Chapter 4

Evolutionary Genetics
MODIFICATION OF LINKAGE INTENSITY
BY NATURAL SELECTION

MASATOSHI NEI

Division of Genetics, National Institute of Radiological Sciences, Chiba, Japan

Received June 8, 1967

GENETIC linkage or linkage of genes is one of the common genetic features of all organisms from viruses to man. It may be of great concern to evolutionary geneticists how genetic linkage was developed and how the intensity of linkage has been adjusted in the evolutionary process. FISHER (1930) suggested that "the presence of pairs of factors in the same chromosome, the selective advantage of each of which reverses that of the other, will always tend to diminish recombination, and therefore, to increase the intensity of linkage in the chromosomes of that species." He also recognized another agency which would increase recombination, namely, the constant spread of advantageous mutations which, unless they occur so seldom that each has become predominant before the next appears, can only come together in the same gamete by means of recombination. This problem was recently discussed at length but still semi-quantitatively by BODMER and PARSONS (1962).

There is a large amount of evidence that linkage intensity is under genetic control (cf. BODMER and PARSONS 1962). In Drosophila ananassae, for example, MORIWAKI (1940) identified a dominant gene, En-2, located on the right arm of the second chromosome, which enhances recombination between almost every pair of loci on the same chromosome in both males and females, while KIKKAWA (1937) found another gene (or genes) which induces crossing over in males in the third chromosome. CLARK and MARGULIES (1965) and HOWARD-FLANDERS and THERIOT (1966) also reported several recombination-deficient mutants in Escherichia coli K-12, which are probably due to single-gene mutation. Further, the effectiveness of artificial selection in reducing or increasing recombination values was reported by DETELFSEN and ROBERTS (1921), PARSONS (1958), and MUKHERJEE (1961) in Drosophila.

The purpose of this paper is to present mathematical models for the development of genetic linkage and the modification of linkage intensity by natural selection. In the present paper we shall consider only those agencies which would increase the linkage intensity. KIMURA (1956) developed a model of a genetic system which leads to closer linkage by natural selection in diploid organisms, but he was concerned mainly with a crossover-reducing mechanism such as an inversion.

1 This investigation was supported in part by a grant from the Scientific Research Fund of the Ministry of Education, Japan, and also by a grant from the Toyo Rayon Foundation.

Haploid system

Model 1: The modification of linkage intensity may occur in two different manners, namely, through structural changes of chromosomes and through selection of modifier genes. When a structural change such as an inversion or translocation occurs, the genes which were located far apart on the same chromosome or on different chromosomes before the change, may come close together. The present model represents this type of modification. Structural changes of chromosomes are known to occur in fungi (cf. Barry 1967) and in bacteria (Jacobs and Wollman 1961; Beckwith and Signer 1966). It should also be noted that a certain type of development of new linkage may be represented by this model, as will be discussed later.

Consider two pairs of genes each with two alleles, A-a and B-b, and suppose that these two pairs of genes are linked with a recombination value \( r \), and that the fitnesses and the frequencies before selection of the four possible genotypes are as follows:

<table>
<thead>
<tr>
<th>Genotype</th>
<th>( AB )</th>
<th>( Ab )</th>
<th>( aB )</th>
<th>( ab )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitness</td>
<td>( W_1 )</td>
<td>( W_2 )</td>
<td>( W_3 )</td>
<td>( W_4 )</td>
</tr>
<tr>
<td>Frequency</td>
<td>( p_1 )</td>
<td>( p_2 )</td>
<td>( p_3 )</td>
<td>( p_4 )</td>
</tr>
</tbody>
</table>

As shown by Felsenstein (1965) and Kimura (1965), the amounts of changes in \( p_1, p_2, p_3, \) and \( p_4 \) per generation are then given by

\[
\begin{align*}
\Delta p_1 &= p_1(W_1 - \bar{W})/\bar{W} - rD \\
\Delta p_2 &= p_2(W_2 - \bar{W})/\bar{W} + rD \\
\Delta p_3 &= p_3(W_3 - \bar{W})/\bar{W} + rD \\
\Delta p_4 &= p_4(W_4 - \bar{W})/\bar{W} - rD
\end{align*}
\]

respectively, where \( \bar{W} = \sum_{i=1}^{4} p_iW_i \) and \( D = (p_1W_1p_3W_4 - p_2W_2p_3W_3)/\bar{W}^2 \).

Now suppose that a new completely linked gene pair, \( \hat{a}b \), arises from \( ab \) by some mechanism, and let \( x \) and \( p_i^* \) be the frequencies of \( \hat{a}b \) and \( ab \), respectively, with \( x + \sum_{i=1}^{4} p_i^* = p_4 \). We assume that the fitness of \( \hat{a}b \) remains the same as that of \( ab \). Then we have

\[
\Delta p_i^* = p_i^*(W_i - \bar{W})/\bar{W} - rD^*
\]

where \( D^* = (p_1W_1p_4^*W_4 - p_2W_2p_3^*W_3)/\bar{W}^2 \). Hence, genotype \( \hat{a}b \) has a selective advantage over \( ab \), if \( rD^* \) is positive. This quantity or \( D^* \) is expected to be positive only when the coupling genotypes are fitter than the repulsion ones. When \( D^* \) or \( D \) is negative, those linkages which occur in \( Ab \) or \( aB \) have a selective advantage. If there is no epistasis, \( D \) or \( D^* \) is expected to be 0 at least after several cycles of random mating, so that \( \hat{a}b \) has no selective advantage over \( ab \).

Anyway, if \( W_4 - \bar{W} \) is positive, genotype \( aB \) is expected to increase in frequency. In the initial stage of increase in \( x \), \( \bar{W} \) is almost constant if the differences between \( W' \)'s are small. In this situation,

\[
\Delta x/x = (W_4 - \bar{W})/\bar{W} = \alpha
\]

where \( \bar{W} \) is the equilibrium value of \( \bar{W} \) before the introduction of \( \hat{a}b \). Therefore,
the frequency of \( \hat{ab} \) in the initial stage of its increase is approximately given by

\[
x_n = x_0 e^{an}
\]

(3)

where \( x_n \) denotes the value of \( x \) in the \( n \)th generation.

