ON TREATING THE CHROMOSOME AS THE UNIT OF SELECTION*

MONTGOMERY SLATKIN

Department of Theoretical Biology, University of Chicago, Chicago, Illinois 60637

Manuscript received July 19, 1971

ABSTRACT

A simple method is presented for approximating the behavior of a multiple genetic system under the action of selection and linkage. The effects of genetic drift and mutation are left out. It is found that the results are close to those obtained from a computer simulation of the same problem by FRANKLIN and LEWONTIN (1970). It is possible to describe the system in terms of a correlation length on the chromosome which measures the degree to which the different parts are bound by linkage. There is also a brief discussion about the higher order correlation coefficients as a measure of the interaction effects.

IN recent years there has been increasing theoretical interest in multiple-locus genetic systems in an effort to understand the structure of natural populations. The limitations of the single-locus theory are apparent, and the work of FRANKLIN and LEWONTIN (1970) and LEWONTIN (1964a,b) among others has shown that the higher order interactions between loci, caused by epistasis and linkage, are important and the single-locus and even the two-locus theory are inadequate predictors of the behavior of more complex systems. In order to understand the fundamental problems of the maintenance of genetic variability within populations and possible rates of evolution of populations, such factors must be considered. The model described here is of the simplest multiple-locus system including the effects of selection and linkage, and as such is only a first step in understanding more complicated problems. However, what is shown is that it is possible to reduce the degrees of freedom of a multiple-locus system in a consistent and useful way without losing its essential properties. In addition, the method of approximation itself leads to a better understanding of the behavior of such systems.

In a recent paper, FRANKLIN and LEWONTIN (1970) described the results of some computer simulations of the effect of selection and linkage on multiple-locus genetic systems. They presented evidence for the fact that when there is a large number of loci, many properties of the system depend on general characteristics of the selection and linkage rather than on the exact number of loci present. In particular, for their simulation of a completely symmetric heterotic model with multiplicative fitness, as the number of loci increased the average correlation between pairs of loci depended only on the total map length of the chromosome and on the fitness of the multiple homozygote. This result suggests that there might be an analytic description of this system which would lead to

* This research has been supported by AEC Contract AT(11-1)-1437.

the same conclusions and which would illustrate its underlying simplicity.

Before going into the details of the present model, I will describe some general properties of simple multiple-locus systems which will make it easier to interpret FRANKLIN and LEWONTIN's result. If the two possible alleles at each of N loci are denoted by 1 and 0, then a gamete is represented by an ordered N-tuple $(i_1, i_2, \ldots, i_N)$ where each of the i's have the value 1 or 0. The frequency of such a gamete is denoted by $g_{i_1 \ldots i_N}$. The frequency of the 1 allele at locus n, p_n, can be written as $\bar{i}_n$, the average value of i_n taken over all of the gametes. For example, with $N=2$,

$$\bar{i}_2 = p_2\, 0\cdot(g_{00}+g_{10}) + 1\cdot(g_{01}+g_{11}) = g_{01}+g_{11}$$

We can take advantage further of the numerical description of the gametes to write the gametic frequencies in terms of average values.

$$\begin{aligned} g_{11\ldots1} &= E(i_1 i_2 \ldots i_N) \\ g_{01\ldots1} &= E[(1-i_1) i_2 \ldots i_N] \end{aligned} \tag{1}$$

etc., where $E(\cdot)$ means the average value of the quantity taken over the population.

LEWONTIN and KOJIMA (1960) introduced the term "linkage disequilibrium" for the degree of non-random association between pairs of loci. They defined it as

$$D = g_{00}g_{11} - g_{10}g_{01} \tag{2}$$

Writing this in terms of average values, we get

$$D = E[(1-i_1)(1-i_2)]E(i_1 i_2) - E[i_1(1-i_2)]E[i_2(1-i_1)] \tag{3}$$

and by expanding and combining terms, the result is

$$D = E(i_1 i_2) - \bar{i}_1\bar{i}_2 = E[(i_1-\bar{i}_1)(i_2-\bar{i}_2)]$$

Therefore D is the covariance and related to the correlation coefficient of i_1 and i_2 called ρ by FRANKLIN and LEWONTIN and defined as

$$\rho = E[(i_1-\bar{i}_1)(i_2-\bar{i}_2)]/[E([i_1-\bar{i}_1]^2)\cdot E([i_2-\bar{i}_2]^2)]^{1/2} \tag{5}$$

