THE INTERACTION OF SELECTION AND LINKAGE.
I. GENERAL CONSIDERATIONS; HETEROTIC MODELS[1]

R. C. LEWONTIN

Department of Biology, University of Rochester, Rochester, N.Y.

WHILE the theory of the genetic changes in a population due to selection is quite well understood for single loci, our theory for multiple-gene characters is in a rudimentary stage. Most of the formulations for multiple-gene characters are simply extensions of single-locus models, extensions which ignore the problem of linkage. There are, however, a few papers in which the role of linkage has been investigated for more or less special cases of selection (KIMURA 1956; LEWONTIN and KOJIMA 1960; BODMER and PARSONS 1962). The results of these investigations were sufficient to show that even for relatively simple cases (two loci, simple symmetrical selective values) linkage might have profound effects on the course of natural selection and, *pari passu*, natural selection may have major effects on the distribution of coupling and repulsion linkage in a population.

The results of the investigations of LEWONTIN and KOJIMA (1960) of the two-locus model can be summarized as follows: (1) If the fitnesses are additive between loci (no epistasis), linkage does not effect the final equilibrium state of the population. (2) If linkage is tighter than the value demanded by the magnitude of the epistasis (the greater the epistasis the greater the value) there may be permanent linkage disequilibrium and alteration of equilibrium gene frequencies. (3) The rate of genetic change with time is affected by the tightness of the linkage. (4) In some cases stable gene frequency equilibria are possible only if linkage is tight enough.

Although these conclusions were based only on two-locus model and for selective values of a fairly restricted sort, they point clearly to the importance of taking linkage into account in understanding the changes of gene frequencies in populations. In fact, some experimental results (an example of which will be given below) can be understood only if the interaction of selection and linkage is taken into account.

The equations describing the interaction between selection and linkage (see below) do not usually have general literal solutions. It is for this reason that the authors cited above have restricted themselves to relatively simple cases. In view of the interesting findings of those previous papers, however, it is worthwhile to explore the subject more intensively. To do so requires the numerical rather than general literal solutions to the equations, but such numerical solutions apply, obviously, only to the particular parameter values chosen. To make such a nu-

merical approach at all useful, it is necessary to cover a variety of models of selection and to vary each model so that an empirical "feel" for general results can be obtained. In this sense, numerical calculations are like experiments: the generality of the results depends upon the variety of conditions of the experiments.

In this and the succeeding two papers of this series, three main types of selection are discussed. While these are not completely exhaustive of all possibilities, they represent the main modes of selection in natural and artificial populations. In this paper I will consider heterotic models, in which heterozygotes at each locus are more fit than homozygotes. In the second paper of the series optimum selection will be examined; that is, selection operating against individuals whose phenotypes deviate from some intermediate optimum. The last paper will deal with unidirectional selection in which an extreme phenotype or genotype is selected against. Since the effect of linkage is rather different in these three cases, separate discussions of each are required.

THE MATHEMATICS OF SELECTION AND LINKAGE

A general treatment of the equations of gene frequency change with linkage for the two locus case is given by KIMURA (1956) for the continuous time model and by LEWONTIN and KOJIMA (1960) for the discrete generation case. The results of these latter authors will be briefly recapitulated here and extended to multiple loci. Let there be two loci with two alleles each denoted by A,a and B,b respectively. There are then four gametic types ab, aB, Ab, AB and these will have the frequencies just after meiosis in any generation g_{00}, g_{01}, g_{10} and g_{11}, respectively. In these subscripts a 0 denotes the lower case letter allele (a or b) and a 1 denotes the upper case allele (A or B). To simplify notation let these four frequencies be, $x_0, x_1, x_2,$ and x_3. The subscripts of the x's are the decimal equivalents of the binary subscripts of the g's. That is, 00 is binary 0, 01 is binary 1, 10 is binary 2, and 11 is binary 3.

Further let
$Z_{ij} =$ the frequency of the zygote formed from the gametes whose frequencies are x_i and x_j
$W_{ij} =$ fitness of genotype whose frequency is Z_{ij}
$W_{i.} = \sum_{ij} W_{ij} x_j$
$\overline{W} = \sum_i W_{i.} x_i$ (the mean fitness)
$R =$ recombination fraction between the loci
$D = x_0 x_3 - x_1 x_2$ (the linkage disequilibrium determinant)

Then, LEWONTIN and KOJIMA have shown that the change in genetic frequency in one generation, Δx_i, is given by

$$(1) \quad \Delta x_i = \frac{x_i(W_{i.} - \overline{W}) - (-1)^i R D W_{12}}{\overline{W}}$$

At gene frequency equilibrium $\Delta x_i = 0$ for all i and this will happen when one of two conditions holds. First if $D = 0$ there is no linkage disequilibrium and for Δx_i to be zero

(2a) $W_{i.} - \overline{W} = 0$

or (2b) $x_i = 0$

for all i.

The second possibility is that $D \neq 0$, that there is permanent linkage disequilibrium in which case when $\Delta x_i = 0$.

(3) $x_i(W_{i.} - \overline{W}) - (-1)^i R D W_{11} = 0$

In this latter case there is a balance between the loss or gain of a gametic type by selection and the gain or loss of that type by recombination.

These relationships can be generalized to more than two loci fairly easily although the resulting equations are rather cumbersome. Again let x_i be the frequency of the gametic type g_i where the subscript of the x's is the decimal equivalent of the binary g subscripts. Thus, for five loci, $g_{00000}, g_{00001}, \ldots g_{11111}$ have the frequencies $x_0, x_1 \ldots x_{31}$. The $n-$ locus generalization of equation (1) has the form

(4) $\Delta x_i = [x_i(W_{i.} - \overline{W}) - \rho(x, R)]/\overline{W}$

where $\rho(x, r)$, the recombination function, has the following complicated and unfortunate form:

$$(5) \ \rho(x, R) = \sum_{i=1}^{\frac{n(n-1)}{2}} \left[R_i \sum_{j,k,l,m \varepsilon R} \frac{(x_j x_k - x_l x_m) W_{jk}}{2^{H-2}} \right]$$

$$+ \sum_{i=\frac{n(n-1)}{2}+1}^{2^{n-1}-1} \left[O_i(-1)^s \sum_{j,k,l,m \varepsilon O} \frac{x_j x_k W_{jk}}{2^{H-2}} \right]$$

with $n =$ number of loci.

$R_i =$ one of the $[n(n-1)]/2$ recombination fractions between two loci among n.

$W_{jk} =$ fitness of a zygote formed from the gametic combination x_j, x_k, ($W_{jk} = W_{lm}$)

$H =$ number of heterozygous loci in the zygote jk

R is the subset of all possible pairs $x_j x_k, x_l x_m$ with the following characteristics:

(a) $x_j x_k$ and $x_l x_m$ must each be capable by some recombination event (including no recombination) of producing the gamete x_i.

