see also (10)] and (7). If the new mutant is of type $ab$, then $D(0) = p_a(0)p_b(0)$ and the $b$ gene will increase in frequency. It is easy to see from (10) that in this case

\[
p_b(n+1) = p_b(0) \left\{ 1 + \frac{F_n}{p_b(0)} \right\}
\]  

(22) (i)
and hence,
\[ p_B(n+1) = p_B(0) - \frac{p_b(0)F_n}{p_a(0)} \]  
(22) (ii)

where
\[ F_n = \sum_{k=0}^{n} \frac{\Delta p_a(k)}{p_A(k)p_a(k)} B_k p_A(0) p_a(0) . \]  
(22) (iii)

Note that \( F_n \) is independent of the frequencies at the \( B \) locus.

As before, one can get approximate results for changes in the frequency of the neutral gene \( b \) and in the linkage disequilibrium \( D \) if the recombination fraction is very small. It follows from (7) that, for \( R \) very small (taking \( R = 0 \)), \( D \) will increase to the value given below in (23), attained when the \( a \) gene is near its equilibrium value.

\[ D_{\text{max}} \approx \hat{p}_a \hat{p}_a \frac{p_b(0)}{p_a(0)} \quad \text{(for } R \text{ small)} \]  
(23)

For the case where the selection is symmetric [\( s_1 = s_2 = s \) in (2)] then the equilibrium \( a \) gene frequency \( \hat{p}_a = 0.5 \), and \( D \) will increase to \( \approx 0.25 \frac{p_b(0)}{p_a(0)} \). If \( p_a(0) \) is small, for example, if \( p_a(0) = 0.02 \) (\( p_b(0) = 0.01 \)) then \( D_{\text{max}} \) will be large, namely 0.125. If \( p_a(0) \) is near the equilibrium \( a \) gene frequency, namely 0.5, say \( p_a(0) = 0.4 \) then \( D_{\text{max}} = 0.006 \).

It follows from (22) that for \( R \) small
\[ p_b(n+1) = p_b(0) \left[ 1 + \frac{S_a}{p_a(0)} \right] \]  
(24)

where \( S_a = \sum_n \Delta p_a(n) \) [\( = \hat{p}_a - p_a(0) \)] is the total change in \( a \) frequency since the occurrence of the new \( b \) mutant. So if \( p_a(0) = 0.05 \) (\( \hat{p}_a = 0.5 \)) then the final \( b \) gene frequency will be \( p_b(n) \approx 10 p_b(0) \) while if \( p_a(0) = 0.2 \), \( p_b(n) \approx 2.5 p_b(0) \).

The approximate changes in \( p_b \) and \( D \) for very small recombination values are given in Table 7. Cases where \( p_b(0) = 0.01 \) are considered for various initial frequencies of the selected \( a \) gene.

Details of actual changes in \( p_b \) and \( D \) for various values of the recombinant fraction \( R \) and initial \( a \) gene frequency \( p_a(0) \) will not be given. The main qualitative result in all cases is that the hitchhiking effect will only be important when the initial \( a \) gene frequency is small. The magnitude of the effect produced is very dependent on the ratio \( p_b(0)/p_a(0) \) (see (23) and (24), also Table 7).

We again consider the average effect on the neutral \( B \) locus in terms of gene frequency and heterozygosity. It is possible to show that the average \( b \) gene frequency after the evolution of the \( a \) gene is \( p_b(0) \), as expected. The average heterozygosity after hitchhiking is given by
\[ H = 2 p_b(0)p_b(0) \left[ 1 - \frac{F^2 p_b(0)}{p_A(0)p_a(0)p_b(0)} \right] \]  
(25)
Approximate changes in frequency of the new mutant neutral gene \( b \) \((p_b)\) and linkage disequilibrium \( D \) for small recombination values and various initial \( a \) gene frequencies \( p_a(0) \).

<table>
<thead>
<tr>
<th>( p_a )</th>
<th>Initial ( D )</th>
<th>( p_b )</th>
<th>Final ( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.0099</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0095</td>
<td>0.1</td>
<td>0.05</td>
</tr>
<tr>
<td>0.1</td>
<td>0.009</td>
<td>0.05</td>
<td>0.025</td>
</tr>
<tr>
<td>0.3</td>
<td>0.007</td>
<td>0.017</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Initially \( p_b(0) = 0.01 \) and \( D(0) = p_a(0) p_b(0) \).

So the overall heterozygosity at the neutral locus is again reduced. However it should be remembered that in this case the original heterozygosity is very small, since \( b \) is a new mutant. The percentage reduction in heterozygosity is very small unless \( p_b(0) \) and \( p_b(0) \) are of the same order of magnitude. However, it should be pointed out that these results apply to the average heterozygosity. When the initial \( a \) gene frequency is small, what in fact happens is that with low probability the heterozygosity will increase significantly while with large probability it will decrease slightly. So in any particular case the effect produced could be quite significant. This is masked by considering the average effect.

There will of course be a pseudo-selection acting on the neutral \( B \) locus while the selected locus evolves towards its equilibrium value. Details will not be given here.

**Case C: The new a mutant is at a selective advantage.**

Maynard Smith and Haigh (1974) have considered the hitchhiking effect when a selectively favorable mutation occurs in a population and is subsequently fixed. They have dealt with a number of questions. The two of relevance to the present approach are the changes in gene frequency at a closely linked neutral locus and the reduction in heterozygosity caused by the evolution of the selected locus. The approach used by Maynard Smith and Haigh can be slightly extended to take account also of effects on the linkage disequilibrium between the selected and neutral loci and of the average effect on heterozygosity caused by hitchhiking.

The selection scheme is as given below in (26)

\[
\begin{array}{ccc}
AA & Aa & aa \\
1 & 1 + hs & 1 + s
\end{array}
\]  

(26)

with \( 0 \leq h \leq 1 \). As before, the \( B \) locus is assumed to be selectively neutral. We suppose that a new mutation \( a \) has just recently occurred. The frequency of the \( a \) gene will be increasing each generation until it becomes fixed in the population. If the neutral and selected loci are in linkage disequilibrium then, as before, the
gene frequencies at the neutral locus will also change each generation during the evolution of the selected locus. It turns out that

\[ p'_s = p_b - \frac{D}{w} [h s(p_a - p_a) + sp_a] \]  

The amount of linkage disequilibrium between the A and B loci is also changing each generation.

\[ D' = \left\{ \frac{p'_A p'_B}{p_A p_a} - R \frac{(1 + h s)}{w} \right\} D \]  

The term \( \frac{p'_A p'_B}{p_A p_a} \) > 1 until \( p_A = p_a = \frac{1}{2} \).