In systems with more loci, higher order interactions become important and we would like a description of the size of those effects. The obvious generalization of D to more loci is

$$\begin{aligned} D_{123} &= E[(i_1-\bar{i}_1)(i_2-\bar{i}_2)(i_3-\bar{i}_3)] \\ D_{1234} &= E[(i_1-\bar{i}_1)(i_2-\bar{i}_2)(i_3-\bar{i}_3)(i_4-\bar{i}_4)] \end{aligned} \tag{6}$$

etc., as the higher order disequilibrium constants. The D's are the same as the correlation functions (L_n) introduced by BENNETT (1954). With this definition the gametic frequencies can be written in terms of the disequilibrium constants and the allelic frequencies by expanding (1). For example,

$$\begin{aligned} g_{11\ldots1} = {} & \bar{i}_1\bar{i}_2 \ldots \bar{i}_N + D_{12}\bar{i}_3 \ldots \bar{i}_N + D_{13}\bar{i}_2\bar{i}_4 \ldots \bar{i}_n + \ldots \\ & + D_{123}\bar{i}_4 \ldots \bar{i}_N + D_{124}\bar{i}_3\bar{i}_5 \ldots \bar{i}_N + \ldots \\ & \quad \vdots \\ & + D_{123\ldots N} \end{aligned}$$

$$g_{01\ldots1} = (1-\bar{\imath}_1)\bar{\imath}_2\ldots\bar{\imath}_N - D_{12}\bar{\imath}_3\ldots\bar{\imath}_N + \ldots \quad (7)$$
$$- D_{123}\bar{\imath}_4\ldots\bar{\imath}_N - \ldots$$
$$\vdots$$
$$- D_{123\ldots N}$$

BENNETT (1954) obtained expressions similar to (7).

In the completely symmetric, heterotic system, one of the equilibrium states is one for which all the allelic frequencies are ½. KARLIN and FELDMAN (1970) have shown that, for the two-locus system, there may be other equilibrium points but for the completely symmetric case ($\gamma=\beta$, $\delta=\alpha$ in their notation), none of the other equilibrium points is stable. This suggests but certainly does not prove that the same result is true for a larger number of loci. Here we will be concerned only with the symmetric equilibrium ($\bar{\imath}_1 = \bar{\imath}_2 = \ldots = ½$) which FRANKLIN and LEWONTIN's simulations indicate is stable. Also, because of the symmetry, the gametes appear in complementary pairs and the frequency of a gamete is the same as the gamete obtained by changing all the 1's to 0's and vice versa (e.g., $g_{10101}=g_{01010}$). In the computation of the D's with an odd number of subscripts, the contribution from a gametic type will be the negative of the contribution from the complementary gamete. Since, in the completely symmetric case, the complementary types occur in equal frequencies, all D's with an odd number of subscripts are 0. Because the allelic frequencies are ½ and the effect of the alleles are the same, there are 2^N equivalent states of the population which can be obtained by relabeling the alleles at the different loci. Therefore, we can choose any one of them to analyze. When there is perfect linkage, two of the gametes will have frequency ½ and the rest 0. This can be any complementary pair but for convenience let us call them $(1,\ldots,1)$ and $(0,\ldots,0)$. All of the nth order disequilibrium constants have the value $½^n$. We can compute $g_{1\ldots1}$ using (7) to get

$$g_{11\ldots1} = \frac{1}{2^N}\left(1+\binom{N}{2}+\binom{N}{4}+\ldots+\binom{N}{N}\right) = \frac{1}{2} \quad (8)$$

and we see that the major contribution to this sum comes not from the pairwise disequilibrium constants but those of order $N/2$, because there are more of them. In other words the higher order interactions dominate and a consideration of two-locus effects would lead to an incorrect result. This suggests that the higher order disequilibrium constants will be important when there is some recombination introduced.

The computations of LEWONTIN (1964a) and FRANKLIN and LEWONTIN (1970) indicate that when permanent linkage disequilibrium is maintained one complementary pair of gametes is most frequent. It is this observation that leads to the present method of analysis. The tables of gametic frequencies presented by FRANKLIN and LEWONTIN (1970) and LEWONTIN (1964a) show that, for these simulations, the populations consist almost entirely of the most common gametic pair and those gametes which can be obtained by a single crossing over. If the

most common type is $(1,\ldots,1)$ and its complement $(0,\ldots,0)$, then the next most common type has the form $(1,\ldots,1,0,\ldots,0)$ and its complement. For convenience we will call the first type A and the second B(i) where the change from 1 to 0 occurs between the ith and $(i+1)$st locus. It seems possible that a careful description of the effect of selection and linkage on the frequencies of these two gametic types would be sufficient to model much of the behavior of this system.