(b) $x_j x_k$ and $x_l x_m$ must both be heterozygous for the two loci corresponding to the R_i.

(c) at the two loci in question $x_j x_k$ must be in the same linkage phase as the gamete to be produced, x_i, while $x_l x_m$ must be in the opposite linkage phase.

$O_i =$ special recombination fractions which are not conventional recombination distances between pairs of genes.

$S =$ total number of exchanges for a given O_i (see below).

These special O_i arise in the following way. When n loci are heterozygous, the gametic output of a given heterozygote is completely specified by the probabilities

of exchange in the $n - 1$ intervals between the genes. If we do not assume independence of the intervals, that is, if we allow interference, there are obviously 2^{n-1} parameters made up of the noncrossovers, and single, double, triple, etc. exchanges. These parameters (probabilities) add to unity however, so there are only $2^{n-1} - 1$ independent parameters. For five loci, for example there are 15 parameters consisting of the four single exchange probabilities (S_1, S_2, S_3, S_4), the six double exchange probabilities $(D_{12}, D_{13}, D_{14}, D_{23}, D_{24}, D_{34})$, four triple exchanges $(T_{123}, T_{124}, T_{234})$ and one quadruple exchange (Q_{1234}). On the other hand there are only $n(n - 1)/2$ conventional recombination fractions among n genes taken two at a time. For five loci there are ten such recombination fractions. Thus, any formulation of the results of crossing over cannot be put only in terms of the usual recombination fractions for $n > 3$. The O_i referred to in equation (5) are extra orthogonal recombination values necessary to make up the full set of $2^{n-1}-1$ recombination values which are linear combinations of the $2^{n-1}-1$ exchange probabilities. Table 1 shows the relationships between the R_i and O_i and the various exchange probabilities for the five locus case. Figure 1 shows the definitions of the ten conventional recombination fractions, R_i.

The 15 equations in Table 1 may be solved for the exchange probabilities with the result shown in Table 2. The Table shows the sign ($+$ or $-$) associated with each R_i and O_i in the linear combination

$$(6) \quad \text{Exchange probability} = \frac{1}{8} \sum_{i=1}^{10} \pm R_i + \frac{1}{8} \sum_{i=1}^{5} \pm O_i.$$
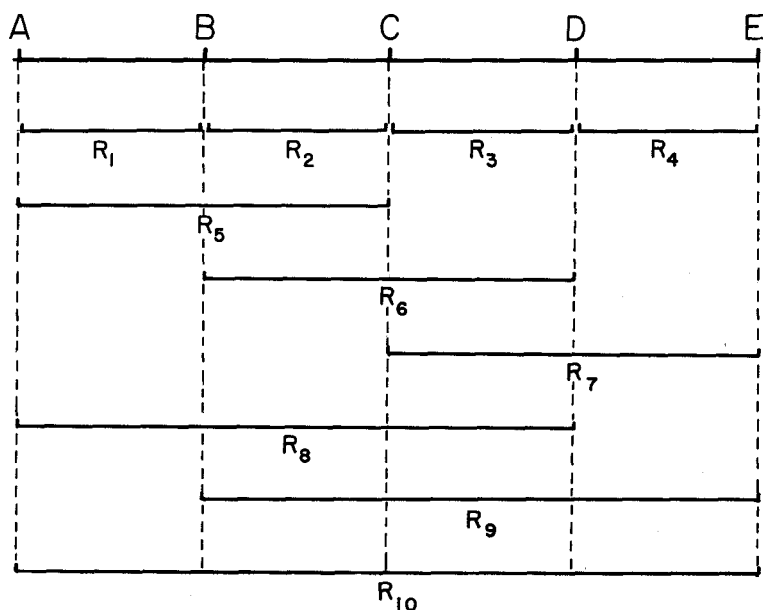

FIGURE 1.—Definition of the ten recombination fractions R_1–R_{10} among five genes A–E as used in the text.

TABLE 1

Relationships between the recombination factions r_i *and* O_i *and the probabilities of single, double, triple and quadruple exchanges for five loci*

$$r_1 = (S_1 + D_{12} + D_{13} + D_{14} + T_{123} + T_{124} + T_{134} + Q)$$
$$r_2 = (S_2 + D_{23} + D_{24} + D_{12} + T_{123} + T_{234} + T_{124} + Q)$$
$$r_3 = (S_3 + D_{23} + D_{13} + D_{34} + T_{123} + T_{134} + T_{234} + Q)$$
$$r_4 = (S_4 + D_{14} + D_{24} + D_{34} + T_{234} + T_{124} + T_{134} + Q)$$

$$r_5 = (S_1 + S_2 + D_{23} + D_{13} + D_{14} + D_{24} + T_{234} + T_{134})$$
$$r_6 = (S_2 + S_3 + D_{24} + D_{12} + D_{13} + D_{34} + T_{124} + T_{134})$$
$$r_7 = (S_3 + S_4 + D_{13} + D_{14} + D_{23} + D_{24} + T_{124} + T_{123})$$
$$O_1 = (S_1 + S_3 + D_{12} + D_{14} + D_{23} + D_{34} + T_{124} + T_{234})$$
$$O_2 = (S_2 + S_4 + D_{12} + D_{23} + D_{14} + D_{34} + T_{123} + T_{134})$$
$$O_3 = (S_1 + S_4 + D_{12} + D_{13} + D_{24} + D_{34} + T_{123} + T_{234})$$

$$r_8 = (S_1 + S_2 + S_3 + D_{14} + D_{24} + D_{34} + T_{123} + Q)$$
$$r_9 = (S_2 + S_3 + S_4 + D_{12} + D_{13} + D_{14} + T_{234} + Q)$$
$$O_4 = (S_1 + S_3 + S_4 + D_{12} + D_{23} + D_{24} + T_{134} + Q)$$
$$O_5 = (S_1 + S_2 + S_4 + D_{13} + D_{23} + D_{34} + T_{124} + Q)$$

$$r_{10} = (S_1 + S_2 + S_3 + S_4 + T_{123} + T_{124} + T_{234} + T_{134})$$

TABLE 2

The linear combinations of the r_i *and* O_i *which are equated to the various exchange probabilities. See text for explanation.*

Exchange probabilities	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8	r_9	r_{10}	O_1	O_2	O_3	O_4	O_5
S_1	+	−	−	−	+	−	−	+	−	+	+	−	+	+	+
S_2	−	+	−	−	+	+	−	+	+	+	−	+	−	−	+
S_3	−	−	+	−	−	+	+	+	+	+	+	−	−	+	−
S_4	−	−	−	+	−	−	+	−	+	+	−	+	+	+	+
D_{12}	+	+	−	−	−	+	−	−	+	−	+	+	+	+	−
D_{23}	−	+	+	−	+	−	+	−	−	−	+	+	−	+	+
D_{34}	−	−	+	+	−	+	−	+	−	−	+	+	+	−	+
D_{13}	+	−	+	−	+	+	+	−	+	−	−	−	+	−	+
D_{24}	−	+	−	+	+	+	+	+	−	−	−	−	+	+	−
D_{14}	+	−	−	+	+	−	+	+	+	−	+	+	−	−	−
T_{123}	+	+	+	−	−	−	+	+	−	+	−	+	+	−	−
T_{124}	+	+	−	+	−	+	+	−	−	+	+	−	−	−	+
T_{234}	−	+	+	+	+	−	−	−	+	+	+	−	+	−	−
T_{134}	+	−	+	+	+	+	−	−	−	+	−	+	−	+	−
Q	+	+	+	+	−	−	−	+	+	−	−	−	−	+	+

The orthogonality of the complete set of 15 R_i and O_i is seen from the table since every row and every column contains eight plus and seven minus terms.