The actual changes in gene frequencies at the neutral locus and the linkage disequilibrium generated can be calculated for any particular values of the selection parameters and recombination fraction. (Methods exactly the same as illustrated for the case of a heterotic selected locus can be used). Details of particular cases will not be given here. However, it should be mentioned, as pointed out by Maynard Smith and Haigh (1974) that the effects produced are qualitatively different for various values of \( h \). (Maynard Smith and Haigh have given approximations for changes in the B gene frequency applicable when \( R \) is very small.)

For the recessive case, \( h = 0 \), the hitchhiking effect is only significant for extremely tight linkage. This follows from the fact that for a recessive gene, the change in frequency of the selected \( a \) gene is initially very slow. By the time the \( a \) gene has reached frequencies such that selection on the homozygotes \( aa \) can be a significant force, the disequilibrium between the A and B loci will have been broken down by recombination, unless the recombination fraction between the two loci is extremely small. Thus the hitchhiking effect will not be very significant when the new selected mutant is recessive.

For the additive and dominant cases, \( h = \frac{1}{2} \) and 1 respectively, the hitchhiking effect can be very significant, both in terms of changes in gene frequency at the neutral locus and build-up of linkage disequilibrium. The value of the linkage disequilibrium by the time the \( a \) gene is near fixation is usually fairly low. This is because the first term in (28) is \(< 1 \) when \( p_a > \frac{1}{2} \) and so the linkage disequilibrium decreases. However, during the evolution of the selected locus, \( D \) can attain very significant values. As illustration consider the case where \( s = 0.05, h = \frac{1}{2}, R = 0.005, p_b(0) = 0.05 \) and hence \( D(0) = 0.0095 \). The final \( b \) gene frequency can be shown to be \( p_b = 0.441 \) and \( D = 0.001 \) but the maximum \( D \) attained is \( D_{max} = 0.098 \) at generation 170.

If the A and B loci are absolutely linked, that is \( R = 0 \), then fixation at the A locus will also result in fixation at the B locus. (This applies to the case where the new mutant is at the selected locus). For very tight linkage, the B locus will
either be fixed or come close to fixation. Maynard Smith and Haigh (1974) considered this aspect of the hitchhiking effect and suggest that it may be an important force in reducing the level of polymorphism in natural populations. Ohta and Kimura (1975) have also considered this problem and conclude that the hitchhiking effect is generally unimportant as a mechanism for reducing heterozygosity. Ohta and Kimura criticize Maynard Smith and Haigh for considering only the one-sided effect of hitchhiking. However, it turns out that Maynard Smith and Haigh (1974) in fact did calculate the average reduction in heterozygosity. They considered the one-sided effect of hitchhiking when \( p_b(0) = 0.5 \) and in this case, by symmetry, the heterozygosity is reduced by the average amount. Most of the examples considered by Ohta and Kimura (1975) deal with the case where the new mutant is at the neutral locus and the selected \( a \) gene is in the process of evolving towards fixation. They find that the hitchhiking effect is not very significant in this case. This is to be expected. It has been pointed out above, for the case where the selection is heterotic, that the hitchhiking effect is a much more significant force when the new mutant is at the selected locus than when it is at the neutral locus. This same result applies for the case when the \( a \) gene is at a selective advantage. I would agree with Haigh and Maynard Smith (1976) that it will be much more common for a selectively favorable mutant to arise closely linked to a pre-existing neutral polymorphism than the other way around. In this case the reduction in heterozygosity at the neutral loci can be very large. For example, when \( s = 0.1 \), \( h = \frac{1}{2} \), the heterozygosity at neutral loci a distance \( R = 0.001 \) from the selected locus is reduced on the average by 81.2\%. As pointed out earlier, it is not possible at this stage to determine the importance of the hitchhiking effect as a mechanism for reducing the total level of heterozygosity in a population. For this we would need to know what fraction of the genome at any time consists of selected loci that are in the process of evolving towards equilibrium or fixation.

**THREE-LOCUS MODEL**

*Case D: The ordering of loci is ABC where B is the selected locus (heterosis).*

A model of greater interest than those considered above is a three-locus model with one locus selected and two neutral loci. We assume there are two alleles at each locus. At the first locus the alleles are \( A \) and \( a \), at the second \( B \) and \( b \) and at the third \( C \) and \( c \). In this case there are eight haplotypes (gametic or chromosome types) possible, namely \( ABC, ABc, AaB, AbC, aBC, aBc, aBC \) and \( abc \). Their frequencies are denoted by \( x_1, x_2, x_3, x_4, x_5, x_6, x_7 \) and \( x_8 \) respectively. We use a simple extension of the two-locus, two-allele representation of haplotype frequencies as given in (1). For the three-locus model, the eight haplotype frequencies can be completely specified by three gene frequencies and four disequilibrium parameters. Three of the disequilibrium parameters measure pairwise disequilibrium, while the fourth is a measure of triple association after taking account of the pairwise associations. The representation of the haplotype frequencies is given in Table 8.
At the first locus the alleles are $A$ and $a$, at the second $B$ and $b$ and at the third $C$ and $c$. $p_A$, $p_B$ etc. denote the frequencies of the alleles $A$ and $B$, respectively, etc. $D_{AB}$ denotes the pairwise linkage disequilibrium between the loci $A$ and $B$ and similarly for $D_{AC}$ and $D_{BC}$. $D_{ABC}$ is the third order linkage disequilibrium after taking account of the pairwise associations. The eight haplotype frequencies can be written as above.

We first give consideration to the case where the $B$ locus is selected and the $A$ and $C$ loci are neutral. The selection scheme (heterosis) is given in (29)

\[
\begin{array}{ccc}
 BB & Bb & bb \\
 1-s_1 & 1 & 1-s_2 \\
\end{array}
\]

with $s_1, s_2 \geq 0$. The recursion system relating the gametic frequencies in the next generation to those in the present is given in Table 9.

The recombination fraction between the $A$ and $B$ loci is denoted $R_{AB}$, that between the $B$ and $C$ loci $R_{BC}$ and that between the $A$ and $C$ loci $R_{AC}$. We assume that there is no interference, so $R_{AC} = R_{AB} + R_{BC} - 2R_{AB}R_{BC}$.