The approximation described above has to be carried out somewhat indirectly. We can write the probability of a change from a 1 to a 0 between locus i and $i+1$ as we read along the gamete as a_i. In other words, if a gamete is chosen at random, then a_i is the probability that $i_{i+1}=0$ given that $i_i=1$. In this symmetric model a_i is also the probability of a change from 0 to 1 but this need not be true in general. The basic assumption is that the a_i are independent of the state of the rest of the chromosome. This is not satisfied exactly but, as will be discussed below, it is equivalent to ignoring the gametes other than the A or B types. Since these other gametic types are relatively infrequent, except near linkage equilibrium (when all the g's are nearly $1/2^N$) this approximation may be useful. Once this assumption is made, the gametic frequencies can be written in terms of the a_i.

$$g_A = \frac{1}{2}\prod_{i=1}^{N-1}(1-a_i) \tag{9}$$

$$g_{B(i)} = \frac{1}{2}(1-a_i)\ldots(1-a_{i-1})a_i\ldots(1-a_N) = g_A\frac{a_i}{1-a_i}$$

For simplicity, we assume that a generation consists of two distinct parts: (1) mating and the production of offspring, and (2) selection on the resulting offspring. In order to calculate how the a_i change during one generation we calculate the change for each of the two parts independently. In order to calculate the change in the a_i under recombination, we must compute D for adjacent pairs of loci. Since $D_{i,i+1}$ does not depend on the relative frequencies at any but the ith and $(i+1)$st loci, it is sufficient to look at the two locus case. By definition

$$D_{12} = g_{11}g_{00} - g_{10}g_{01}$$

and in the completely symmetric case $g_{11}=g_{00}$ and $g_{10}=g_{01}$, as pointed out above, so

$$D_{12} = \frac{1}{2}(g_{11}-g_{10}) \tag{10}$$

$$= \frac{1}{2}\left[\frac{1}{2}(1-a_i) - \frac{1}{2}a_i\right] = \frac{1}{4}(1-2a_i)$$

or, in general

$$D_{i,i+1} = \frac{1}{4}(1-2a_i). \tag{11}$$

For future reference, we can calculate the D_{ij} for all pairs of loci. Proceeding in a similar way for three loci.

$$D_{13} = (g_{111}+g_{101})(g_{010}+g_{000}) - (g_{011}+g_{001})(g_{110}+g_{100})$$
$$= \frac{1}{2}(g_{111}+g_{101}-g_{011}-g_{001}) \tag{12}$$
$$= \frac{1}{4}(1-2a_1)(1-2a_2)$$

By induction, we have

$$D_{ij} = \frac{1}{4}(1-2a_i)(1-2a_{i+1}) \ldots (1-2a_{j-1}), \quad i<j. \tag{13}$$

GEIRINGER (1944) has shown that in the two-locus system D changes under recombination according to

$$D \rightarrow D(1-R) \tag{14}$$

If we assume that the probability of crossing over between any adjacent pair of loci is independent of the rest of the chromosome, then we can use this result to compute this component of the change in the a_i

$$(1-2a_i) \rightarrow (1-2a_i)(1-R_i)$$

or

$$a_i \rightarrow a_i+R_i(1-2a_i)/2 \tag{15}$$

where R_i is the recombination fraction between the ith and $(i+1)$st loci. We know that the assumption of independence of crossing over is not strictly correct but we will use it as a first approximation and in order to compare the results obtained here with those of FRANKLIN and LEWONTIN's simulations, in which the same assumption was made.