The necessity of introducing extra orthogonal parameters in the study of linkage of more than three genes has also been discussed by JONES (1960) and in a very illuminating paper of SCHNELL (1961). (It is a curious sidelight on the progress of science that within a period of two or three years, three of us have found it

necessary to develop independently nearly identical linkage theory for three quite different purposes in a field that was first exploited mathematically 20 years ago by GEIRINGER [1944].)

It is from Table 2 that equations (4) and (5) have been constructed. For example, a contribution to the gamete class g_{11000} from the zygote g_{11111}/g_{00000} would occur when a single exchange in the region 2–3 occurred. The contribution will be one half of the probability of the appropriate exchange, S_2. The frequency of the zygote g_{11111}/g_{00000} after selection is $2x_0x_{31}W_{0,31}$ so that the net contribution to gamete class g_{11000} is, from Table 2

$$\frac{X_0 X_{31}}{8} (-R_1 + R_2 - R_3 - R_4 + R_5 + R_6 - R_7 + R_8 + R_9 + R_{10} - O_1 + O_2 - O_3 - O_4 + O_5) W_{0,31}$$

The last term in equation (5) will be positive or negative depending upon whether S is even or odd. To understand the significance of S we must interpret the orthogonal recombination values O_i as simultaneous recombinations in non-adjacent *intervals* rather than between pairs of genes (SCHNELL 1961). Looking at Table 1, O_1 is the recombination fraction for *intervals* 1 and 3, O_2 for 2 and 4, etc. Then S is the number of recombinations in these *intervals* for a given O_i necessary to produce the gamete in question from the zygote x_jx_k. For example, to produce a gamete g_{01001} from the zygote g_{00000}/g_{11111} we must have an exchange in regions 1, 2 and 4. From Table 1 we see that O_1 concerns exchanges in intervals 1 and 3. Then for O_1, $S = 1$. On the other hand O_5 concerns regions 1, 2 and 4 so that for O_5, $S = 3$.

Using equations (4) and (5) it is possible to follow the change in gametic frequencies generation after generation or else to examine equilibrium conditions by setting the $\Delta x_i = 0$. Gene frequencies, rather than gametic frequencies are found by summing the appropriate gametic frequencies over all other loci.

The measure of linkage disequilibrium. Another question of interest is the intensity of linkage disequilibrium among the loci either at gene frequency equilibrium or at some intermediate stage of population evolution. For the two-locus case, the measure of linkage disequilibrium usually used is the gametic determinant since this appears explicitly in the equation for change in gametic frequency (equation 1) and because $D = 0$ when there is complete linkage equilibrium. This measure can also be used for multilocus cases by computing separate D values for each pair of loci. Thus the linkage disequilibrium for loci 1 and 3 in the five-locus case would be $D_{13} = g_{1.1..} \, g_{0.0..} - g_{1.0..} \, g_{0.1..}$ where the dot subscripts mean summation over those loci. Formulas for higher order disequilibrium are given by BENNETT (1954).

One difficulty about this measure is that it is sensitive to the gene frequencies so that changes in D reflect both real changes in the intensity of the linkage correlation, but also changes in gene frequency. If p_1 and p_2 are the gene frequencies of A and of B respectively, then at linkage equilibrium the frequency of a gametic type is the product of the appropriate gene frequencies. That is

$$g_{11} = p_1 p_2 \qquad\qquad g_{10} = p_1 (1 - p_2)$$
$$g_{01} = (1 - p_1) p_2 \qquad\qquad g_{00} = (1 - p_1) (1 - p_2)$$

for the two-locus case. It follows from these relationships that the gametic determinant can be written

$$D = [p_1 p_2 + e][(1 - p_1)(1 - p_2) + e] - [p_1(1 - p_2) - e]$$
$$[(1 - p_1) p_2 - e] = e$$

where e is the deviation of the actual gametic frequency from linkage equilibrium. The largest positive value e can take is $p_1(1 - p_2)$ or $p_2(1 - p_1)$, whichever is smaller, while the largest negative value e can take is either $p_1 p_2$ or $(1 - p_1)(1 - p_2)$, whichever is smaller. For example if $p_1 = p_2 = .50$, then e can be as great as $\pm .25$ while if $p_1 = .10$ and $p_2 = .70$, e must be in the limits $- .07$ to $+ .03$. A simple measure of the intensity of linkage disequilibrium then is the ratio of D to the maximum possible e for given gene frequencies. This relative value of disequilibrium, D', is given in the succeeding sections along with D when appropriate.

Numerical solutions by "Genetic Operators." It is clear that equations (4) and (5) are impossible to work with from a practical standpoint. For five loci, for example, equation (5) alone has 660 terms, so that equation (4) has 693 separate terms and there are 32 simultaneous equations like this to be solved. Previous work has been somewhat restricted in its generality precisely because of the practical mathematical difficulties of handling so many very cumbersome equations. General literal solutions to such equations are usually impossible to find except in the very simplest cases. A reasonable insight into two-locus models has been gained in the works previously cited, but even there the most general two-locus models could not be handled. While the theoretical population geneticist would prefer to state his results in general and usable symbolic terms, we have reached an impasse which can only be broken by a more empirical, numerical approach.

Even the decision to attack only specific numerical examples has not made the problem much easier, practically, because even high speed computers cannot cope easily with so many large equations. The method I have used is to bypass equations (4) and (5) completely and go to a more basic method of *genetic operators*. This method is to consider an initial vector of gametic frequencies $[g]_0$ which is transformed to a new vector $[g]_1$, by a transformation T. That is

$$(7) \quad [g]_1 = T([g]_0)$$

We wish to do two things. First we would like to apply the transformation n times so that we can get the gametic frequencies after n generations. Second, we would like to find the value of the vector components, g^*, such that

$$\mathbf{T}([g^*]_0) = [g^*]_0$$

That is, $[g^*]$ is the *equilibrium vector* so that T is an *identity operator* for $[g^*]$. There will, in general, be more than one such vector, but we are interested in those cases in which more than one of the components is non-zero. Thus, the 2^n trivial solutions of the form $[0, 0, 0, \ldots, 1, \ldots, 0]$ in which all gene frequencies are fixed, are known in advance to be solutions but we want to know if there are any others.