As an example, consider the case where \( r = 0.5 \), \( W_1 = 1 \), \( W_2 = 1 - s_a \), \( W_3 = 1 - s_b \), and \( W_4 = 1 + t \), \( s_a \), \( s_b \), and \( t \) being small positive constants. In the absence of linked gene pair \( \hat{ab} \), there is an unstable equilibrium for the gene frequency of \( a \) or \( b \), below which these gene frequencies or \( p_a \) cannot increase even if \( ab \) is the fittest genotype. The equilibrium values of gene frequencies of \( a \) and \( b \) are approximately given by \( s_b/(s_a + s_b + t) \) and \( s_a/(s_a + s_b + t) \) respectively (Caow and Kimura 1965). Below these equilibrium values, the genes \( a \) and \( b \) are maintained only by mutation, so that the frequency of \( ab \) is very small. However, once the linked gene pair \( \hat{ab} \) is introduced, it is expected to increase rapidly after some period of stochastic changes. In this case

\[
\Delta x = x(1 - x) t/W
\]

\[
\approx x(1 - x) t
\]

Hence, the frequency of \( \hat{ab} \) in the \( n \)th generation is approximately given by

\[
x_n = x_0 e^{ia}/(1 - x_0 + x_0 e^{ia})
\]

(4)

Therefore, \( x_n \) increases logistically. For example, if \( t = 0.06 \) and \( x_0 = 0.001 \), about 230 generations are required for \( x_n \) to reach 0.99.

Model 2: In this model it is assumed that the recombination value between a pair of loci is controlled by another modifier locus and the modification of linkage intensity is brought about by an increase in the frequency of the modifier gene. Consider two loci each with two alleles, \( A-a \) and \( B-b \), and suppose that the recombination value between these two loci is modified by another locus, i.e. \( M-m \). For simplicity we assume that this modifier locus is independent of the other two loci and has no biological function other than modifying the recombination value. For the loci \( A \) and \( B \) there are four possible genotypes, i.e., \( AB \), \( Ab \), \( aB \), and \( ab \), the fitnesses of which are denoted again by \( W_1 \), \( W_2 \), \( W_3 \) and \( W_4 \), respectively. Let the frequencies of \( AB \), \( Ab \), \( aB \), and \( ab \) \textit{after selection} in a generation be \( P_1 \), \( P_2 \), \( P_3 \), and \( P_4 \) respectively \( (P_1 + P_2 + P_3 + P_4 = 1) \), and the frequencies of genes \( M \) and \( m \) be \( x \) and \( \gamma := (1 - x) \). Note that \( P_i = p_i W_i/W \). Then, under random mating the frequencies of genotypes \( ABM \), \( ABm \), \( AbM \), etc. will be \( P_1 x, P_2 y, P_3 x, \text{etc.} \), respectively.

There are ten possible mating types for the loci \( A \) and \( B \), as given in Table 1. Each of these mating types can be subdivided into three groups according to the mating type for the modifier locus, i.e. \( M \times M \), \( M \times m \), and \( m \times m \) with frequencies \( x^2, 2xy, \) and \( y^2 \) respectively. Except for the two mating types \( AB \times ab \) and \( Ab \times aB \), this subdivision is irrelevant to the production of genotypes in the next generation. With \( AB \times ab \) and \( Ab \times aB \) the frequencies of offspring genotypes are dependent on the recombination value, and let the recombination values for matings \( M \times M, M \times m \), and \( m \times m \) be \( r_2 \), \( r_3 \), and \( r_0 \), respectively, with \( r_2 \leq r_3 \leq r_0 \).

In bacteria, the recombination of genes is accomplished by any of several processes—sexual mating, transduction, and transformation, but the recombination at a particular time is almost always restricted to a short segment of chro-
mosome. If, therefore, the modifier locus is located far-off from the A and B loci, the pairing of modifier genes does not usually take place in association with the pairing of the A and B loci. Thus, each mating type for the A and B loci is subdivided into only two groups, M and m, which show recombination values of, say, \( r'_1 \) and \( r'_0 \) respectively. However, this situation can be covered by the above treatment with a simple transformation \( r'_0 - r'_1 = 2(r_0 - r_1) = 2(r_1 - r_0) \).

Since the modifier locus is independent of the other two loci, the frequency of genotype \( ABM \) before selection in the next generation becomes
\[
x[P_2^2 + P_1P_2 + P_2P_3 + P_3P_4(1 - x_{r_2} - y_{r_1}) + P_2P_3(x_{r_2} + y_{r_1})] = x(P_1 - r_2D)
\]
where \( r_m = x_{r_2} + y_{r_1} \) and \( D = P_1P_4 - P_3P_5 \), which is identical to the previous \( D \).

For other genotypes we have

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>( AbM )</td>
<td>( x(P_3 + r_2D) )</td>
</tr>
<tr>
<td>( abM )</td>
<td>( x(P_3 + r_2D) )</td>
</tr>
<tr>
<td>( Abm )</td>
<td>( y(P_4 - r_2D) )</td>
</tr>
<tr>
<td>( Abm )</td>
<td>( y(P_4 + r_2D) )</td>
</tr>
<tr>
<td>( abm )</td>
<td>( y(P_4 + r_2D) )</td>
</tr>
</tbody>
</table>

where \( r_m = x_{r_2} + y_{r_0} \). Thus, the average fitness of genotype \( M \) becomes
\[
\bar{W}_M = (P_1 - r_2D)W_1 + (P_2 + r_2D)W_2 + (P_3 + r_2D)W_3 + (P_4 - r_2D)W_4
= \bar{W}_E - r_2D\epsilon
\]
where \( \bar{W}_E \) is \( \sum_{i=1}^4 P_iW_i \) and \( \epsilon = W_1 - W_2 - W_3 + W_4 \). On the other hand, the fitness of genotype \( m \) is given by
\[
\bar{W}_m = \bar{W}_E - r_2D\epsilon
\]
and the average fitness of population by

\[ \bar{W} = x \bar{W}_m + \gamma \bar{W}_m = \bar{W}_m - \bar{r} D \varepsilon \]  

(7)

where \( \bar{r} = x r_m + \gamma r_m \).