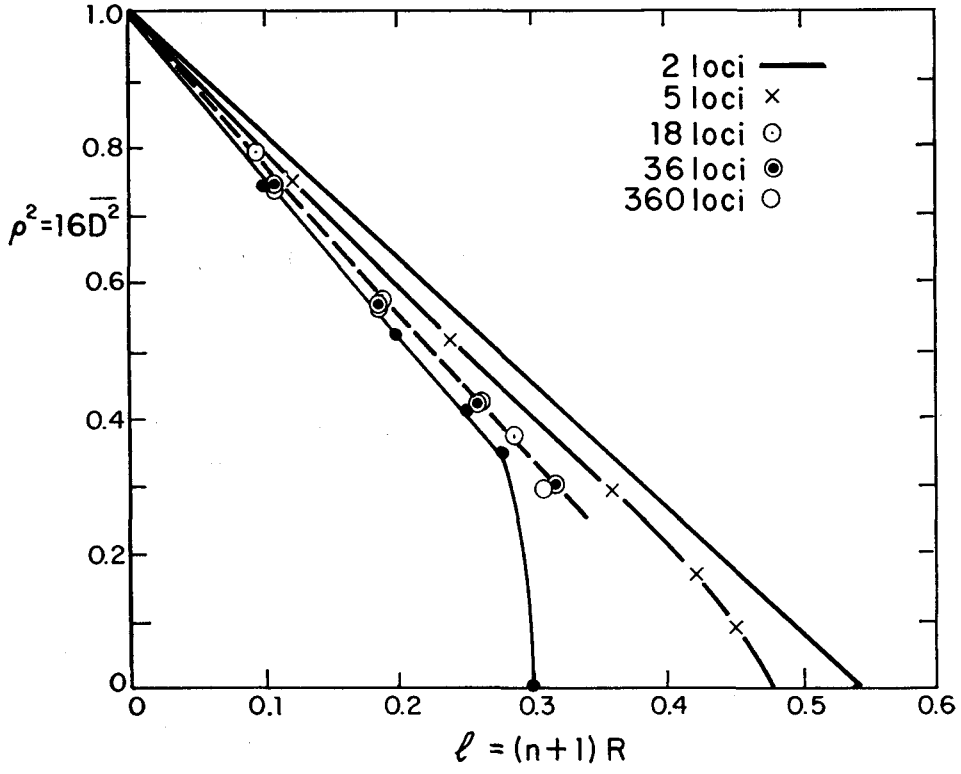
Next we must calculate the change in the a_i under selection. In the completely symmetric system, the fitness of a diploid individual depends only on the number of homozygous (or heterozygous) loci. We can write the fitness of an individual with i homozygous loci as W_i. From the two-locus calculation of KARLIN and FELDMAN, we expect that the equilibrium of such a system is stable only if the selection is heterotic so that the cases of interest are those for which W_i is monotonically decreasing from the fitness of the multiple heterozygote. We can arbitrarily set W_0, the relative fitness of the complete heterozygote, to be 1.

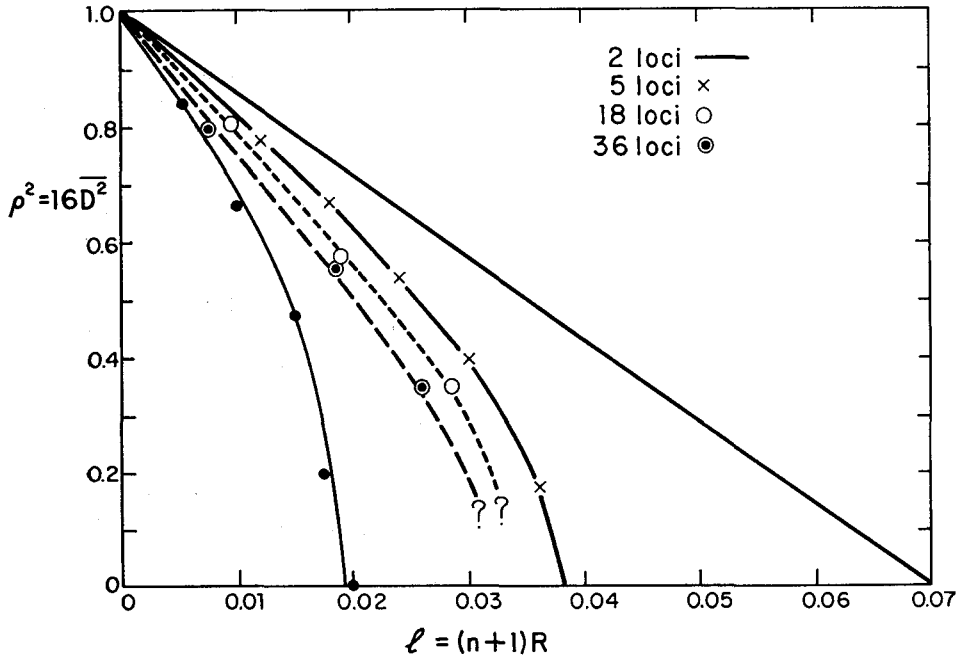
To calculate the change in a_i under selection we use the fact that the change in the ratio of gametic frequencies is determined by the ratio of the marginal fitness of the two gametes. If we let F denote the marginal fitness, which is the average fitness of a gamete taken over the population, then

$$\frac{g_A}{g_{B(i)}} \rightarrow \frac{F_A}{F_{B(i)}} \frac{g_A}{g_{B(i)}} \tag{16}$$

after selection. Formula (16) is exact in an infinite, randomly mating population and does not depend on the approximations used in this model. From (9)

$$\frac{g_A}{g_{B(i)}} = \frac{1-a_i}{a_i} \tag{17}$$


1.0
0.8
0.6
0.4
0.2
0
$\rho^2=16\overline{D^2}$
2 loci
5 loci
18 loci
36 loci
360 loci
0
0.1
0.2
0.3
0.4
0.5
0.6
$\ell = (n+1)\,R$