The operation T is really a sequence of operations and the flow from $[g]_0$ to $[g]_1$ can be described as in Figure 2. M is the mating operator and for the case
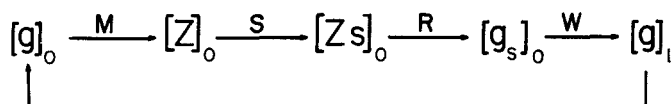
$$[g]_0 \xrightarrow{\ M\ } [Z]_0 \xrightarrow{\ S\ } [Zs]_0 \xrightarrow{\ R\ } [g_s]_0 \xrightarrow{\ W\ } [g]_1$$

FIGURE 2.—The genetic transformation, T, broken up into its components during a single generation.

of random mating it is the row by column multiplication of the vector $[g]_0$ by its transpose $[g]'_0$ to produce the symmetric zygotic matrix $[Z]$.

(8) $[Z] = M([g]_0) = [g]_0 \times [g]'_0$

S is the selection operator and consists in multiplying each element Z_{ij} of the zygotic matrix $[Z]$ by an appropriate fitness value W_{ij} bearing in mind that $W_{ij} = W_{ji}$ and $Z_{ij} = Z_{ji}$.

The matrix of selected zygote frequencies $[Z_s]$ is then used to produce a new vector of selected gametes $[g_s]$ by the recombination operator, R. This operator consists in multiplying a given element of $[Z_s]$ by each of the 2^{n-1} exchange probabilities (including the probability of no exchange) and adding the result of each multiplication to the appropriate element of $[g_s]$. This is then repeated for each element in $[Z_s]$. Finally, $[g_s]$ is converted to $[g]_1$, by the scalar multiplication $[g]_1 = [g_s] \dfrac{1}{\sum\limits_i g_{si}}$. This normalizes the selected g_{si}, bringing their sum back to unity. The sum of the unnormalized g_{si} is $\overline{W}$, the mean fitness of the population.

The operations described above are particularly easy to perform in a binary digital computer.

The procedure described above produces the generation by generation change in the vector $[g]$. To study equilibrium, only a slight modification is necessary. The unnormalized vector $[g_s]$ can be expressed as

(9) $\left[\sum\limits_j \sum\limits_k a_{ijk} g_j g_k \right] = [g_s]$

where the a_{ijk} are the elements of the transformation, T, introduced by the operators S and R. Then, by postulating some initial vector we can solve the vector equation

(10) $\left[\sum\limits_j \sum\limits_k a_{ijk} g_j g_k - g_{si} \right]_0 = 0$

by the standard iteration method of NEWTON.

HETEROTIC SELECTION

There is increasing evidence (WALLACE 1958) that if heterosis at the locus level is important in natural populations, the degree of heterosis at each locus is a function of the amount of heterozygosity at other loci. That is, there is epistatic interaction in the determination of heterozygote superiority. Such epistasis will cause an interaction of linkage and selection at gene frequency equilibrium as shown by LEWONTIN and KOJIMA (1960). In the rest of this paper we will examine the results of this interaction for some two-locus and five-locus models. The

dynamics of two-locus models has been discussed by LEWONTIN and KOJIMA, but solutions of their equations were possible only in certain restricted cases. The present paper will examine some two-locus heterotic models in which symmetry of fitness is not assumed, so that numerical solutions are necessary or in which the stability of equilibria is not obvious and considerable numerical calculation was necessary to test the stability of the equilibria.

Two-locus models: Model 1, whose fitnesses are given in Table 3(a), is a simple heterotic model with epistatic interaction. Each locus shows heterosis in every combination with the other locus, but the heterosis is not additive between loci. Thus, the degree of heterosis at locus *A* is greater when in the presence of *Bb* than in the presence of *BB* or *bb*. A similar inequality holds for the *B* locus, where heterosis is also more pronounced when *A* is heterozygous than when it is homozygous. This is then a *cumulative* heterotic model, fitness increasing more rapidly the more loci that are made heterozygous.

The stable equilibria for Model 1 are given in Table 4. In all results that follow only stable equilibria are given. What is shown are the frequencies of the four gametic types, the gene frequencies of the two loci, the linkage disequilibrium parameter, D, the relative disequilibrium, D', and the mean fitness of the equilibrium population, $\overline{W}$. Several points are worth noting. First, the stable equilibrium of gametic frequencies *does not correspond to linkage equilibrium even with free recombination* $(R = .50)$ although the degree of linkage disequilibrium is small. That is, there is a significant excess of coupling or repulsion at equilibrium no matter how much recombination goes on. Second, there are pairs of solutions for each recombination value, one corresponding to an excess of coupling $(D$ positive) and one to an excess of repulsion $(D$ negative). These paired solutions which are only shown for the tighter linkage values are not symmetrical as was the case in the symmetrical models discussed by LEWONTIN and KOJIMA. Thus there are two possible sets of equilibrium *gene frequencies* for each value of re-

TABLE 3

Relative fitnesses of the nine genotypes for two-locus heterotic models

		AA	Aa	aa
(a) Model 1: asymmetric heterotic model with epistasis				
	BB	.40	.60	.30
	Bb	.60	1.00	.50
	bb	.50	.70	.40
(b) Model 2: asymmetric partially heterotic model with epistasis				
	BB	.5000	.5000	.3750
	Bb	.5625	1.0000	.3125
	bb	.3750	.4375	.3750
(c) Model 3: mixed overdominance, underdominance model				
	BB	.90	.20	.90
	Bb	.20	1.00	.20
	bb	.90	.20	.90

TABLE 4

Results of Model 1. Symbols are as explained in the text

R	g_{00}	g_{01}	g_{10}	g_{11}	p	r	D	D'	$\overline{W}$
.00	.50000	.00000	.00000	.50000	.50000	.50000	+.25000	+1.00000	.70000
	.00000	.58333	.41667	.00000	.58333	.41667	—.24306	—1.00000	.70836
.01	.46225	.05195	.01777	.46805	.51420	.48002	+.21543	+.92384	.69014
	.02359	.55936	.38914	.02791	.58295	.41273	—.21700	—.90191	.70378
.02	.42023	.10875	.03871	.43231	.52898	.45894	+.17746	+.82093	.68044
	.04984	.53246	.35855	.05915	.58230	.40839	—.18797	—.79042	.68902
.03	.37049	.17398	.06621	.38932	.54447	.43670	+.13272	+.66717	.67088
	.08051	.50089	.32332	.09528	.58140	.40383	—.15449	—.65799	.67950
.04									
	.11793	.46211	.28148	.13848	.58004	.39941	—.11374	—.49096	.67038
.06									
	.20082	.37418	.19621	.22879	.57500	.39703	—.02747	—.12033	.65954
.08									
	.21773	.35566	.18039	.24622	.57339	.39819	—.01054	—.04616	.65882
.10									
	.22172	.35125	.17676	.25032	.57297	.39848	—.00659	—.02886	.65878
.30									
	.22703	.34539	.17195	.25563	.57242	.39898	—.00135	—.00591	.65862
.50									
	.22766	.34473	.17141	.25620	.57239	.39907	—.00076	—.00327	.65862

combination, one corresponding to a coupling equilibrium and one to a repulsion equilibrium. Moreover, gene frequencies change with recombination.