Therefore, the genotype or gene frequency of \( M \) after selection is given by

\[ x' = x \frac{\bar{W}_m}{\bar{W}} \]  

(8)

and the amount of change in gene frequency per generation becomes

\[ \Delta x = (xy/\bar{W}) (r_m - r_M) D \varepsilon . \]  

(9)

Thus, if \( D \) and \( \varepsilon \) are both positive or both negative, the frequency of the modifier gene always increases.

In an arbitrary random mating population, the signs of \( D \) and \( \varepsilon \) are not necessarily the same. However, Felsenstein (1965) and Kimura (1965) showed that, in the absence of a modifier gene, the sign of \( D \) is quickly adjusted to be the same as that of \( \varepsilon \) under continuous selection. Therefore, it can be said that the frequency of a recombination-reducing gene generally increases if the loci of which the recombination value is to be modified show epistasis. In other words, for linkage intensity to be modified there must be epistasis, as in Model 1. Note also that, if \( r_M \geq r_i \geq r_0, r_m - r_M \) and consequently \( \Delta x \) in (9) become negative, so that a modifier gene which enhances recombination can not increase in frequency. This implies that natural selection never increases recombination, if the fitnesses of genotypes remain unchanged.

Kimura (1965) further showed that, in the absence of a modifier gene, if the population reaches a quasi linkage equilibrium, i.e. the state where \( \Delta \log(P_1P_4/P_2P_3) = 0 \) or \( P_1P_4/P_2P_3 = 1 \pm \varepsilon/r_0 \) approximately, the epistatic variance is given by \( V_{ep} = r_0 WD \). Note that this state is quickly attained if \( \varepsilon \) is much smaller than \( r_0 \). Now suppose that in course of time a new mutation occurs at the modifier locus from \( m \) to \( M \) and its frequency \( x \) starts to increase. In the initial stage of the increase where \( x \) is small, \( r_m - r_M \) is approximately equal to \( cr_0 \), where \( c \) is the proportion of recombination value reduction in heterozygotes, i.e. \( (r_0 - r_1)/r_0 \). Therefore, the rate of increase of \( x \) is approximately given by

\[ \Delta x = xy(cV_{ep})/\bar{W}^2 \]  

(10)

which further approximates to

\[ xy(cV_{ep})/\bar{W} \]  

(10')

if selection is mild and \( \bar{W} \) is close to 1. The rate of increase of \( x \) is obtained also by using a continuous time model and Malthusian fitness parameters instead of Darwinian fitnesses as used here (cf. Kimura 1956; Felsenstein 1965). It becomes

\[ \frac{dx}{dn} = xy(cV_{ep}) \]  

(11)

Formula (10') or (11) shows an important property that in the initial progress of the modifier gene the selective difference between the modifier and wild-type genes is equal to the proportion of recombination value reduction in heterozygotes multiplied by the epistatic variance of the loci of which the recombination value is to be modified. It is of interest to see that this property is similar to Fisher's (1930) fundamental theorem of natural selection, i.e. the rate of increase
of fitness is equal to the additive genetic variance at that time. When \( r_0, r_1, \) and \( r_2 \) are all large compared with \(|\varepsilon|\) and a quasi linkage equilibrium holds, the rate of change of \( x \) in the whole process of the modification of recombination value may be given by (10) or (11), replacing \( c = (r_0 - r_1)/r_0 \) by \( c' = (r_m - r_u)/\bar{r} \).

In some specific cases formula (9) or (10) reduces to a simpler form. As an example, consider the case where \( W_1 = W_4 = 1 - s \) and \( W_2 = W_3 = 1 \) \((1 > s > 0)\). In the absence of modifier gene the equilibrium genotype frequencies of \( AB, Ab, aB, \) and \( ab\) are given by

\[
\hat{P}_1 = \hat{P}_4 = \frac{1}{4} \left[ 1 + \left( \frac{a}{2} \right) + a\beta - \sqrt{1 + a^2 (\frac{1}{4} + \beta^2)} + a(1 + a\beta) \right] \tag{12a}
\]
\[
\hat{P}_2 = \hat{P}_3 = \frac{1}{2} - \hat{P}_1 \tag{12b}
\]

where \( a = r/(1 - r) \) and \( \beta = (1 - s)/s \). This equilibrium is stable only if \( P_2 = P_3 \)

Jain and Allard (1966) called this type of equilibrium isoplethic. (Finney 1952) used this word for a different meaning. In this connection it should be noted that in the haploid system there is no mechanism for maintaining a strictly stable polymorphism except frequency-dependent selection or alternation of selective advantage between \( AB \) and \( ab \) due to environmental changes in different generations. If we superimpose one of these two factors into the present model, then the equilibrium would become stable.

Now suppose that a new mutation arises which reduces the recombination value. After some stochastic changes, this gene is expected to increase in frequency and eventually replace its original type gene in an infinite population. During this replacement the genotype frequencies of \( AB, Ab, aB, \) and \( ab \) change from the values for \( \alpha = r_0/(1 - r_0) \) to those for \( \alpha = r_2/(1 - r_2) \) in (12 a, b). The selective difference between the modifier and original type genes is dependent on \( x \) except for the case of \( s = 1 \) and \( r_0 - r_1 = r_1 - r_2 \) (additive gene action). For example, if \( s = 0.5, r_0 = 0.1, r_1 = 0.05, \) and \( r_2 = 0.0 \), the selective advantage of modifier gene, which is given by \( (r_m - r_u)Dx \), changes from 0.0113 to 0.0125 as \( x \) increases from 0 to 1. If, on the other hand, \( r_0 = 0.1, r_1 = r_2 = 0.0 \), i.e. the modifier gene shows a complete dominance, \( r_m - r_u \) decreases as \( x \) increases, while \( D \) increases very slightly, so that the selective advantage of the modifier gene gradually decreases as time proceeds, namely, from 0.05 to 0. Figure 1 shows the corresponding changes in \( x \) and \( \bar{r} \). These changes were obtained numerically by using formula (8) and the following expressions for the frequencies of \( AB, Ab, aB \) and \( ab \) in the next generation.

\[
P_1' = (P_1 - \bar{e}D)W_1/\bar{W} \tag{13a}
\]
\[
P_2' = (P_2 - \bar{e}D)W_2/\bar{W} \tag{13b}
\]
\[
P_3' = (P_3 - \bar{e}D)W_3/\bar{W} \tag{13c}
\]
\[
P_4' = (P_4 - \bar{e}D)W_4/\bar{W} \tag{13d}
\]

It is seen from Figure 1 that the rates of changes in \( x \) and \( \bar{r} \) when \( x \) is small are larger in the case of complete dominance than in the case of incomplete dominance but gradually become smaller in the former than in the latter.