1.0
0.8
0.6
0.4
0.2
0
$\rho^2=16\overline{D^2}$
2 loci
5 loci
18 loci
36 loci
? ?
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
$\ell = (n+1)R$

and it remains to describe the F's in terms of the a_i. If we define g_C as the frequency of the remaining gametic types, we can write

$$F_A = (W_0+W_N)g_A + \sum_{i=1}^{N-1} g_{B(i)}\ (W_i+W_{N-i}) + g_C\,\tilde{W} \tag{18}$$

and

$$F_{B(i)} = (W_i+W_{N-i})g_A + \sum_{j=1}^{N-1} g_{B(j)}\ (W_{|i-j|} + W_{N-|i-j|}) + g_C\,\tilde{W} \tag{19}$$

where the effects of the crosses with all the other gametic types are included in the last term.

At this point we can see the relationship between the assumption about the independence of the a_i and the assumption that all of the gametic types other than the A and B types are unimportant. There are nearly enough of the a_i, $N-1$, to exactly describe the frequencies of the N pairs of gametes of the A and B types. Once the a_i are known and g_A and $g_{B(i)}$ are calculated, the approximate frequencies of the other types can be computed directly. These will not be exact but if the sum of their frequencies is small compared to g_A and $\sum_{i=1}^{n-1} g_{B(i)}$, then the error made is unimportant and for our purposes the assumption of independ- of the a_i is adequate. We approximate further on this basis by including the effects of the crosses with all other gametic types in a single term and assume that it is the same for the A and B types. The last term on the right hand sides of (18) and (19) is the total frequency of the remaining types, g_C, multiplied by an estimate of their average effect, $\tilde{W}$. Since all terms in an exact computation of the marginal fitness would be of the form $g \ldots (W_j+W_{N-j})$, $\tilde{W}$ should be nearly the average value of W_i. However, gametes with a more equal distribution of 1's and 0's will be more frequent, so a better value to use is a "triangular" average.

$$\tilde{W} = \sum_{i=1}^{N} \sum_{j=i}^{N-i} W_j \Big/ N(N+1)/2 \tag{20}$$

In the practice the results are relatively insensitive to the exact choice for $\tilde{W}$. This supports the hypothesis that an exact treatment of the other gametes is not essential.

We are now in a position to compute the equilibrium values for a_i for different forms of the selection function and different map lengths. Despite the simplifying assumption which has been made, an analytic solution is still impossible in non-trivial cases. To test the validity of this approach, the results from this model are compared with the results from FRANKLIN and LEWONTIN's paper (Figure 1). In these examples, a multiplicative fitness function is used

$$W_i = (1-s)^i = e^{[i\ ln(K/N)]}\ K = (1-s)^N \tag{21}$$

FIGURE 1.—A comparison of the results from the approximate calculation with those from the simulation by FRANKLIN and LEWONTIN (1970). In each case the lowest curve is the result of the present calculation. (a) $K = .0225$; (b) .4832. (Redrawn from FRANKLIN and LEWONTIN [1970] with permission of the authors.)