Third, the mean adaptive value of the population is highest when there is close linkage and it is higher for repulsion equilibrium than for coupling equilibria. The ratio of fitness at complete linkage to that with free recombination is 1.075, not an immense increase due to the linkage. The most profound change in the population due to linkage is in the genotypic distribution at equilibrium which is in turn a reflection of the very large differences in gametic frequency from one value of linkage to another.

Model 2, whose fitnesses are given in Table 3b and whose results are shown in Table 5 is slightly different from Model 1. Again there is *cumulative heterosis* but here the heterosis disappears in one case: when B is homozygous A shows complete dominance. The results of this model show the same features as Model 1 with a few exceptions. In this case it is the coupling rather than the repulsion equilibrium which have the highest fitnesses. In addition, the general effects of linkage are a great deal stronger. The value of D is ten times greater when there is free recombination than was the case in Model 1, and the ratio of mean fitnesses with complete linkage and free recombination is 1.22. In addition, there is a very strong effect of linkage on the equilibrium *gene* frequencies. In Model 1 the gene frequencies at equilibrium were

$$R = .00 \qquad \hat{p} = .58333 \qquad \hat{r} = .41667$$
$$R = .50 \qquad \hat{p} = .57239 \qquad \hat{r} = .39907$$

TABLE 5

Results of Model 2. Symbols are as explained in the text

R	g_{00}	g_{01}	g_{10}	g_{11}	p	r	D	D'	$\overline{W}$
.00	.55556	.00000	.00000	.44444	.55556	.55556	+.24691	+1.00000	.72223
	.00000	.50000	.50000	.00000	.50000	.50000	—.25000	—1.00000	.68750
.01									
	.01664	.48928	.48593	.00815	.50592	.50257	—.23762	—.96684	.67849
.02	.54063	.02385	.01668	.41884	.56448	.55731	+.22604	+.93128	.70255
	.03563	.47750	.47063	.01624	.51313	.50626	—.22415	—.90940	.66738
.03	.53282	.03652	.02543	.40523	.56934	.55825	+.21499	+.89423	.68779
	.05457	.46552	.45443	.02548	.52009	.50900	—.21016	—.89190	.65730
.05	.51637	.06352	.04396	.37615	.57989	.56033	+.19144	+.81325	.67350
	.10201	.43688	.41605	.04506	.53889	.51806	—.17717	—.79727	.63669
.07									
	.16945	.39738	.36821	.07036	.56683	.53226	—.13225	—.65273	.61463
.075									
	.19509	.38244	.34280	.07967	.57753	.53789	—.11556	—.59187	.60815
.10	.46805	.14242	.09854	.29099	.61047	.56659	+.12216	+.55351	.62830
.15	.41262	.21957	.15828	.20953	.63219	.57090	+.05170	+.24621	.59970
.20	.38645	.24803	.18406	.18146	.63448	.57051	+.02447	+.11734	.59356
.35	.36977	.26391	.19969	.16663	.63368	.56946	+.00891	+.04271	.59138
.50	.36582	.26743	.20328	.16347	.63325	.56910	+.00544	+.02606	.59101

for the repulsion equilibria. This is a very small effect. However, in Model 2 the results are

$$R = .00 \qquad \hat{p} = .55556 \qquad \hat{r} = .55556$$
$$R = .50 \qquad \hat{P} = .63325 \qquad \hat{r} = .56910$$

which represents a considerable change for the first locus.

The fitnesses for third model to be considered are given in Table 3c. This is a symmetrical model of the kind considered by LEWONTIN and KOJIMA, but has certain peculiarities which require careful investigation. Here there is strong heterosis at one locus provided the other locus is heterozygous, but selection *against* the heterozygote when the other locus is homozygous. This model is not presented as representing a particular natural situation, but rather to show the intricacy of the possible interactions between linkage and natural selection.

The results for this model are given in Table 6. These values were computed by formula (18) of LEWONTIN and KOJIMA

$$X_i = \frac{1}{4} \pm \frac{1}{4} \sqrt{1 + \frac{4rd}{b+c-a-d}}$$

for symmetrical fitness models and were checked by the method of genetic operators. There was perfect agreement between them. The startling feature of the results is the existence of three distinct regions of solutions. From complete linkage to $R = .10$ there is a stable equilibrium of gene frequencies with both loci held at a frequency of .50, but with very intense *linkage disequilibrium*. As with all such symmetrical models there are two complementary equilibria, one in coupling (D positive) and one in repulsion (D negative). Above $R = .10$ and be-

R. C. LEWONTIN

TABLE 6

Results of Model 3

R	g_{00}	g_{01}	g_{10}	g_{11}	p	r	D'	$\overline{W}$
.00	.50000	0	0	.50000	.50000	.50000	1.00000	.95000
.01	.49667	.00333	.00333	.49667	.50000	.50000	.98658	.94000
.02	.49324	.00676	.00676	.49324	.50000	.50000	.97297	.93000
.03	.48979	.01021	.01021	.48979	.50000	.50000	.95916	.92000
.04	.48629	.01371	.01371	.48629	.50000	.50000	.94516	.91000
.06	.47913	.02087	.02087	.47913	.50000	.50000	.91651	.89000
.08	.47174	.02826	.02826	.47174	.50000	.50000	.88694	.87000
.10	.46409	.03591	.03591	.46409	.50000	.50000	.85636	.85000
.10 to .375			no stable equilibrium of gene frequencies					
.375 to .50	.25000	.25000	.25000	.25000	.50000	.50000	0	.57500

low $R = .375$ *there is no stable equilibrium of any kind.* That is, the gene frequencies go to fixation under natural selection. Then when R exceeds .375 there is a stable equilibrium with gene frequencies at .50 and perfect *linkage equilibrium.* Thus we have a case where either tight or loose linkage results in the maintenance of genetic variation, but intermediate linkage results in a loss of genetic variation.

Five-locus models: For the five-locus models the following simplifying assumptions have been made: (1) The loci are interchangeable in their effects; for example, the genotypes
$$\frac{01110}{01010}, \frac{11001}{11000}, \frac{00101}{00111}, \text{ etc.}$$
are indistinguishable in their fitnesses since each one is homozygous 0/0 at two loci, homozygous 1/1 at two loci and heterozygous 1/0 at one locus. (2) There is some heterosis for each locus, irrespective of whether the other four loci are heterozygous or homozygous. This assures that all *gene* frequencies will come to a stable equilibrium of gene frequencies at an intermediate value. (3) In view of (2) there is no loss of generality by further specifying that 0/0 and 1/1 homozygotes have equal fitness so that the gene frequencies at each locus come to equilibrium at $p = q = .50$.