When \( s = 1, e = -2, \) and \( D \) is always \(-\frac{1}{4}\) irrespective of the value of \( x \), provided that the initial values of \( P_1, P_2, P_3, \) and \( P_4 \) are 0, 0.5, 0.5, and 0, respectively, which are the equilibrium values in the absence of modifier gene. Further, if
Figure 1.—Changes in the modifier gene frequency and the mean recombination value due to natural selection; haploid system. Fitness, $W_1 = 1$, $W_2 = W_3 = 0.5$, $W_4 = 1$. The solid line refers to the case of $r_0 = 0.1$, $r_1 = 0.05$, $r_2 = 0.0$ and the broken line to the case of $r_0 = 0.1$, $r_1 = r_2 = 0.0$.

$r_0 - r_1 = r_1 - r_2 = \Delta r$, the value of $r_m - r_M$ is independent of $x$ and given by $\Delta r$. Therefore, the rate of change in $x$ is given by

$$\Delta x = (xy/W) (\Delta r/2)$$

(14)

Since $W$ is $1 - r_0/2 + \Delta rx$, the rate of change in $x$ may be represented by

$$\frac{dx}{dn} = \frac{xy\Delta r}{2(1 - r_0/2 + \Delta rx)}$$

(14’)

Solution of this equation gives

$$u_n = u_0 + (\Delta r/2)n$$

(15)

where

$$u = [1 - (r_0/2)] \log_e [x/(1 - x)] - \Delta r \log_e (1 - x).$$

Thus, if $\Delta r$ is large, the frequency of modifier gene increases fairly rapidly. For example, if $r_0 = 0.1$, $r_1 = 0.05$, and $r_2 = 0$ with $\Delta r = 0.05$, $x$ increases from 0.001 to 0.999 in about 540 generations. The mean recombination value correspondingly decreases from 0.0999 to 0.0001. Therefore, the modification of linkage intensity is much more rapid in this case than in the case of $s = 0.5$ (cf. Figure 1). If the modifier gene shows a complete dominance over the original gene, i.e. $r_1 = r_2$, the quantity $r_m - r_M$ is no longer independent of $x$, and it is difficult to obtain an analytical solution. However, a numerical computation has shown that the changes in $x$ and $r$ are of the same pattern as those for $s = 0.5$, though the modification of recombination value is again more rapid in this case.

In the above examples, the $A$ and $B$ loci remain polymorphic not only in the process of modification of recombination value but also after the completion of this process. Particularly when $s = 1$, the genotype frequencies for these two loci
remain the same during the whole process of modification of recombination value. In all cases, however, the mean fitness for these loci gradually increases as the recombination value is reduced.

The modification of linkage intensity also occurs when the A and B loci show a transient polymorphism or they are going to fixation, if these show epistasis. Consider an example in which \( W_1 = 1, \ W_2 = W_3 = 0.5, \) and \( W_4 = 1.1. \) As mentioned previously, this set of fitnesses creates an unstable equilibrium below which the frequency of \( ab \) cannot increase in the absence of a modifier gene. If the recombination value is allowed to be modified, the increase or decrease of the \( ab \) frequency depends on the initial values of \( x, P_3, P_9, P_4, \) as well as on the values of \( r_0, r_1, \) and \( r_2. \) Table 2 shows the results of a numerical computation, in which the initial values of \( x, P_3, P_9, \) and \( P_4 \) were 0.01, 0.49, 0.21, 0.21, and 0.09, respectively, with \( r_0 = 0.1, \) and \( r_1 = r_2 = 0. \) It is seen that in this case the frequency of \( ab \) gradually increases and the population eventually becomes monomorphic, consisting of only \( ab. \) During this process \( x \) increases from 0.01 to 0.025, and the mean recombination value is reduced from 0.098 to 0.095. Therefore, the modification of recombination value is very slight. Of course, if \( x_0 \) is close to 0.5, the modification is more rapid and greater, though it is never completed unless the fixation of the modifier gene is assisted by random genetic drift. For example, if \( x \) starts from 0.5, it increases up to 0.598 before the population becomes monomorphic, while \( r \) changes from 0.025 to 0.016. Moreover, if \( r_0 \) and \( r_1 \) are sufficiently large and the gene frequencies of \( a \) and \( b \) are small, the genotype \( ab \) is not established in the population but eventually eliminated. In this case the modification of recombination value becomes even smaller.

**Diploid system**

**Model 1:** This model corresponds to Model 1 in the haploid system and is essentially the same as Kimura’s (1956). Consider two loci \( A-a \) and \( B-b \) as before, and let \( p_1, p_2, p_3, \) and \( p_4 \) be the frequencies after meiosis of the four types of gametes \( AB, Ab, aB, \) and \( ab \) respectively. We denote the fitness of the genotype formed by the union of the \( i \)th and \( j \)th gametes by \( W_{ij} \) as shown in Table 3. Then,

**TABLE 2**

An example showing the modification of linkage intensity when the A and B loci show a transient polymorphism