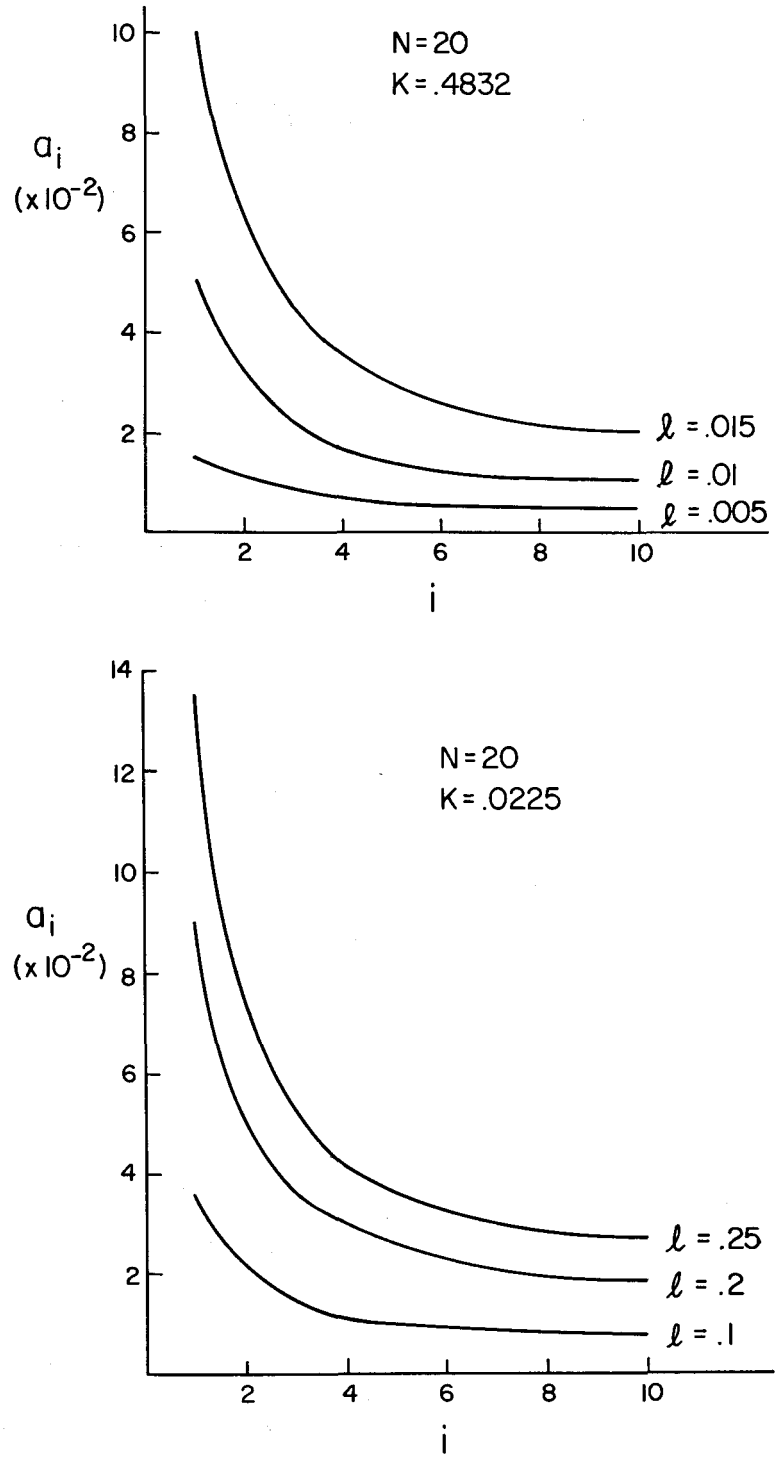


FIGURE 2.—A plot of a_i vs. i for different map lengths. By symmetry $a_i = a_{N-i}$

and two choices for K are shown. The computations were carried by choosing initial values for the a_i and directly iterating equations (15) and (16) in turn, using (18) and (19) to calculate the marginal fitness, until an equilibrium was reached. The calculation of $\bar{\rho}$ was done directly by averaging over the D_{ij} for all pairs of i and j and using the fact that in this case $p = 4D$. The computations were made for N=20 but selected cases were tested with more loci (N=40 and N=60) with no significant change in the curve. In several cases, when different initial values of the a_i were used for the same choice for the selection function and total map length, the same equilibrium values were obtained. The results from this model are consistently lower than those from the simulations. It is not clear at this time how much difference is due to drift induced by the finite population size in the simulation and how much is due to the inadequacy of the approximation.

Figure 2 shows a plot of the a_i as a function of i for different values of the parameters. In all cases there is a significant "edge effect" where the loci at the two ends are much less tightly bound then those in the middle. This result is not surprising and is consistent with the gametic frequencies given by FRANKLIN and LEWONTIN (1970). It is clear that a "circular" chromosome of the same total map length and with the same selection applied to it would be more tightly linked than a linear chromosome.

An equivalent although not necessarily better formulation of the system can be made in terms of continuous rather than discrete variables. A gamete is described by fields of 0's and 1's of different lengths and at different locations. If x is the length variable on the chromosome, then the probability of a change from 0 to 1 or 1 to 0 in $(x,x+dx)$ is $a(x)dx$. Under recombination, $a(x)$ goes into $a(x)+R(x)$ where $R(x)$ is the probability of recombination in the same region. In the limit as dx approaches 0, the effect of reverse recombination is negligible, as can be seen by proceeding directly from (15) replacing a_i by $a(x)dx$. By direct analogy with the discrete model, we can calculate the change in $a(x)$ under selection. After selection we have

$$a(x) \to F_{B(x)}\, a(x)/F_A \tag{22}$$

where

$$F_A = [W(0)+W(1)]g_A + \int_0^1 [W(x)+W(1-x)]g_{B(x)} + g_C \widetilde{W} \tag{23}$$

$$F_{B(x)} = [W(x)+W(1-x)]g_A + \int_0^1 [W(|x-y|)+W(1-|x-y|)]g_{B(y)} + g_C \widetilde{W}$$

and where

$$g_A = e^{-\int_0^1 a(x)dx}, \quad g_{B(x)} = a(x)g_A dx \tag{24}$$

$W(x)$ is the fitness of an individual with a fraction x of the chromosome pair homozygous. Although there is no analytic solution to this set of equations, there is a simple interpretation which makes it a useful description. We can calculate