Suppose that a new mutation $b$ has just recently occurred. The frequency of the selected $b$ gene will be increasing each generation until equilibrium is reached. As before, we find that the gene frequencies at the neutral $A$ and $C$ loci are also changing each generation and also the amount of linkage disequilibrium between the neutral and selected loci. The changes each generation are given as follows:

\[
p_A' = p_A - \frac{(s_1 p_B - s_2 p_b)}{\bar{w}} D_{AB} \quad (30) (i)
\]

\[
p_C' = p_C - \frac{(s_1 p_B - s_2 p_b)}{\bar{w}} D_{BC} \quad (30) (ii)
\]

\[
D_{AB}' = \left[ \frac{p_A' p_B'}{p_B' p_B} - \frac{R_{AB}}{\bar{w}} \right] D_{AB} \quad (30) (iii)
\]

\[
D_{BC}' = \left[ \frac{p_A' p_C'}{p_B' p_B} - \frac{R_{BC}}{\bar{w}} \right] D_{BC} \quad (30) (iv)
\]

Note that these changes are exactly as for the two locus case [see (5) and (7)].

However, the most important finding from the present model is that $D_{ABC}$, the...
Recursion system for the three-locus model when the B locus is selected [see (29)] and the A and C loci are neutral

\[ \bar{w}_{x} = x_{1}(1-s_{p_{B}}) - R_{1}(1-s_{q_{B}}) + 1/2(\alpha_{3} + \beta_{3}) \]

\[ \bar{w}_{x}' = x_{2}(1-s_{p_{B}}) + R_{1}(1-s_{q_{B}}) + 1/2(\alpha_{3} + \beta_{3}) \]

\[ \bar{w}_{x}'' = x_{3}(1-s_{p_{B}}) - R_{1}(1-s_{q_{B}}) - 1/2(\alpha_{3} + \beta_{3}) \]

\[ \bar{w}_{x}''' = x_{4}(1-s_{p_{B}}) + R_{1}(1-s_{q_{B}}) + 1/2(\alpha_{3} + \beta_{3}) \]

The first term is the analogue of the factor involved when we consider the disequilibrium between a neutral and selected locus [see (30)(iii) and (iv)]. The other terms involved are somewhat complicated. However, it turns out that even if \( D_{AC} = 0 \) initially, it can increase in frequency. So in a sense the effect on the linkage disequilibrium between the two neutral loci can be greater than between a neutral and selected locus, as in the latter case, once \( D = 0 \) it remains thus. The reason for this large buildup of disequilibrium between the two neutral loci is that the changes in gene frequency at the neutral loci are more correlated than the changes in gene frequency at the selected locus and one of the neutral loci.

We consider the case where the new mutant is of type \( abc \). The frequency of the gametic types \( AbC, Abc, abC \) are initially zero. The occurrence of the new linkage disequilibrium between the two neutral loci, is also greatly affected by the evolution of the selected locus.

\[ D_{AC}' = \left[ \frac{p_{B}'p_{p} - \frac{R_{3}}{\bar{w}}} {p_{B}p_{p}} \right] D_{AC} + \frac{R_{3}} {\bar{w}} (s_{1}\beta_{1} + s_{2}\beta_{2}) + \frac{\Delta p_{B}} {p_{B}p_{p}} (\beta_{3} - \beta_{1}) \]

\[ + \left( \frac{\Delta p_{B}} {p_{B}p_{p}} \right)^{2} [\beta_{1}p_{B} + \beta_{3}p_{B}] \]

where \( \Delta p_{B} = p_{B}' - p_{B}, \beta_{1}, \beta_{2} and \beta_{3} are defined in Table 9. \( (D_{AC}) = \beta_{1} + \beta_{2} + \beta_{3} \). The first term is the analogue of the factor involved when we consider the disequilibrium between a neutral and selected locus [see (30)(iii) and (iv)]. The other terms involved are somewhat complicated. However, it turns out that even if \( D_{AC} = 0 \) initially, it can increase in frequency. So in a sense the effect on the linkage disequilibrium between two neutral loci can be greater than between a neutral and selected locus, as in the latter case, once \( D = 0 \) it remains thus. The reason for this large buildup of disequilibrium between the two neutral loci is that the changes in gene frequency at the neutral loci are more correlated than the changes in gene frequency at the selected locus and one of the neutral loci.

We consider the case where the new mutant is of type \( abc \). The frequency of the gametic types \( AbC, Abc, abC \) are initially zero. The occurrence of the new
The hitchhiking effect creates linkage disequilibrium between the neutral and selected loci as follows

\[
D_{AB}(0) = p_b(0)p_A(0) \quad (32) \text{(i)}
\]
\[
D_{BC}(0) = p_b(0)p_C(0) \quad (32) \text{(ii)}
\]

These values will of course be quite small as the \( b \) gene frequency \( p_b(0) \) is small. However, it is the existence of this initial linkage disequilibrium which is responsible for the hitchhiking effect. As the \( b \) gene increases in frequency towards its equilibrium value, the \( a \) and \( c \) genes will also increase in frequency being pulled along with the \( b \) gene. The values of the linkage disequilibria \( D_{AB} \) and \( D_{BC} \) will also change each generation. These results will be exactly as for the two locus case already discussed.

The main point of interest in this section is the disequilibrium generated between the two neutral loci. The main qualitative results are listed below, with some numerical examples.

1. The disequilibrium between the neutral (polymorphic) loci \( A \) and \( C \) can initially be zero but still increase to quite significant values. This is illustrated in the example of Table 10. The selection on the \( B \) locus is of the order 5% \([s_1 = s_2 = 0.05 \text{ in (29)}]\) and the \( B \) locus is halfway between the \( A \) and \( C \) loci with \( R_1 = R_2 = 0.005 \). As the \( b \) gene increases in frequency from 0.01 to 0.49 (near its equilibrium value of 0.5), \( D_{AC} \) builds up in value then decreases. At generation 80, for example, \( D_{AC} = 0.052 \) \( (p_a = 0.22, p_c = 0.42) \), while at generation 160 \( D_{AC} = 0.032 \) \( (p_a = 0.34, p_c = 0.52) \).