<table>
<thead>
<tr>
<th>Generation</th>
<th>( x )</th>
<th>( P_3 )</th>
<th>( P_4 )</th>
<th>( P_4 ) ( = P_3 )</th>
<th>( P_4 )</th>
<th>( D )</th>
<th>( \bar{r} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.01000</td>
<td>0.49000</td>
<td>0.21000</td>
<td>0.09000</td>
<td>0.00000</td>
<td>0.09801</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.01137</td>
<td>0.75096</td>
<td>0.01590</td>
<td>0.21723</td>
<td>0.16288</td>
<td>0.09774</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.01379</td>
<td>0.65448</td>
<td>0.01917</td>
<td>0.30718</td>
<td>0.20068</td>
<td>0.09726</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.01734</td>
<td>0.48227</td>
<td>0.02194</td>
<td>0.47385</td>
<td>0.22804</td>
<td>0.09656</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.02159</td>
<td>0.32126</td>
<td>0.01723</td>
<td>0.73341</td>
<td>0.16994</td>
<td>0.09573</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.02481</td>
<td>0.00910</td>
<td>0.00105</td>
<td>0.98881</td>
<td>0.01086</td>
<td>0.09510</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.02493</td>
<td>0.00000</td>
<td>0.00000</td>
<td>1.00000</td>
<td>0.00000</td>
<td>0.09508</td>
<td></td>
</tr>
</tbody>
</table>

**Fitnesses:** \( W_1 = 1, W_2 = W_3 = 0.5, W_4 = 1.1; \) recombination values, \( r_0 = 0.1, r_1 = r_2 = 0.0. \)
MODIFICATION OF LINKAGE INTENSITY 633

TABLE 3

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>AB</th>
<th>Ab</th>
<th>ab</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>$W_{11}$</td>
<td>$W_{12}$</td>
<td>$W_{13}$</td>
</tr>
<tr>
<td>Ab</td>
<td>$W_{21}$</td>
<td>$W_{22}$</td>
<td>$W_{23}$</td>
</tr>
<tr>
<td>aB</td>
<td>$W_{31}$</td>
<td>$W_{32}$</td>
<td>$W_{33}$</td>
</tr>
<tr>
<td>ab</td>
<td>$W_{41}$</td>
<td>$W_{42}$</td>
<td>$W_{43}$</td>
</tr>
</tbody>
</table>

$W_{ij} = W_{ji}$.

the amounts of change in gamete frequencies per generation are given by

\[
\Delta p_1 = \frac{p_1(W_{11} - \bar{W}) - rD_w}{\bar{W}}
\]

\[
\Delta p_2 = \frac{p_2(W_{12} - \bar{W}) + rD_w}{\bar{W}}
\]

\[
\Delta p_3 = \frac{p_3(W_{13} - \bar{W}) + rD_w}{\bar{W}}
\]

\[
\Delta p_4 = \frac{p_4(W_{14} - \bar{W}) - rD_w}{\bar{W}}
\]

as shown by Kimura (1956, 1965) and Lewontin and Kojima (1960), where $W_i = \sum_{j=1}^{4} p_{ij}W_{ij}$, $\bar{W} = \sum_{i=1}^{4} p_iW_i$, and $D_w = p_1p_3W_{14} - p_2p_3W_{23}$.

Now suppose that a new completely linked gene pair $\bar{a}b$ is formed from $ab$ and let the frequencies of $\bar{a}b$ and $ab$ be $x$ and $p_x^4$, respectively, with $x + p_x^4 = p_x$. The amounts of changes in $p_x^4$ and $x$ are then given by

\[
\Delta p_x^4 = \frac{[p_x^4(W_{44} - \bar{W}) - rD_w^x]}{\bar{W}}
\]

\[
\Delta x = x(W_{44} - \bar{W})/\bar{W}
\]

(16a)

(16b)

where no crossing over is assumed to occur in such genotypes as $AB/\bar{a}b$ and $ab/\bar{a}b$ and $D_w^x$ denotes $p_3p_x^4W_{14} - p_2p_x^4W_{23}$. Formula (16b) is equivalent to Kimura's (1956) formula (8). The initial progress of gamete $\bar{a}b$ is again given by formula (3), replacing $W_4$ by $\bar{W}_4$, i.e. the equilibrium value of $W_4$. Therefore, this type of modification of linkage intensity is fairly rapid, if $\bar{W}_4 - \bar{W}$ is large.

Model 2: This model represents the modification of linkage intensity by selection of modifier genes. As before, we consider two loci each with two alleles, $A-a$, and $B-b$, and designate the fitnesses of genotypes as given in Table 3. Let $p_{ij}$ be the frequency after selection of the genotype formed by the union of the $i$th and $j$th gametes, namely $p_{ij} = p_i p_j W_{ij}/\bar{W}$. Then, $P_i = \sum_{j=1}^{4} p_{ij}$ is the frequency of the $i$th type chromosome before meiosis. This will be called chromosome frequency. We will designate by $x$ and $y' (= 1 - x)$ the frequencies of modifier gene $M$ and its allelic gene $m$ respectively. Each genotype for loci $A$ and $B$ can be subdivided into three groups according to the genotype of the modifier locus, i.e. $MM$, $Mm$, and $mm$. But this subdivision is irrelevant to the gamete types produced, except for double heterozygotes. We again assume that the modifier locus is independent of both loci $A$ and $B$. In the computation of gamete frequencies Table 1 can be used if we replace mating type by genotype and genotype produced by gamete produced. Thus, $AB \times AB$, $AB \times Ab$, etc. are replaced by $AB\bar{A}B$, $AB\bar{A}b$, etc., and $P_x^3$, $2P_1P_2$, etc. by $P_{11}$, $2P_{12}$, etc. It will be easily
seen that the frequencies of various gamete types become as given in Table 4. In this table, \( D \) denotes \( P_{14} - P_{23} \) or \( D_0/W \).