D between any two locations x and y on the chromosome directly from (13) by replacing a_i by $a(x)dx$.

$$D(x,y) = \frac{1}{4}\Pi(1-2a(x))dx$$

$$\cong \frac{1}{4}e^{-\Sigma 2a(x)dx} = \frac{1}{4}e^{-2\int_x^y a(x)dx}$$

If we define the average of $a(x)$ between x and y as

$$\bar{a}(x,y) = \frac{2}{y-x}\int_x^y a(x)dx \tag{25}$$

then we can write

$$D(x,y) = e^{-\bar{a}(x,y)(y-x)} \tag{26}$$

Since $a(x)$ has units 1/length (it is the probability of a change from 0 to 1 or 1 to 0 per unit length on the gamete), $\bar{a}$ has the same units. Therefore $1/\bar{a}$ is a length which we can interpret as the "correlation length" on the gamete. If we define $l_c(x,y) = 1/\bar{a}(x,y)$, then pairs of loci which are closer than l_c are highly correlated ($D \sim 1/4$) while loci which are further apart are uncorrelated ($D \sim 0$). It is possible for l_c to be greater than the length of the chromosome, in which case we could say that the whole system is tightly bound and acting nearly as a unit. At least in the case of multiplicative fitness, selection acts to increase l_c while recombination acts to decrease it. Equilibrium is reached when the amount of increase is balanced by the amount of decrease in each generation. Figure 3 shows a plot of $\bar{l}_c$ as a function of map length where $\bar{l}_c$ is the average of $l_c(x,y)$.

There are some properties of the system which can be derived from this model without detailed calculations. In order for there to be a balance between recombination and selection, selection must act to decrease $a(x)$ (or the a_i), since recombination always increases $a(x)$. For this to occur $F_{B(x)} < F_A$ for all x. In other words, the marginal fitness of the most common gamete must be greater than the marginal fitness of all of the less common types. Since the assignment of the 1's

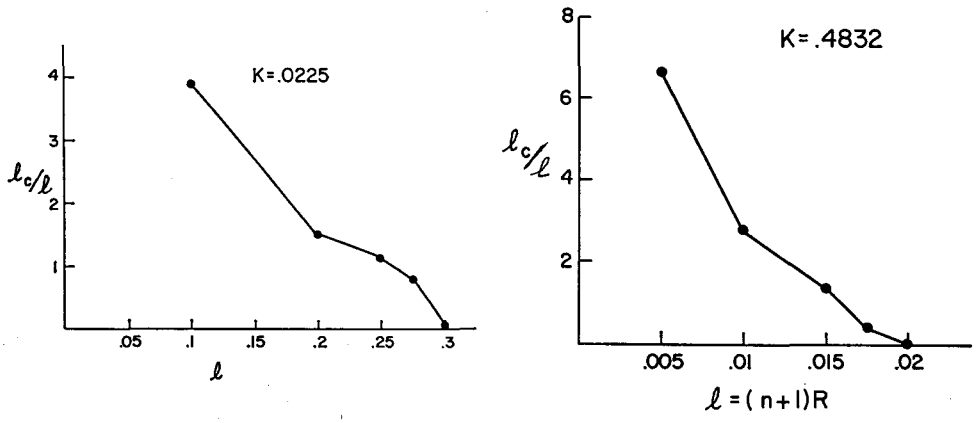


FIGURE 3.—A plot of $\bar{l}_c$ vs. l for the two strengths of selection.

and 0's is arbitrary, this means that the A's have the advantage only because they are most frequent. Therefore the gain in fitness from the crosses between the most common types must be greater than the gain in fitness from the crosses of the B types with the A's. To illustrate this, let us consider three possible forms for $W(x)$. If $W(x)$ is linear in x, $W(x) = 1-cx$ (or $W_i = 1-ic/N$ in the discrete model), then the average fitness of a cross of a gamete with a complementary pair is always the same because

$$\frac{1}{2}[W(x)+W(1-x)] = 1 - c/2 \tag{27}$$

Since all contributions to the marginal fitness are of the above form, $F_A = F_{B(x)}$ for all x and the most common types do not have the advantage under selection. In this case permanent linkage disequilibrium is not possible as has been shown for the two locus model by LEWONTIN and KOJIMA (1960).