These assumptions do not restrict the generality of the results, but have been made in order to make the problem more manageable. With five loci there are $3^5 = 243$ different genotypes and each one could be given a unique fitness. Assumption 1 reduces this number to 21, since it would be impossible to explore the immense variety of possibilities with 243 different fitnesses. The three restrictions together result in there being only six different fitnesses depending upon the number of loci heterozygous. Table 7 gives the fitnesses of these genotypes for two models of heterotic selection to be discussed.

Both models show *cumulative* heterosis. That is at each increase in heterozygosity there is a more than linear increase in fitness, this more than linear increase representing the epistatic interaction among the loci. Let W_1, W_2 and W_3 repre-

TABLE 7

Fitnesses of genotypes with different numbers of loci heterozygous for the two five-locus models with heterosis

Number of loci heterozygous	Model 4		Model 5	
	W	E	W	E
0	.06	..	.03	..
1	.09	..	.06	..
2	.18	.33	.12	.25
3	.33	.18	.24	.25
4	.54	.11	.48	.25
5	.81	.07	.96	.25

sent the fitnesses of three successively greater degrees of heterozygosity. Then

$$e = \frac{W_1 + W_3 - 2W_2}{W_3}$$

is a measure of the *relative epistatic effect* of increasing heterozygosity. As Table 7 shows, Model 1 was chosen to have *decreasing epistatic interaction* with added heterozygosity, so that the increase from four to five loci heterozygous is accompanied by an increase in fitness nearly equal to that found in the increase from three to four loci heterozygous. Models 2 however, shows a *constant epistatic interaction* from level to level of heterozygosity.

Each model has been examined for the equilibrium conditions of gametic frequencies. Because of the symmetry of the models, gene frequencies at equilibrium always equal .50 at all loci, and reciprocal gametic types always have equal frequencies. That is: $g_{11111} = g_{00000}$, $g_{11110} = g_{00001}$, etc.

The results, then, show only the frequencies of the first 16 gametic types g_{00000} through g_{01111}. In addition the values of the relative linkage disequilibrium parameters among all pairs of loci are given, D'_{12} through D'_{45}. Since the gene frequencies all equal .50, these D' values are always four times the equivalent D values. Finally $\overline{W}$, the mean fitness is also given.

Tables 8 and 9 show this information for the five-locus models investigated for different values of recombination. The model is that the five genes are equally spaced along the linkage map with a linkage distance R between *adjacent* genes. Thus, for $R = .05$ the total linkage distance between Loci 1 and 5 is .20.

In both models certain features are common. Linkage is only effective for fairly small recombination values between adjacent loci. However, there is a *cumulative effect along* the chromosome so that the outside genes are in linkage disequilibrium even though they are quite far apart on the linkage map. Selection holds Loci 1 and 2 out of linkage equilibrium and also Loci 2 and 3 with the result that 1 and 3 are also out of equilibrium, and so on down the chromosome. The closer the loci, the greater the linkage disequilibrium, the values of D' being in the order

$$D'_{12} > D'_{13} > D'_{14} > D'_{15}$$

In addition, there is a small effect of *absolute position* in the linkage maps as far

TABLE 8

Results of Model 4. Symbols are as explained in the text

Gametes	R between adjacent loci									
	.000	.002	.01	.02	.03	.0325	.0338	.0350	.0367	$\geq .04$
00000	.50000	.48860	.43945	.36690	.27027	.23742	.21789	.19440	.13407	.03125
00001	0	.00339	.01718	.03474	.05166	.05534	.05694	.05822	.05762	.03125
00010	0	.00002	.00057	.00286	.00886	.01164	.01346	.01578	.02222	.03125
00011	0	.00226	.01155	.02363	.03585	.03876	.04015	.04144	.04259	.03125
00100	0	.00001	.00038	00199	.00636	.00846	.00987	.01170	.01718	.03125
00101	0	.00000	.00002	.00024	.00154	.00248	.00322	.00433	.00872	.03125
00110	0	.00001	.00044	.00224	.00707	.00938	.01092	.01293	.01894	.03125
00111	0	.00226	.01155	.02363	.03585	.03876	.04015	.04144	.04259	.03125
01000	0	.00002	.00057	.00286	.00886	.01164	.01346	.01578	.02222	.03125
01001	0	.00000	.00002	.00027	.00173	.00278	.00361	.00485	.00977	.03125
01010	0	.00000	.00000	.00003	.00037	.00072	.00104	.00159	.00441	.03125
01011	0	.00000	.00002	.00024	.00154	.00248	.00322	.00433	.00872	.03125
01100	0	.00001	.00044	.00224	.00707	.00938	.01092	.01293	.01894	.03125
01101	0	.00000	.00002	.00027	.00173	.00278	.00361	.00485	.00977	.03125
01110	0	.00002	.00061	.00310	.00960	.01264	.01463	.01722	.02462	.03125
01111	0	.00039	.01718	.03474	.05166	.05534	.05694	.05822	.05762	.03125
D'_{12}	1.00000	.98624	.92456	.82492	.66980	.60900	.57032	.52092	.37572	0
D'_{13}	1.00000	.97716	.87740	.72616	.51652	.44312	.39904	.34556	.20048	0
D'_{14}	1.00000	.96812	.83232	.63808	.39688	.32116	.27804	.22828	.11316	0
D'_{15}	1.00000	.95480	.76984	.52892	.27384	.20512	.16868	.12928	.05040	0
D'_{23}	1.00000	.99076	.94800	.87396	.74676	.69320	.65816	.61224	.46984	0
D'_{24}	1.00000	.98164	.89940	.76796	.57192	.49948	.45496	.40004	.25184	0
D'_{25}	1.00000	.96812	.83232	.63808	.39688	.32116	.27804	.22828	.11316	0
D'_{34}	1.00000	.99076	.94800	.87396	.74676	.69320	.65816	.61224	.46984	0
D'_{35}	1.00000	.97716	.87740	.72616	.51652	.44312	.39904	.34556	.20648	0
D'_{45}	1.00000	.98624	.92456	.82492	.66980	.60900	.57032	.52092	.37572	0
$\overline{W}$	.43500	.42852	.40251	.36941	.33395	.32930	.31838	.31218	.29851	.28500

as linkage distance is concerned, Loci 1 and 2 are equivalent to Loci 2 and 3, yet in all cases D'_{12} is smaller than D'_{23}. In general

$$D'_{12} = D'_{45} < D'_{23} = D'_{34}$$

and $D'_{13} = D'_{35} < D'_{24}$

That is, a pair of loci in the middle of the linkage group is held in greater disequilibrium than a pair of loci near the ends of the linkage group.