The mean fitnesses of genotypes \( MM, Mm, \) and \( mm \) in the next generation are obtained as follows:

\[
\begin{align*}
\bar{W}_{MM} &= (P_1 - r_M D)^2 W_{11} + 2 (P_1 - r_M D) (P_2 + r_M D) W_{12} + 2 (P_3 - r_M D) (P_4 - r_M D) W_{14} + (P_2 + r_M D)^2 W_{22} + 2 (P_2 + r_M D) (P_3 + r_M D) W_{23} + 2 (P_3 - r_M D) (P_4 - r_M D) W_{33} + 2 (P_4 - r_M D)^2 W_{44} = W_R - 2 r_M D \varepsilon_A + r_M^2 D^2 \varepsilon_D \\
\bar{W}_{Mm} &= W_R - (r_M + r_m) D \varepsilon_A + r_M r_m D^2 \varepsilon_D \\
\bar{W}_{mm} &= W_R - 2 r_m D \varepsilon_A + r_m^2 D^2 \varepsilon_D
\end{align*}
\]

where \( W_R = \sum_{i=1}^{4} P_i \sum_{j=1}^{4} W_{ij} \) and \( E_i = W_{11} - W_{12} - W_{13} + W_{14} \). The average fitnesses of gene \( M \) and \( m \) (\( \bar{W}_M \) and \( \bar{W}_m \)) and the total population (\( \bar{W} \)) then become:

\[
\begin{align*}
\bar{W}_M &= x \bar{W}_{MM} + y \bar{W}_{Mm} = W_R - (r_M + r) D \varepsilon_A + r_M r \bar{D}^2 \varepsilon_D \\
\bar{W}_m &= W_R - (r_m + \bar{r}) D \varepsilon_A + r_m \bar{r} \bar{D}^2 \varepsilon_D \\
\bar{W} &= W_R - 2 \bar{D} \varepsilon_A + \bar{D}^2 \varepsilon_D
\end{align*}
\]

Therefore, the gene frequency of \( M \) in the next generation is given by \( x \bar{W}_M/\bar{W} \), and the amount of change per generation is

\[
\Delta x = \frac{x y}{\bar{W}} (r_m - r_M) D (\varepsilon_A - \bar{D} \varepsilon_D).
\]

The quantity \( \varepsilon_A - \bar{D} \varepsilon_D \) in the above expression is equivalent to KIMURA’s (1965) \( \bar{e} \), and he showed that the sign of \( D \) is quickly adjusted to be the same as that of \( \bar{e} \) and the population again reaches quasi linkage equilibrium when selection is mild. Therefore, the frequency of a modifier gene generally increases if the loci of which the recombination value is to be modified show epistasis. When there is no epistasis, both \( \varepsilon_A \) and \( \varepsilon_D \) are 0, so that no modification of linkage intensity is possible.

As shown by KIMURA (1965), the chromosomal epistatic variance in the state of quasi linkage equilibrium is given by \( \bar{V}_{eD} = 2 r_0 \bar{W} \bar{D} \bar{e} \) in the absence of a modifier gene. Thus, in the initial progress of the modifier gene we again have
\[ \Delta x \approx \frac{xy}{W} \left( \frac{c}{2} V_{ep} \right) \]  

(21)

or

\[ \frac{dx}{dn} = xy \left( \frac{c}{2} V_{ep} \right) \]  

(22)

where \( c \) is the same as before. Therefore, the rule established for the haploid system holds true also for the diploid system with a slight modification, i.e., substituting the epistatic variance by half the epistatic variance.

There are numerous types of balanced polymorphisms involving epistasis in diploid populations. Fisher (1930) considered a particular type of epistasis which would be favorable for reducing the recombination value. The fitness matrix given in (a) in Table 5 is one of this type of epistasis. The equilibrium gamete frequencies and their stability conditions for this type of epistasis were investigated by Kimura (1956). The present model corresponds to the case of \( s = 0.1 \) and \( t = 0.2 \) in his formula, and the equilibrium frequencies of \( AB, Ab, aB, \) and \( ab \ after \ meiosis \) \( (\hat{p}_1, \hat{p}_2, \hat{p}_3, \) and \( \hat{p}_4) \) are given by

\[
\hat{p}_1 = \hat{p}_4 = \frac{1}{2} \left[ \frac{1}{2} - \frac{(r/s)}{1/4 + (r/s)^2} \right] \\
\hat{p}_2 = \hat{p}_3 = \frac{1}{2} - \hat{p}_1
\]

where \( r \) is the recombination value in the absence of a modifier gene. In this case stable equilibria obtain only if \( r < 0.0375 \). For other values of \( r \), however, the isoplethic equilibria apply if \( p_2 = p_3 \) or \( P_2 = P_3 \). Note that \( p_1, p_2, \) etc. can be transformed into \( P_1, P_2, \) etc. by the formulas in the column ‘Total’ in Table 4. The selective advantages of a modifier gene in the initial stage of its increase when \( r_1 \) is small, \( r_0 D (s_A - r_0 D e_0) \), for various values of \( r_0 \) are given in column (a) in Table 6. It is clear that the selective advantage is very small for any value of \( r_0 \), so that a large number of generations are required for the complete modification of linkage intensity. A stronger type of epistasis, where \( s \) and \( t \) in Kimura’s formula get large, creates a larger selective advantage of the modifier gene than the above, but the stability condition becomes more severe (the critical recombination value below which the stable equilibria exist becomes smaller), though the isoplethic equilibria are possible for any degree of epistasis.

The second set of genotype fitnesses (b) in Table 5 represents an intermediate optimum model with heterozygote advantage. This model was first studied by Wright (1952, 1965) in relation to Mather’s (1943) concept on the relational

---

**TABLE 5**

**Fitness matrices used for computations of the selective advantage of recombination-reducing gene**

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>(a) Aa</th>
<th>aa</th>
<th>AA</th>
<th>(b) Aa</th>
<th>aa</th>
<th>AA</th>
<th>(c) Aa</th>
<th>aa</th>
<th>AA</th>
<th>(d) Aa</th>
<th>aa</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB</td>
<td>.9</td>
<td>1.0</td>
<td>.7</td>
<td>.60</td>
<td>.95</td>
<td>1.00</td>
<td>.5000</td>
<td>.5625</td>
<td>.3750</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bb</td>
<td>.8</td>
<td>1.0</td>
<td>.8</td>
<td>.95</td>
<td>1.10</td>
<td>.95</td>
<td>.5000</td>
<td>1.0000</td>
<td>.4375</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>bb</td>
<td>.7</td>
<td>1.0</td>
<td>.9</td>
<td>1.00</td>
<td>.95</td>
<td>.60</td>
<td>.3750</td>
<td>.3125</td>
<td>.3750</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Model due to (Kimura) (Wright) (Lewontin) (Haldane)
balance of polygenes. Later, Bodmer and Parsons (1962) and Parsons (1963) stressed that this type of model is particularly favorable for bringing about the relational balance of genes and consequently for reducing recombination values. The selective advantages of recombination-reducing gene for various initial values of $r_0$ are given in Table 6 (b). It is seen that the selective advantage is larger in this model than in model (a). In this and the following models of epistasis there are stable equilibria of gamete frequencies before and after the modification of recombination value.