For a fitness function which is convex upward (curve (2) in Figure 4), the average fitness of the crosses betweeen the A types $[W(0)+W(1)]/2$ is always less than the average of a cross of a B type with a pair of A's $[W(x)+W(1-x)]/2$ so that $F_A < F_{B(x)}$ for all x. Therefore $a(x)$ is increased by selection and permanent linkage disequilibrium cannot be maintained. It is only in the case where $W(x)$ is concave upward (curve (3) in Figure 4) that the proper condition is satisfied. The average fitnesses of the crosses between the two A types $[W(0)+$

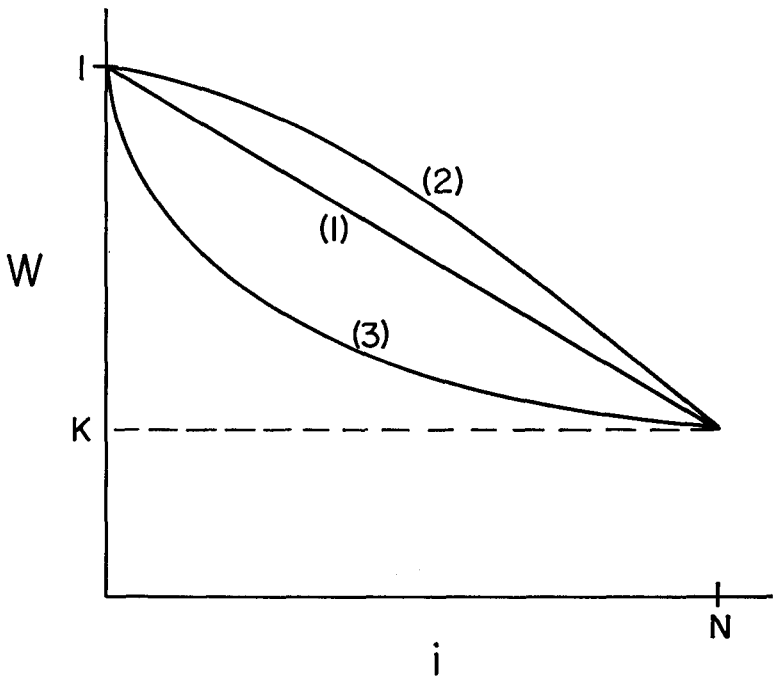


FIGURE 4.—A graph of the three types of heterotic selection functions for a completely symmetric model.

$W(1)]/2$ is greater than that for any other crosses $[W(x)+W(1-x)]/2$ so that the most common type has the advantage and can be maintained by selection. This above argument does not depend on the particular representation of the population or on the assumptions here, but only on the fact that at equilibrium the gametes come in complementary pairs in this special symmetric model.

In conclusion, the basic properties of the symmetric multiple-locus system can be described rather simply in terms of higher order disequilibrium constants or the correlation lengths, $l_c(x,y)$. In either description, it is clear that when permanent linkage disequilibrium is maintained in a population, the higher order interactions are important and the chromosome tends to act as a unit. The degree to which this is true in any given system is a measure of whether the gene or the chromosome is the unit of selection, or, more accurately, what parts of the genome can be said to be acting in unison. Further work on this problem will have to be directed at the effect of relaxing some of the symmetry conditions imposed on the present model. The equality of frequency of complementary gametes depends on the equality of effect of the two alleles. When this condition is changed, some of the simplicity of the present description is lost but it is possible that by introducing a different probability of a change from a 1 to a 0 on a gamete (say b_i), that the same general technique could be used. The calculations of LEWONTIN (1964a) indicate that such a system is not fundamentally different from the completely symmetric one. When the loci have unequal effects but the two alleles at each locus are equivalent, then the complementary gametes have the same frequency but the computation of the marginal fitness is much more complicated. Still more general systems will probably require a more detailed description than can be provided by this technique.

I wish to thank RICHARD LEWONTIN of the University of Chicago for help and encouragement at every stage of this work and CURTIS STROBECK and JESSE KRAKAUER for critically reading the manuscript.

LITERATURE CITED

BENNETT, J. H., 1954 On the theory of random mating. Ann. Eugenics **18**: 311–317.

FRANKLIN, I. and R. C. LEWONTIN, 1970 Is the gene the unit of selection? Genetics **65**: 707–734.

GEIRINGER, H., 1944. On the probability theory of linkage in Mendelian heredity. Ann. Math. Statis. **15**: 25–50.

LEWONTIN, R. C., 1964a The interaction of selection and linkage. I. General considerations and heterotic models. Genetics **49**: 49–67. ——, 1964b The interaction of selection and linkage. II. Optimum models. Genetics **50**: 595–609.

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

KARLIN, S. and M. W. FELDMAN, 1970 Linkage and selection: Two-locus symmetric viability model. Theoret. Pop. Biol. **1**: 39–71.