A third feature, also seen in the two-locus models, is that the mean fitness, $\overline{W}$, of the population at equilibrium is greater for linked cases than for unlinked ones. This is a result of the simultaneous selection in a single individual of several deleterious homozygotes. When there is linkage disequilibrium, the death of one organism removes from the population homozygous genotypes at several loci in a greater frequency than when the loci associate at random, with the result that fewer individuals need be selected against: the segregation "load" is less and the mean fitness is greater.

TABLE 9

Results of Model 5. Symbols are as explained in the text

Gametes	R between adjacent loci									
	000	.01	.02	.03	.04	.05	.06	.063	.0645	.065
00000	.50000	.46199	.42053	.37444	.32183	.25904	.17488	.13627	.09817	.03125
00001	0	.01083	.02193	.03316	.04418	.05411	.05997	.05874	.05413	.03125
00010	0	.00016	.00074	.00201	.00438	.00863	.01675	.02119	.02567	.03125
00011	0	.00775	.01572	.02384	.03192	.03947	.04495	.04515	.04336	.03125
00100	0	.00010	.00048	.00133	.00299	.00611	.01254	.01642	.02087	.03125
00101	0	.00000	.00003	.00013	.00044	.00135	.00458	.00754	.01213	.03125
00110	0	.00013	.00061	.00166	.00363	.00723	.01443	.01869	.02344	.03125
00111	0	.00775	.01572	.02384	.03192	.03947	.04497	.04515	.04336	.03125
01000	0	.00016	.00074	.00201	.00438	.00863	.01675	.02119	.02567	.03125
01001	0	.00000	.00003	.00015	.00050	.00155	.00524	.00859	.01370	.03125
01010	0	.00000	.00000	.00001	.00006	.00029	.00164	.00341	.00700	.03125
01011	0	.00000	.00003	.00013	.00044	.00135	.00458	.00754	.01213	.03125
01100	0	.00013	.00061	.00166	.00363	.00723	.01443	.01869	.02344	.03125
01101	0	.00000	.00003	.00015	.00050	.00155	.00524	.00859	.01370	.03125
01110	0	.00019	.00088	.00234	.00504	.00985	.01905	.02410	.02912	.03125
01111	0	.01083	.02193	.03316	.04418	.05411	.05997	.05874	.05413	.03125
D'_{12}	1.00000	.95476	.90300	.84164	.76508	.66172	.49236	.39660	.28448	0
D'_{13}	1.00000	.92352	.83888	.74296	.63072	.49236	.29912	.20836	.11928	0
D'_{14}	1.00000	.89284	.77752	.65208	.51380	.35836	.17452	.10408	.04720	0
D'_{15}	1.00000	.85140	.69840	.54184	.38376	.22808	.08192	.03984	.01344	0
D'_{23}	1.00000	.96744	.92944	.88300	.82260	.73604	.58104	.48584	.36680	0
D'_{24}	1.00000	.93572	.86316	.77876	.67656	.54488	.34880	.25104	.15068	0
D'_{25}	1.00000	.89284	.77752	.65208	.51380	.35836	.17452	.10408	.04720	0
D'_{34}	1.00000	.96744	.92944	.88300	.82260	.73604	.58104	.48584	.36680	0
D'_{35}	1.00000	.92352	.83888	.74296	.63072	.49236	.29912	.20836	.11928	0
D'_{45}	1.00000	.95476	.90300	.84164	.76508	.66172	.49236	.39660	.28448	0
$\overline{W}$	.49500	.45688	.41927	.38203	.34491	.30738	.26720	.25240	.24021	.22781

A final feature of these models, also seen in two-locus models, is the existence of multiple equilibrium conditions. Tables 8 and 9 only show a single equilibrium array of gametes for each value of R. Actually there are 15 additional arrays for each case which are complementary to the one shown. This is because the model is symmetrical with a 0/0 homozygote equivalent in fitness to a 1/1 homozygote and all loci identical in effect. Thus by interchanging 0 and 1 at any locus a new array can be produced. The sign of the D values will depend upon the array but the absolute values will be the same, as will the mean fitness, $\overline{W}$. When 0/0 and 1/1 homozygotes have identical fitness there is no meaning to "coupling" or "repulsion" gametes and the signs can be ignored.

The difference between the two models is in the direction expected. The first model in which there is decreasing epistasis with increasing heterozygosity shows no effect of linkage above $R = .04$. The critical value of R above which there is no effect of linkage was not determined exactly but it lies between .038 and 0.04. In Model 5 where the epistasis is constant, the critical value is approximately .065.

In Model 4 the ratio of fitness with complete linkage to fitness with free recombination is .435/.285 = 1.53, while in Model 5 this ratio is 2.17.

Correspondence to Experiments

The models discussed above make specific predictions about the course of selection in a population segregating for several genes in the same linkage group. If there is heterosis and epistasis it is predicted that there will be permanent linkage disequilibrium when recombination is within certain limits, and that genes at opposite ends of the linkage group map be held out of linkage equilibrium by their association with genes between them in the sequence. Any *exact* comparison of prediction to experiment is virtually impossible with five loci, since the fitnesses of all the genotype would be virtually impossible to measure. However, the qualitative aspects of these predictions can be tested in experimental populations in which five loci are segregating, and Table 10 shows such experimental results. This table comes from the experiments of DR. GRACE B. CANNON of the Zoology Department of Washington University, St. Louis, to whom I am deeply grateful for allowing me to make use of her experimental data and analysis. These data have since been published (CANNON 1963).

The experiment involves five mutant markers on the third chromosome of *Drosophila melanogaster*. These markers have the following linkage relations
se–32.5–ss–5.5–k–6.7–e–20.4–ro
The original populations were made up by introducing into a wild-type population a few chromosomes with the constitutions: se + + + +, + ss k e + and + + + + ro. The initial gene frequencies are shown in Table 10 for three repli-

TABLE 10

Results of five-locus experiments in Drosophila melanogaster *with genes*
se, ss, k, e and ro. *Data of* DR. GRACE B. CANNON

		Population and week								
		Population 20			Population 21			Population 22		
		0	28	50	0	28	50	0	28	50
(a) Gene frequencies										
se		.007	.102	.058	.007	.044	.073	.005	.026	.037
ss		.012	.052	.216	.012	.078	.203	.009	.106	.186
k		.012	.026	.200	.012	.100	.177	.009	.092	.175
e		.012	.013	.174	.012	.133	.219	.009	.106	.181
ro		.007	.064	.084	.007	.066	.094	.005	.026	.048
(b) D and D' values										
ss–k	D		+.0247	+.1408		+.0610	+.1166		+.0693	+.1328
	D'		+1.0000	+.8980		+.6616	+.8265		+.8426	+.9323
k–e	D		−.0003	+.1182		+.0781	+.1231		+.0823	+.1173
	D'		−1.0000	+.8491		+.9008	+.8905		+1.0000	+.8184
ss–e	D		+.0123	+.1154		+.0588	+.0907		+.0810	+.1039
	D'		+1.0000	+.8459		+.8695	+.5721		+.8547	+.7052

cate populations. Two subsequent samples were taken from the populations, one after 29 weeks and one after 50 weeks, or approximately 15 and 25 generations after introduction of the mutant genes. The gene frequencies at these times are shown in Table 10a. Dr. CANNON's samples were taken in such a way that *gametic* frequencies could be estimated, and the results for the three sample periods are given in Table 10, in terms of D'_{ij}, the relative linkage disequilibrium parameters between gene pairs.