2. The position of the \( B \) locus seems to have very little effect on the magnitude of \( D_{AC} \). The position does of course alter the changes in \( p_a, p_c, D_{AB} \) and \( D_{BC} \). For example, if the starting conditions are exactly as for Table 10 except that the

**TABLE 10**

*The effect of a selected locus (B) on closely linked neutral loci, with \( s_1 = s_2 = 0.05 \) and \( R_1 = R_2 = 0.005 \)*

<table>
<thead>
<tr>
<th>Gen.</th>
<th>( p_b )</th>
<th>( p_a )</th>
<th>( p_c )</th>
<th>( D_{AB} )</th>
<th>( D_{AC} )</th>
<th>( D_{BC} )</th>
<th>( D_{ABC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.01</td>
<td>0.05</td>
<td>0.30</td>
<td>0.0095</td>
<td>0.000</td>
<td>0.007</td>
<td>-0.0067</td>
</tr>
<tr>
<td>40</td>
<td>0.07</td>
<td>0.10</td>
<td>0.33</td>
<td>0.048</td>
<td>0.023</td>
<td>0.035</td>
<td>-0.024</td>
</tr>
<tr>
<td>80</td>
<td>0.23</td>
<td>0.22</td>
<td>0.42</td>
<td>0.115</td>
<td>0.052</td>
<td>0.085</td>
<td>-0.029</td>
</tr>
<tr>
<td>120</td>
<td>0.39</td>
<td>0.31</td>
<td>0.49</td>
<td>0.124</td>
<td>0.047</td>
<td>0.092</td>
<td>-0.011</td>
</tr>
<tr>
<td>160</td>
<td>0.46</td>
<td>0.34</td>
<td>0.52</td>
<td>0.106</td>
<td>0.032</td>
<td>0.078</td>
<td>-0.003</td>
</tr>
<tr>
<td>213</td>
<td>0.49</td>
<td>0.35</td>
<td>0.52</td>
<td>0.081</td>
<td>0.019</td>
<td>0.060</td>
<td>-0.001</td>
</tr>
</tbody>
</table>

There are three loci \( A, B \) and \( C \). The \( B \) locus is selected. The selection scheme is as in (29) with \( s_1 = s_2 = 0.05 \). Locus \( B \) is assumed to be halfway between \( A \) and \( C \) with \( R_1 = R_2 = 0.005 \). The notation for gene frequencies and linkage disequilibrium is as described in Table 8. A new mutation \( b \) has just recently occurred and we assume the new mutant is of type \( abc \). We consider the evolution of the three loci as the \( b \) gene increases in frequency from 0.01 to 0.49 (near its equilibrium value). The neutral loci \( A \) and \( C \) are assumed to be already polymorphic as indicated by the initial frequencies.
B locus is no longer halfway between A and C, instead \( R_1 = 0.001, R_2 = 0.009 \), then the values of \( D_{AC} \) generated are nevertheless the same as those in Table 10.

(3) The value of \( D_{AC} \) can initially be negative and then become positive. For example, if the starting conditions are exactly as for Table 10 except that initially \( D_{AC} = -0.005 \), the values of \( D_{AC} \) generated are very similar to those in Table 10, and the hitchhiking effect is the predominant force.

The main results of the two-locus model also obtain. These are reiterated.

(4) The effect of the selected locus on closely linked neutral loci lasts for as long as it takes the selected locus to reach its equilibrium value. For smaller selective values the effect will last a longer time but will not be as strong, as there is also more time for recombination to reduce the effect of the pulling along of the gene frequencies together. In Table 11 there is 10% selection acting \((s_1 = s_2 = 0.1)\) on the B locus. All the initial conditions are the same as in Table 10 to facilitate a direct comparison. The disequilibrium \( D_{AC} \) generated between the neutral loci is much greater for the case of stronger selection.

Thus, for 10% selection \( D_{AC} = 0.086 \) at generation 60 (Table 11) whereas for 5% selection the value of \( D_{AC} \) does not get as large as this (Table 10). However, the effect of the selected locus on the whole lasts for a longer time with the lower value of selection. Even with very weak selection, significant values of \( D \) will be generated if the loci are tightly linked. This is illustrated in Table 12 where the selection is of the order 1% and \( R_1 = R_2 = 0.0001 \). At generation 800 the linkage disequilibrium between the two neutral loci is 0.14. For tighter linkage the disequilibrium generated would be even greater.

(5) In all cases, \( D_{AC} \) eventually goes to zero, but for cases where a significant effect is produced this is usually well after the gene frequencies have reached their equilibrium values (see Tables 10, 11 and 12).

(6) It should be noted that the hitchhiking effect can also cause \( D_{ABC} \), the third order interaction after taking account of pairwise interactions, to increase in value. Since the theory for estimating and testing third order interaction is just being developed (HILL 1976; THOMSON and BODMER in preparation) this aspect of the hitchhiking effect will not be discussed here.

All the above examples have been concerned with the case where the new

<table>
<thead>
<tr>
<th>Gen.</th>
<th>( p_b )</th>
<th>( p_a )</th>
<th>( p_c )</th>
<th>( D_{AB} )</th>
<th>( D_{AC} )</th>
<th>( D_{BC} )</th>
<th>( D_{ABC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.01</td>
<td>0.05</td>
<td>0.30</td>
<td>0.0095</td>
<td>0.00</td>
<td>0.007</td>
<td>-0.0067</td>
</tr>
<tr>
<td>20</td>
<td>0.07</td>
<td>0.10</td>
<td>0.34</td>
<td>0.055</td>
<td>0.030</td>
<td>0.041</td>
<td>-0.031</td>
</tr>
<tr>
<td>40</td>
<td>0.25</td>
<td>0.25</td>
<td>0.45</td>
<td>0.145</td>
<td>0.080</td>
<td>0.107</td>
<td>-0.043</td>
</tr>
<tr>
<td>60</td>
<td>0.40</td>
<td>0.37</td>
<td>0.53</td>
<td>0.169</td>
<td>0.086</td>
<td>0.125</td>
<td>-0.041</td>
</tr>
<tr>
<td>80</td>
<td>0.47</td>
<td>0.41</td>
<td>0.57</td>
<td>0.158</td>
<td>0.072</td>
<td>0.116</td>
<td>-0.006</td>
</tr>
<tr>
<td>103</td>
<td>0.49</td>
<td>0.42</td>
<td>0.58</td>
<td>0.140</td>
<td>0.057</td>
<td>0.103</td>
<td>-0.002</td>
</tr>
</tbody>
</table>

Exactly as for Table 10 except that \( s_1 = s_2 = 0.1 \) \((R_1 = R_2 = 0.005)\).
TABLE 12

The effect of a selected locus (B) on closely linked neutral loci, with $s_1 = s_2 = 0.01$ and $R_1 = R_2 = 0.0001$