The third model (c), which is a case of cumulative overdominance, was numerically investigated by Lewontin (1964), and the selective advantages of a modifier gene in this case can be obtained from Table 5 in his paper. Table 6

![Figure 2](image-url)

**Figure 2.** Changes in the modifier gene frequency and the mean recombination value due to natural selection; diploidal system. Fitness, $W_{11} = 0.5, W_{12} = 0.5625, W_{13} = 0.5, W_{14} = 1, W_{23} = 0.375, W_{24} = 1, W_{24} = 0.4375, W_{33} = 0.3750, W_{34} = 0.3125, W_{44} = 0.375$. The solid line refers to the case of $r_0 = 0.1, r_1 = 0.05, r_2 = 0.0$ and the broken line to the case of $r_0 = 0.1, r_1 = r_2 = 0.0$. 

---

**Table 6**

*Selective advantages of the recombination-reducing gene in the initial stage of progress*

<table>
<thead>
<tr>
<th>Initial recombination value</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.00022</td>
<td>0.00039</td>
<td></td>
<td>0.0048</td>
</tr>
<tr>
<td>0.02</td>
<td>0.00036</td>
<td>0.00071</td>
<td>0.00256</td>
<td>0.0092</td>
</tr>
<tr>
<td>0.05</td>
<td>0.00056</td>
<td>0.00133</td>
<td>0.00491</td>
<td>0.0199</td>
</tr>
<tr>
<td>0.10</td>
<td>0.00065</td>
<td>0.00178</td>
<td>0.00466</td>
<td>0.0296</td>
</tr>
<tr>
<td>0.20</td>
<td>0.00068</td>
<td>0.00203</td>
<td>0.00665</td>
<td>0.0187</td>
</tr>
<tr>
<td>0.50</td>
<td>0.00070</td>
<td>0.00214</td>
<td>0.00200</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

a, b, c, and d refer to the fitness matrices in Table 5.
shows that this model is more favorable than model (b) in modifying the recombination value, although the intensity of selection (genetic load) is also higher in this case. Figure 2 shows the changes in \( x \) and \( \bar{r} \) in successive generations when \( x_0 = 0.01, r_0 = 0.1 \) and \( r_2 = 0.0 \).

The last model (d) represents a case of balanced lethals. As shown by Haldane (1962), the equilibrium \( \bar{r} \) and \( r_0 \) are all 0.25 if \( r > \frac{1}{4} \), and if \( r < \frac{1}{4} \) they are

\[
\bar{P}_1 = \bar{P}_4 = \bar{P}_2 = \bar{P}_3 = \frac{1}{2} [1 \pm (1 - 4r)^{\frac{1}{2}} (1 - r)]
\]

Thus, the initial selective advantage \( r_0 D(\bar{r}_A - \bar{r}_D) \) is 0 when \( r_0 > \frac{1}{4} \) and \( r_0 (1 - 2r_0)^2 (1 - r_0)^2 / 2(1 - r_0)^2 \) when \( r_0 < \frac{1}{4} \). Table 6 shows that the selective advantage of a modifier gene is very high in this case and thus this type of balanced lethals is expected quickly to reduce the recombination value between them if modifiers exist. A numerical computation has shown that the modifier gene frequency changes from 0.01 to 0.99 in about 190 generations when \( r_0 = 0.1, r_1 = 0.05 \), and \( r_2 = 0.0 \).

As in the case of haploid system the modification of linkage intensity is possible also when the \( A \) and \( B \) loci are going to fixation. One example is given in Table 7 where the fitness set represents an optimum model involving no stable equilibria, namely, \( W_{11} = W_{44} = 0.6, W_{12} = W_{14} = W_{24} = W_{34} = 0.9 \), and \( W_{14} = W_{22} = W_{23} = W_{33} = 1.0 \). The initial values of \( P_1, P_2, P_3, \) and \( P_4 \) employed were 0, 0.5001, 0.4999, and 0, respectively, with \( r_0 = 0.1, r_1 = 0.0, \) and \( r_2 = 0.0 \). It is seen from Table 7 that in this case the selection of a modifier gene is not so effective as with stable polymorphic loci.

DISCUSSION

Development of linkage: At present it is not known how and when the linkage of genes was developed in the evolutionary history of life. The fact that bacteria and viruses have genetic linkage suggests that it occurred in the very early stage.

**TABLE 7**

<table>
<thead>
<tr>
<th>Generation</th>
<th>( x )</th>
<th>( P_1 )</th>
<th>( P_2 )</th>
<th>( P_3 )</th>
<th>( P_4 )</th>
<th>( D )</th>
<th>( \bar{r} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.01000</td>
<td>.00000</td>
<td>.50010</td>
<td>.49990</td>
<td>.00000</td>
<td>-.25000</td>
<td>.09801</td>
</tr>
<tr>
<td>10</td>
<td>.01038</td>
<td>.10981</td>
<td>.39031</td>
<td>.39007</td>
<td>.10981</td>
<td>-.14177</td>
<td>.09794</td>
</tr>
<tr>
<td>50</td>
<td>.01142</td>
<td>.13383</td>
<td>.36656</td>
<td>.36578</td>
<td>.13383</td>
<td>-.11173</td>
<td>.09773</td>
</tr>
<tr>
<td>80</td>
<td>.01224</td>
<td>.13375</td>
<td>.36720</td>
<td>.36529</td>
<td>.13375</td>
<td>-.11180</td>
<td>.09757</td>
</tr>
<tr>
<td>100</td>
<td>.01282</td>
<td>.13367</td>
<td>.36808</td>
<td>.36458</td>
<td>.13367</td>
<td>-.11188</td>
<td>.09746</td>
</tr>
<tr>
<td>200</td>
<td>.01615</td>
<td>.13265</td>
<td>.40311</td>
<td>.33160</td>
<td>.13265</td>
<td>-.11174</td>
<td>.09680</td>
</tr>
<tr>
<td>400</td>
<td>.01945</td>
<td>.00002</td>
<td>.99996</td>
<td>.00000</td>
<td>.00002</td>
<td>-.00000</td>
<td>.09615</td>
</tr>
</tbody>
</table>