Table 10b shows only the D' values for the three middle genes, *ss*, *k* and *e* for the following reason. The frequency of the two outside genes *se* and *ro* is quite low so that gametes carrying them are rare. As a result estimates of linkage disequilibrium involving these genes are variable and unreliable. The significance of D values can be tested by a two-by-two chi-square, testing the association between two loci. Such tests performed on the results of the last generation (Week 50) give probabilities between .1 and .8 for all associations involving the outside genes *se* and *ro*. That is, there is no significant linkage disequilibrium involving these genes. The three D values among the central three genes, however, are highly significant with probabilities much less than .01 for all three D values in all three populations. It is only the significant D values that are shown in Table 10b.

As a whole, the results are clear-cut and in excellent agreement with the theoretical prediction. All mutant genes increased in frequency during the time of the experiment. Whether or not they have reached equilibrium it is impossible to say, but the fact that they have increased is remarkable in the face of our usual assumption that visible mutants are deleterious. Either these mutations are heterotic, or there is gene-frequency-dependent selection keeping them in the population, or else they are tightly linked to other genes affecting fitness. Whatever the explanation, the mutants are maintained at high frequency.

When we look at Table 10 we see that accompanying the maintenance of these genes is the maintenance of pronounced linkage disequilibrium of the three closely linked genes *ss* *k* and *e* which are all within a map length of 12 centimorgans. Without taking selection into account we would expect the linkage disequilibrium among these genes to have decayed considerably in the 25 generations of the experiment. Taking into account the lack of recombination in males, we would expect for *ss* and e that

$$D'_{25} = (.939)^{25} D'_0 = .207$$

whereas D'_{25} is actually .846 (Population 20). The same lack of decay is also seen for the *ss–k* and *k–e* intervals. On the other hand the outside genes *se* and *ro* are not in intense disequilibrium with each other or with the middle three genes, since the chi-square tests on all three populations show that the only linkage disequilibrium values significantly different from zero are among the three middle genes *ss*, *k* and *e*.

Another way of seeing the selection of linked blocks of genes is shown in Table 10a. The three middle genes have all increased in frequency together while the two loosely linked genes *se* and *ro* have much lower frequencies. Thus selection is operating on the unrecombined block as a whole.

General Implications of the Results

One of the most important and difficult problems in population genetics at the present time is the extent to which heterosis accounts for the observed genetic variation in populations. A telling argument that has been advanced against widespread heterosis is that the cost (genetic load) to a population of maintaining many heterotic genes is unbearable, especially if we assume that this load acts through loss of zygotes. Thus if a single locus has a heterotic loss of say, 5 percent, then only .95 of the population survives. Adding a second locus like this leaves only .9025 of the population and so on, so that 100 such heterotic loci could be maintained only at a cost of killing .995 of the population. What the results of the investigation of linkage show is that this load on the population is considerably reduced if the genes are linked to each other so that simultaneous elimination of homozygotes at different loci can occur. In particular, the fitness in Model 2 of the four-locus models can be more than doubled by this linkage effect.

As I have repeatedly pointed out, epistasis is required in order for linkage to be important in natural selection. Can interaction between genes be sufficiently strong and widespread to make linkage important? The answer is clearly, yes, because epistasis *as defined in this context always occurs on the scale of adaptive values in nature*. By epistasis (or interaction) we mean a deviation from arithmetic additivity of the effects of genes at two different loci. But arithmetic additivity is not a usual property of fitnesses. For example if homozygosity for allele A at one locus results in a fitness of .5, and homozygosity for allele B at a second locus also results in a fitness of .5, then perfect additivity requires the fitness of $AA\ BB$ to be zero whereas in fact it is likely to be about .25. Fitnesses tend to be *multiplicative* rather than *additive* and this fact, in itself, gives rise to interaction as we have defined it. Model 5 of the five-locus models (Table 7) is a perfectly multiplicative model, each substitution of a homozygous genotype cutting the fitness by one half. In one sense there is no interaction in this model since the genes can be thought of as acting independently, yet the effect is to produce considerable epistasis on the additive scale and this is sufficient to make linkage important. The fact that Dr. CANNON's data on five loci chosen at random shows such pronounced effects of linkage is added evidence for the importance of taking linkage into account in our formulations.

Finally, the five-locus models add an important observation not seen in the two-locus models. This is the *cumulative* effect of the linkage along the chromosome. Even if two genes are loosely linked, they may be held out of linkage equilibrium with each other if loci between them are out of linkage equilibrium because of selection.

SUMMARY

The general problem of the interaction between linkage and selection has been examined for a number of multilocus models. General equations and a method of "genetic operators" are given for the solution of this problem. Specific numerical cases for two- and five-locus models exhibiting heterosis have been examined with the following results: (1) Loci may be kept in permanent linkage disequilibrium despite gene frequency equilibrium, by natural selection. (2) When epistasis between loci is strong enough disequilibrium will be maintained for genes that are completely unlinked. (3) The epistasis which results in disequilibrium can be generated by simple multiplicative fitnesses, a common situation, if not the most common. (4) The linkage disequilibrium results in higher mean fitness. (5) In multiple-locus models genes quite far apart on the chromosome may be held out of linkage equilibrium by genes between them along the chromosome. The effect is then *cumulative* along the chromosome. (6) Some experiments with Drosophila are reviewed and the predictions of the models are upheld in the experiments.

LITERATURE CITED

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

BODMER, W. F., and P. A. PARSONS, 1962   Linkage and recombination in evolution. Advan. Genet. **11**: 1–87.

CANNON, G. B., 1963   The effects of natural selection on linkage disequilibrium and relative fitness in experimental populations of *Drosophila melanogaster*. Genetics **48**: 1201–1216.

GEIRINGER, H., 1944   On the probability theory of linkage in Mendelian heredity. Ann. Math. Statist. **15**: 25–57.

JONES, R. M., 1960   Linkage distributions and epistacy in quantitative inheritance. Heredity **15**: 153–159.

KIMURA, M., 1956   A model of a genetic system which leads to closer linkage by natural selection. Evolution **10**: 278–287.

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

SCHNELL, F. W., 1961   Some general formulations of linkage effects in inbreeding. Genetics **46**: 947–957.

WALLACE, B., 1958   The average effects of radiation-induced mutations on viability in *D. melanogaster*. Evolution **12**: 532–552.