<table>
<thead>
<tr>
<th>Gen.</th>
<th>$p_B$</th>
<th>$p_A$</th>
<th>$p_C$</th>
<th>$D_{AB}$</th>
<th>$D_{AC}$</th>
<th>$D_{BC}$</th>
<th>$D_{ABC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.01</td>
<td>0.05</td>
<td>0.30</td>
<td>0.0095</td>
<td>0.000</td>
<td>0.007</td>
<td>-0.0067</td>
</tr>
<tr>
<td>200</td>
<td>0.06</td>
<td>0.10</td>
<td>0.34</td>
<td>0.055</td>
<td>0.032</td>
<td>0.041</td>
<td>-0.034</td>
</tr>
<tr>
<td>400</td>
<td>0.23</td>
<td>0.25</td>
<td>0.45</td>
<td>0.161</td>
<td>0.104</td>
<td>0.118</td>
<td>-0.061</td>
</tr>
<tr>
<td>600</td>
<td>0.38</td>
<td>0.40</td>
<td>0.55</td>
<td>0.214</td>
<td>0.139</td>
<td>0.157</td>
<td>-0.035</td>
</tr>
<tr>
<td>800</td>
<td>0.46</td>
<td>0.46</td>
<td>0.60</td>
<td>0.220</td>
<td>0.140</td>
<td>0.162</td>
<td>-0.014</td>
</tr>
<tr>
<td>1094</td>
<td>0.49</td>
<td>0.49</td>
<td>0.63</td>
<td>0.215</td>
<td>0.133</td>
<td>0.158</td>
<td>-0.004</td>
</tr>
</tbody>
</table>

As for Table 10 except that $s_1 = s_2 = 0.01$ and $R_1 = R_2 = 0.0001$.

mutant is of type abc and $p_a < p_A, p_c < p_C$. This gives the maximum effect. Other combinations of starting conditions have been considered; some give significant effects and others do not. Thus the hitchhiking effect will not always produce significant changes in gene frequency or build-up of linkage disequilibrium. Very roughly we can say that if the new selected mutant occurs in a chromosome carrying a gene at the neutral locus which is in fairly low frequency, then a very significant effect will be produced. This means that in a proportion of cases (not particularly small) the hitchhiking effect is a predominant force. It should also be pointed out that, depending on initial conditions, the amount of disequilibrium generated between more distant loci can be greater than that between tightly linked loci. However, other things being equal, the hitchhiking effect decreases monotonically with increasing recombination fraction.

Case E: The ordering of loci is ABC where C is the selected locus.

To conclude the discussion on the three-locus model, consideration will be given to the case where the C locus is selected and the A and B loci are neutral. The selection scheme is as before, namely heterosis, and is given in (33).

\[

cC \quad Cc \quad cc
\]

\[
1-s_1 \quad 1 \quad 1-s_2
\]  \hspace{1cm} (33)

with $s_1, s_2 \geq 0$. As before, the recombination fraction between the A and B loci is $R_1$, between the B and C loci $R_2$, and that between the A and C loci $R_3$ ($R_3 = R_1 + R_2 - 2R_1R_2$). We suppose that a new mutation c has just recently occurred and we want to determine the effect produced on the neutral A and B loci as the c gene evolves towards equilibrium.

The changes in gene frequency at the neutral loci and the pairwise disequilibria $D_{AC}$ and $D_{BC}$ are exactly as expected from two-locus theory. As in the previous case, the linkage disequilibrium between the two neutral loci is also greatly affected by the evolution of the selected locus.

\[
D_{AB}' = \left[ \frac{p_d p' c}{p_c p_c} - \frac{R_1}{w} \right] D_{AB} + \frac{R_1}{w} \left[ s_1 a_1 + s_2 a_3 \right] \\
+ \frac{\Delta p_c}{p_c} \left( a_3 - a_1 \right) + \left( \frac{\Delta p_c}{p_c} \right)^2 \left( a_1 p_c + a_3 p_c \right) \hspace{1cm} (34)
\]
The effect of a selected locus (C) on closely linked neutral loci, with $s_1 = s_2 = 0.05$ and $R_1 = 0.002$ and $R_2 = 0.001$

<table>
<thead>
<tr>
<th>Gen.</th>
<th>$p_c$</th>
<th>$p_a$</th>
<th>$p_b$</th>
<th>$D_{AB}$</th>
<th>$D_{AC}$</th>
<th>$D_{BC}$</th>
<th>$D_{ABC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.01</td>
<td>0.05</td>
<td>0.30</td>
<td>0.000</td>
<td>0.0095</td>
<td>0.007</td>
<td>-0.0067</td>
</tr>
<tr>
<td>40</td>
<td>0.07</td>
<td>0.10</td>
<td>0.34</td>
<td>0.032</td>
<td>0.052</td>
<td>0.041</td>
<td>-0.032</td>
</tr>
<tr>
<td>80</td>
<td>0.23</td>
<td>0.23</td>
<td>0.45</td>
<td>0.095</td>
<td>0.135</td>
<td>0.117</td>
<td>-0.054</td>
</tr>
<tr>
<td>120</td>
<td>0.39</td>
<td>0.35</td>
<td>0.55</td>
<td>0.120</td>
<td>0.159</td>
<td>0.149</td>
<td>-0.031</td>
</tr>
<tr>
<td>160</td>
<td>0.46</td>
<td>0.39</td>
<td>0.59</td>
<td>0.117</td>
<td>0.146</td>
<td>0.149</td>
<td>-0.017</td>
</tr>
<tr>
<td>213</td>
<td>0.49</td>
<td>0.41</td>
<td>0.61</td>
<td>0.107</td>
<td>0.125</td>
<td>0.142</td>
<td>-0.010</td>
</tr>
</tbody>
</table>

There are three loci $A$, $B$ and $C$. The $C$ locus is selected. The selection scheme is as in (33) with $s_1 = s_2 = 0.05$. The recombination fraction between the $A$ and $B$ loci $R_1 = 0.002$ and between the $B$ and $C$ loci $R_2 = 0.001$. The notation for gene frequencies and linkage disequilibrium is as described in Table 8. A new mutation $c$ has just recently occurred and we assume the new mutant is of type $abc$. We consider the evolution of the three loci as the $c$ gene increases in frequency from 0.01 to 0.49 (near its equilibrium value). The neutral loci $A$ and $B$ are assumed to be already polymorphic as indicated by the initial frequencies.

where $\Delta p_c = p_c' - p_c$, $D_{AB} = \alpha_1 + \alpha_2 + \alpha_3$ with the $\alpha$'s as defined in Table 9.