Fitness, \( W_{14} = W_{44} = 0.6, W_{12} = W_{14} = W_{24} = W_{34} = 0.9, \) and \( W_{14} = W_{22} = W_{23} = W_{33} = 1.0 \); recombination values, \( r_0 = 0.1, r_1 = r_2 = 0 \).
of organic evolution. The linkage of genes may occur in two different ways. One
is that a gene (or genes) is first duplicated, resulting in an end-to-end connection
of the duplicated and original genes (tandem duplication), and in course of time
the two genes are gradually differentiated owing to mutation so as to have differ-
ent biochemical or biological functions. In this case linkage is made first and the
recombination mechanism is developed later. In the second way, the duplicated
gene is not connected with the original gene from the beginning but exists inde-
pendently of the latter, and in course of time some type of epistasis or gene inter-
action arises and then the linkage of genes, which may occur by some mechanism,
becomes selectively advantageous and increases in frequency. In this case the
duplicated gene may not necessarily have originated as duplicated gene of its
partner, but may be an already differentiated gene derived from another cell by
means of an unequal recombination. Anyway, this case is represented by Model 1
in the section *Haploid system*, where \( r = \frac{1}{2} \).

However, the advantage of linkage in the early stage of organic evolution may
not always be due to epistasis, but linkage *per se* may be advantageous. For ex-
ample, linked genes might be physicochemically more stable than unlinked genes,
or the replication of DNA or RNA molecule as gene material might be easier
when genes are linked than when they exist independently. In the replicon hy-
pothesis of Jacob, Brenner and Cuzin (1963) the replication of DNA molecule
occurs as an unit of many genes (cistrons). Further, in the first model of the
development of linkage in the above paragraph, the duplicated gene pair might be
selectively advantageous over the single gene owing to increased formation of an
enzyme. In these cases the change in the frequency of linked gene pair \( AB \) or \( AA \)
(\( x \)) may be described by

\[
\frac{dx}{dn} = \frac{x(1-x)s}{1-s(1-x)}
\]

where \( s \) is the selective advantage of \( AB \) or \( AA \) over the unlinked gene pair \( AB \)
or single gene \( A \). In these cases, therefore, the development of linkage is fairly
rapid unless \( s \) is very small.

*Modification of linkage intensity:* In the treatment of modification of linkage
intensity in Model 2, we assumed the existence of modifier genes. As mentioned
previously, there are a large number of investigations which indicate that recom-
binations between gene loci are under genetic control. The recombination-modi-
fying genes so far identified presumably control the formation of some enzyme(s)
involved in the recombination event, and the physical distance between gene loci
would not be affected in this case. The positive interference between recombin-
ation events generally observed in higher organisms, or the lack of recombination
in male Drosophila perhaps reflects this type of recombination control. The fact
that environmental and genetic factors affect crossing over mainly in the centro-
mere region in Drosophila supports this concept (see Bodmer and Parsons
1962). The same mechanism might be involved in the influence of an inversion
chromosome on crossing over in other chromosomes as observed by Levine and
Levine (1954) in *Drosophila pseudoobscura*. Presumably, the formation of re-
combinational enzyme(s) is affected by inversion through a position effect.
There is, however, another type of recombination control, namely physical changes of gene arrangement of chromosome structure. Translocations and inversions are typical examples. In this case the changes of recombination values are more direct. However, chromosome rearrangements may sometimes destroy a gene complex or operon already established, resulting in a so-called position effect (cf. Lewis 1967), so that they are not always selectively advantageous. It seems that the real control of recombination values has been mediated both by selection of modifier genes and by chromosome rearrangements. If the degree of epistasis is the same, the modification of recombination values is more rapid in the latter process than in the former. However, even in the former process the modification of linkage intensity may proceed more rapidly than the modification of dominance as visualized by Fisher (1928).

So far, gene complexes, gene clusters, or operons appear to be more common in bacteria and bacteriophages than in fungi and other higher organisms. Horowitz (1965) and Lewis (1967) postulate that operons have often arisen by a process of repeated tandem duplications accompanied by gradual functional differentiation of the daughter genes, and operons for the primitive biosynthetic pathways of living organisms seem to have become fragmented in higher organisms. The tight linkage between the $\beta$ and $\delta$ loci of human hemoglobin (Boyer et al. 1963), for example, advocates this hypothesis for the origin of operons or gene complexes. If, however, operons or gene complexes are a unit of strong gene interaction as shown by Stahl and Murray (1966) in the T2 and T4 phages, the gradual tightening of linkage between structural genes would also have been possible, as shown in this paper. If organisms live in a relatively constant environment for a long time with little evolution, as bacteria and bacteriophages might have done, this possibility cannot be ruled out. In organisms which have evolved rapidly, recombinations would have been necessary for combining various advantageous mutations into the same individual, as indicated by Fisher (1930), Muller (1932), and Crow and Kimura (1965). Therefore, from this point of view too, the scarcity of operons or gene complexes in higher organisms is expected. Further, even if gene complexes had originated as tandem duplications, gene interactions would have played an important role in maintaining them in populations for a long time.

SUMMARY

Two different mathematical models for the modification of linkage intensity by natural selection are presented. In the first model the changes of recombination values are assumed to occur by structural changes of chromosome such as translocation and inversion, while in the second they are brought about by selection of modifier genes. In both models there must be epistasis or gene interaction for the modification of linkage intensity to proceed, the speed of the modification depending upon the degree of epistasis, and this modification always occurs in the direction of decreased recombination, if the fitnesses of genotypes remain constant. If the degree of epistasis is the same, the modification of linkage intensity is more rapid in the first model than in the second. In general, stable polymor-
phisms with epistasis are more favorable for establishing close linkage than unstable and transient polymorphisms with epistasis. Two different models for the evolution of linkage are proposed.

LITERATURE CITED