The main point of interest in this case is that significant disequilibrium can be generated between the neutral $A$ and $B$ loci even when the selected locus is not between the two neutral loci. In fact, $R_2$ can be quite large compared with $R_1$ and a significant effect still produced. This is illustrated by considering $D_{AB}$ in Tables 13 and 14. In Table 13, $R_1 = 0.002$, $R_2 = 0.001$ while in Table 14 $R_1 = 0.002$, $R_2 = 0.01$. Even though $R_2$ is significantly larger in the second example (Table 14 $R_2 = 0.01$) $D_{AB}$, the disequilibrium generated between the two neutral loci $A$ and $B$, is still significant and nearly as large as that generated in the first example (Table 13 $R_2 = 0.001$). As $R_2$ gets larger, the effect becomes insignificant.

As for the two-locus case, one can also consider the situation where the selected

TABLE 14

The effect of a selected locus (C) on closely linked neutral loci, with $s_1 = s_2 = 0.05$ and $R_1 = 0.002$ and $R_2 = 0.001$

<table>
<thead>
<tr>
<th>Gen.</th>
<th>$p_c$</th>
<th>$p_a$</th>
<th>$p_b$</th>
<th>$D_{AB}$</th>
<th>$D_{AC}$</th>
<th>$D_{BC}$</th>
<th>$D_{ABC}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.01</td>
<td>0.05</td>
<td>0.30</td>
<td>0.000</td>
<td>0.0095</td>
<td>0.007</td>
<td>-0.0067</td>
</tr>
<tr>
<td>40</td>
<td>0.07</td>
<td>0.09</td>
<td>0.33</td>
<td>0.026</td>
<td>0.036</td>
<td>0.029</td>
<td>-0.023</td>
</tr>
<tr>
<td>80</td>
<td>0.23</td>
<td>0.17</td>
<td>0.40</td>
<td>0.067</td>
<td>0.065</td>
<td>0.056</td>
<td>-0.033</td>
</tr>
<tr>
<td>120</td>
<td>0.39</td>
<td>0.21</td>
<td>0.44</td>
<td>0.082</td>
<td>0.053</td>
<td>0.050</td>
<td>-0.023</td>
</tr>
<tr>
<td>160</td>
<td>0.46</td>
<td>0.23</td>
<td>0.45</td>
<td>0.081</td>
<td>0.034</td>
<td>0.034</td>
<td>-0.014</td>
</tr>
<tr>
<td>213</td>
<td>0.49</td>
<td>0.23</td>
<td>0.45</td>
<td>0.074</td>
<td>0.018</td>
<td>0.020</td>
<td>-0.007</td>
</tr>
</tbody>
</table>

As for Table 13 except that $R_1 = 0.002$, $R_2 = 0.01$ ($s_1 = s_2 = 0.05$).
locus is at a selective advantage and evolving towards fixation. Again, significant disequilibrium can be generated between the neutral loci.

Based on the results of the three-locus model, it is obvious that the observation of linkage disequilibrium between loci is not necessarily an indication of selection acting directly on these loci. It could merely indicate selection acting on closely linked loci. The model has considerable flexibility in terms of ways of creating linkage disequilibrium. The selection need not be acting solely on one locus. One must also consider the possibility of the cumulative effect of weak selection acting on a number of loci.

DISCUSSION

A. Selection and linkage disequilibrium

In the introduction the following question was raised: under a theory of selection, is there any quantity which one would expect to be observed consistently and whose existence would then indicate the action of selection? The obvious candidate to consider is linkage disequilibrium. There are a number of possible mechanisms which can give rise to significant disequilibrium. To determine the usefulness of $D$ as a measure of selection, we need to consider the relative importance of each of these mechanisms in creating linkage disequilibrium. The following is an attempt to identify all these possible mechanisms.

1. Selection: It is well known that certain types of selective interaction between genes can lead to equilibria with $D \neq 0$ (Kimura 1956; Lewontin and Kojima 1960; Bodmer and Felsenstein 1967; Karlin and Feldman 1970; Feldman, Franklin and Thomson 1974). It should also be pointed out that even if the selective regime is such that the equilibrium value of $D$ will be zero, it is nevertheless possible that $D$ will be maintained at nonzero values, for quite a considerable length of time, while the genes are evolving towards their equilibrium value (Thomson, in preparation).

2. Migration and admixture: Linkage disequilibrium may also be created by the mixing of two populations which are genetically different. Feldman and Christiansen (1975) have considered the equilibrium properties and rates of approach to equilibrium of two deterministic models of population subdivision. They have shown that clines in gene frequencies and significant linkage disequilibrium can be maintained for relatively long periods before equilibrium is reached. The extent of the clines and the amount of linkage disequilibrium generated by migration depend very strongly on the differences in gene frequencies in the initial populations, the recombination fraction between the two loci and the rate of migration. Closely related to the study of migration patterns is the question of admixture. It has been shown (Thomson, Bodmer and Bodmer 1976) that the existence of linkage disequilibrium for the haplotype HLA-A1, B8 (in the human histocompatibility system HLA) in an American black population appears to be quite adequately explained by admixture. This, of course, gives no explanation for the existence of persistent linkage disequilibrium for the haplotype HLA-A1, B8 in Caucasian populations.
3. **Drift:** Linkage disequilibrium between two segregating loci may be built up in a population by random drift due to small population size. Even though the expected value of the linkage disequilibrium is zero in small populations, the variance of the linkage disequilibrium may be large (Hill and Robertson 1968; Karlin and McGregor 1968). So although the average of \( D \) may be close to zero, the actual observed values could be quite different. If the population size is very small or the linkage between the two loci extremely tight, significant disequilibrium may be created by drift. In many cases these conditions will not be satisfied and drift can be ruled out as an explanation of an observed linkage disequilibrium. As pointed out by Lewontin (1974), drift cannot explain disequilibrium of the same sign in all populations, as it would be necessary to invoke a degree of migration between populations that was sufficient to make them all one large population. But, if this were the case, then the effective population number would be too large for linkage disequilibrium to occur by genetic drift.

4. **Inbreeding:** Very high levels of inbreeding are necessary to cause significant linkage disequilibrium (Degos and Bodmer 1972). Inbreeding levels in most animal populations are sufficiently low that this is not likely to be a major force. There are large numbers of plant populations which are predominantly inbreeding. The models discussed in this paper are not applicable to such populations.

5. **Selection operating on a closely linked locus:** The above results show that the evolution of a selected locus can generate linkage disequilibrium between closely linked neutral loci. A particular feature of the analysis is the demonstration of substantial, relatively long-term, transient effects. The model has considerable flexibility in terms of ways of creating linkage disequilibrium. The selection need not be acting solely on one locus. Disequilibrium can also be generated by the cumulative effect of weak selection acting on a number of loci. One of the rationales for considering the effect of a selected locus on closely linked neutral loci was the relevance of the models for interpreting data in the human histocompatibility system (HLA). We wanted to consider the effect of selection acting on, say, the postulated immune response genes within the HLA system and not in the HLA-A and B loci themselves. Based on the above results, it seems possible that the disequilibrium observed between certain alleles of the HLA system is not necessarily an indication of selection acting directly on the A and B loci. It could merely indicate selection acting on closely linked loci (Thomson, Bodmer and Bodmer 1976).

It is clearly difficult to be sure in any particular case which mechanism, or combination of mechanisms, is the most appropriate to explain an observation of linkage disequilibrium. In most populations mechanisms (3) and (4) can be ruled out. Significant levels of disequilibrium can be generated by mechanism (5). In this context, the existence of linkage disequilibrium does not necessarily indicate selection acting directly on the loci involved. However, in relation to the question of using \( D \) as a measure of selection, this mechanism indicates the action of selection at a closely linked locus, so in that sense is a measure of selec-
ASPECTS OF THE HITCHHIKING EFFECT

This leaves us to consider the question of whether or not migration and admixture can create significant values of $D$, for sufficiently long periods of time. As pointed out above, the amount of linkage disequilibrium generated by migration is very dependent on the differences in gene frequencies in the initial populations, the recombination fraction between the two loci and the rate of migration. If historical details of population movement and the possible extent of original genetic differences are known, then one could calculate the values of disequilibrium expected at present. Some preliminary results have been obtained with respect to whether or not migration is the predominant cause of observed clines of gene frequencies and linkage disequilibrium across Europe in the HLA system (THOMSON, BODMER and BODMER 1976). It is argued that since the HLA system is so variable it seems very unlikely that the difference in frequencies for a particular antigen between two populations would be large enough for migration to create the linkage disequilibrium observed for, say, the haplotype HLA-A1, B8. (In a typical Northern European population the frequency of HLA-A1 is 0.16 and of HLA-B8 is 0.13 and the linkage disequilibrium $D = 0.06$. The maximum possible value of linkage disequilibrium, for these particular allele frequencies, is approximately 0.1). KARLIN (personal communication) has also pointed out that the discrete stepping stone model may give a slower rate of decay than a continuous model of migration. The most likely situation in the HLA system seems to be that some form of selection has been involved in creating the observed linkage disequilibrium. However, migration is most likely an important factor in establishing the clines of linkage disequilibrium which are observed.

The relative importance of migration and selection remains unresolved. The theory of migration and clines needs to be further extended. It is not possible to give any definitive answer at the moment as to the use of $D$ as an indicator of selection. In many ways it appears to be not a very useful parameter. The fact that disequilibrium can be created in a number of ways not involving any selection means, that in any particular case, it is difficult to determine the cause of an observed disequilibrium. One should also remember that not all selection schemes lead to linkage disequilibrium. Also there may be overlap of stability classes, so that under the same conditions one population could go to an equilibrium with $D = 0$ and another to an equilibrium with $D \neq 0$. In particular cases, one may be able to rule out migration as a predominant force in the population, in which case consistent observation of linkage disequilibrium would most likely be due to selection. However, with criteria as strict as this it would mean that most cases of selection would usually go undetected.

B: Inversions

The interactions of linkage and selection are not only important for equilibrium properties of populations, but must also be considered in terms of the condi-
tions required for the initial increase of new gametic combinations. Bodmer and Parsons (1962) were the first to emphasize this approach. In this regard I would like to put forward an idea which has resulted from the work of the previous sections on the effect of a selected locus on closely linked neutral loci. The idea is that the existence of inversions may have very little to do with their recombination modification effect. What, in fact, we may be observing is the hitchhiking along of the whole inverted region with one or more selected loci. If an inversion occurs and within it is contained a selected gene, or a complex of selected genes, evolving towards equilibrium, then the frequency of the inversion in the population will increase, merely as a result of the frequency of this selected gene, or complex of genes, which occurs in the inversion, increasing. One may reasonably argue that this would be a fairly rare event. However, the point is that when such a rare event did occur, the effect produced would be very significant. Inversions may be the visible result of hitchhiking. In any particular situation it would be much more complicated than this. One must weigh up the effects of the recombination reduction and selection. It is impossible to construct a model which would be a realistic representation of the true situation. Many of the inversions studied in Drosophila contain a large number of loci and many complex interactive forces must be involved in both the establishment and maintenance of these inversions.

It is obvious that an inversion may be established by the hitchhiking effect. Some simple models have been looked at and it appears that selection is usually a much stronger force than recombination modification. However, at this stage it is impossible to give any estimate of the relative importance of the roles of these two forces in the establishment of inversions.

It was mentioned earlier that in a number of studies looking for linkage disequilibrium in Drosophila, most of the associations found have involved inversions. This observation would be in agreement with the establishment of inversions via the hitchhiking effect. The disequilibrium would be built up by hitchhiking, exactly as indicated previously. Inversion heterozygosity strongly suppresses recombination, even outside the limits of the inversion. So disequilibrium built up by the establishment of the inversion will be maintained for very long periods. The disequilibrium observed would not necessarily indicate selection acting on the particular locus associated with the inversion, although in the present context it would indicate selection acting somewhere in the inverted region. The establishment of inversions as a direct result of their recombination modification effect (Feldman 1972; Charlesworth and Charlesworth 1973b) would also result in a build-up of linkage disequilibrium. The magnitude of the disequilibrium which would be generated is unknown.

C: Nonequilibrium properties of populations

Consider again the length of time for which an effect is produced when a selected locus is evolving towards its equilibrium value. For heterotic selection of order 1%, it takes over 1000 generations for the new selected mutant to reach its equilibrium value (see Table 6). (In human populations 1000 generations is
approximately 25,000 years.) All this time the selected locus is having an effect on closely linked loci. As indicated previously, the actual number of loci affected may be extremely large. This raises the question, and this is by far one of the more important aspects of the present work, of whether it is ever valid to look at equilibrium properties of populations. If it takes a long time to reach the equilibrium, then by the time the population has reached the equilibrium value most likely the selection regime, either on that locus or one nearby, has altered and in effect the situation is never static.

We must seriously consider the possibility that populations are never at equilibrium. The difficulty that this raises from the point of view of population genetics is that the value of various parameters as a population is evolving towards equilibrium may be quite different from the parameter values expected at equilibrium. EWENS and FELDMAN (1976) have pointed out that even under the neutral theory, the rate at which stationarity is achieved is usually very slow. Even if stationarity has been achieved, selection at nearby loci will perturb the loci under consideration away from the stationary state, making tests of neutrality invalid.

So at any given time a number of loci is a population will not be in a stationary state. The actual number is of course completely unknown. This line of argument also implies that some of the variation observed in natural populations may represent a transient phase in the fixation of selectively advantageous mutants.

D: Pseudo-selection

Laboratory experiments are often started with a very small sample of genomes from natural populations. This creates linkage disequilibrium. The results of the hitchhiking effect emphasize the fact that this will lead to difficulties when trying to distinguish the effect of the locus being followed from surrounding loci with which it is linked. STAM (1975) has demonstrated how linkage disequilibrium causes pseudo-selection at a neutral locus in Tribolium populations. Considerable care should be taken to include a large number of founding genomes in laboratory populations when attempting to measure selective effects on a particular locus. Replicate experiments using different proportions of founding stocks and different initial frequencies of the locus being followed should aid in determining if an observed selective force is due to hitchhiking.

It has been demonstrated that the hitchhiking effect will lead to transient linkage disequilibrium. In populations with small generation time it may be possible to detect that the disequilibrium is transient (an example of transient linkage disequilibrium in Drosophila reported by O'BRIEN and MacIntrYRE (1971) may be the result of a hitchhiking effect). However, in populations with long generation times, for example man, the transient nature of the disequilibrium may not be detectable.

E. Observations of linkage disequilibrium

One can ask why so little linkage disequilibrium has been found. The reason is probably because to find significant disequilibrium it is necessary to look at
very closely linked loci. (For all the mechanisms mentioned above which create linkage disequilibrium, other things being equal, the linkage disequilibrium generated decreases monotonically with increasing recombination fraction. It follows from this that one would be less likely to observe significant disequilibrium between loosely linked than between tightly linked loci. Of course, if the forces between loosely linked loci are sufficiently strong, then significant disequilibrium will still be created.) Many of the studies looking for linkage disequilibrium possibly do not include a small enough region of the genome. It would be interesting to have more data on very tightly linked complexes of genes. A problem arises though if we look at very closely linked loci. Possibly one will always observe some disequilibrium, not necessarily created by selection acting directly on the loci, but by any one of the mechanisms mentioned above.

Karlin (1975) has argued that under general two-locus selection schemes, a polymorphic equilibrium should have $D \neq 0$. His argument is based on the fact that the existence of a polymorphic equilibrium with $D = 0$ implies special algebraic relations among the selection coefficients, as shown by Bodmer and Felsenstein (1967). (When the effects of two loci are additive or multiplicative these relations are satisfied.) However, although we know that some epistatic interaction between loci exists, there must still presumably be a large proportion of loci which are evolving virtually independently of each other. So, if we chose two loci at random, the probability that there is strong interaction between them may be fairly low. In this event the selection scheme acting on the loci would be close to additive or multiplicative and $D = 0$, or in fact $D$ close to zero, would be the usual case. In this regard, if we are looking for associations between loci which are not very tightly linked then, as a first step, it would seem more appropriate to look for associations between functionally related enzymes. The chances of epistatic interaction would possibly be higher in this case.

It has already been mentioned that the frequent observation of linkage disequilibrium between allozyme polymorphisms and inversions in Drosophila may be explained as a hitchhiking effect. This is caused by the establishment of the inversion in the population via the evolution of one or more selected loci in the inversion. Another possible explanation of the observed association is as follows. Since inversions may contain a large number of loci, the chances of another locus interacting epistatically with the whole inverted region is much greater than the chance of two loci chosen at random interacting epistatically. Hence, the observed linkage disequilibrium could be a result of selection acting directly on the observed polymorphic locus and the inverted region.

F: *The effect of selection on the whole genome*

One of the aims of this paper was to consider the evolution of the whole genome. We have found that the evolution of a selected locus may significantly alter the dynamics of a very large number of closely linked loci. Presumably, if the selection acting on a particular locus is strong, its effect can override the selective effects on all nearby loci. It would follow from this that the maximization of all traits in a population is not possible, since evolution of the genome is
not the independent evolution of all the loci. The overall effect, as regards evolutionary trends, is unknown.

We have shown that as a selected gene evolves towards equilibrium, or fixation, it has a significant effect on the level of heterozygosity at closely linked neutral loci. On the average, the heterozygosity at the neutral locus will always be reduced. One prediction from this result is that bursts of very strong selection on a large number of loci may lead to evolutionary bottlenecks, due to a significant proportion of the genome being made homozygous via the hitchhiking effect. The reduction in genetic variability would possibly make the population less able to adapt to subsequent changes in selective pressures. This prediction is of course extremely speculative.

There are many questions we cannot answer. We know that selection operates, but what proportion of loci, at any particular time, are selected? What are realistic selection values to use? Are the models themselves anywhere near reality? The models we consider obviously cannot give a definitive answer in any particular situation. Hopefully they are of considerable value in giving estimates of the effect that will be produced in a population under given assumptions. Obviously we must develop more realistic models, especially those involving variable selection coefficients and also multi-allelic, multi-locus models. Our scope, however, may be limited. The selection which makes for evolutionary change is probably of a small order of magnitude and therefore very difficult to detect. Also, to detect the effect of selection we need to test parameter values predicted from the theory. But to construct an adequate theory we need to know what proportion of loci are selected. This tautology may be a serious stumbling block. Some of these difficulties may be alleviated as our knowledge of the biochemistry and function of particular allelic variants increases. With this knowledge our model building may become less abstract.

I would like to thank Ellen Solomon for many helpful discussions in the course of preparing this paper. Also, for their comments and critical reading of the manuscript, I thank Walter Bodmer, Bryan Clarke, Ian Craig, Mary Ellis, Warren Ewens, Peter Goodfellow, John Maynard Smith, David Roberts and Nick Wedd.

LITERATURE CITED


Histocompatibility Testing, 1972 Edited by Jean Dausset and Jacques Colombani, Munksgaard, Copenhagen.

Histocompatibility Testing, 1975 Edited by F. Kissmeyer-Nielsen, Munksgaard, Copenhagen.


Jennings, H. S., 1917 The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage. Genetics 2: 97–154.
ASPECTS OF THE HITCHHIKING EFFECT 787


Corresponding editor: W. J. Ewens