THE EVOLUTIONARY DYNAMICS OF COMPLEX POLYMORPHISMS[1,2,3]

R. C. LEWONTIN

Department of Biology, University of Rochester

AND

KEN-ICHI KOJIMA

Department of Genetics, North Carolina State College

Received March 7, 1960

The theory of balanced polymorphism which has been elaborated by population geneticists, notably Sewall Wright and R. A. Fisher, has in the main been concerned with the effects of single loci. As theory generally goes apace with experiment, this accent on single locus polymorphisms has been due to the plethora of observational evidence relating to simple cases. It is sufficient to note the vast effort made by Dobzhansky and his co-workers in their elucidation of the inversion polymorphism of the third chromosome of *Drosophila pseudoobscura*.

In recent years, however, a few cases have come to light of polymorphisms involving more than one Mendelian unit. Among these are the inversions on different chromosomes found in *D. robusta* studied by Levitan (1955 and 1958), the shell color of *Cepaea nemoralis* reported by Lamotte (1951) and by Cain and Sheppard (1952), the complex mimicry pattern in certain butterflies (Sheppard, 1959) and the inversions in two chromosomes of the grasshopper *Moraba scurra* analyzed by White (1957) and Lewontin and White (1960).

The study of effects of natural selection on single locus polymorphisms must take into account only inter-allelic effects such as additivity and dominance. In multi-locus polymorphisms, however, two new factors add more complexity. They are recombination between loci and interactions between the loci in determining fitnesses of genotypes (*epistasis* in the broad sense).

Some theoretical work has been done on selection in multi-locus systems, notably the paper of Kimura (1956) on a special case of epistasis and heterosis. Kimura's equations (p. 279) are, however, quite general in their applicability and are completely analogous to our equations 12a–12d although arrived at from a different point of view.

It is our purpose in this article to present the basic theoretical considerations for the analysis of multiple polymorphisms, by examining in some detail the evolutionary dynamics of two-locus systems. It is believed, in view of the results to be presented, that no great gain in understanding would be made by an extremely unwieldy analysis of more complex systems.

LINKAGE EQUILIBRIUM AND DISEQUILIBRIUM

Essential to a discussion of the two-locus problem is the concept of *linkage equilibrium*. This problem has been dealt with by Geiringer (1944) and can be summarized as follows: Suppose that there are two segregating loci, A–a and B–b. There will then be four possible gametic types: AB, Ab, aB and ab. Let the frequencies of these four gametic types be g_{11}, g_{10}, g_{01}, g_{00}, respectively. These frequencies satisfy the following

[1] Part of the work reported here was done under Grant No. RG-6223 of the Division of Research Grants of the National Institutes of Health.

[2] This work was also supported by a National Science Foundation Grant.

[3] Journal Paper No. 1138 of the North Carolina Agricultural Experiment Station.

relationship:

$g_{11}+g_{10}=p$, the frequency of allele A,
$g_{01}+g_{00}=(1-p)$, the frequency of allele a,
$g_{11}+g_{01}=r$, the frequency of allele B,
$g_{10}+g_{00}=(1-r)$, the frequency of allele b.

Linkage equilibrium will be defined by the condition that each gametic frequency is the product of the appropriate gene frequencies. That is, at linkage equilibrium

$$\hat{g}_{11}=pr, \qquad \hat{g}_{10}=p(1-r),$$
$$\hat{g}_{01}=(1-p)r, \quad \hat{g}_{00}=(1-p)(1-r). \quad (1)$$

It can be shown that at a non-equilibrium condition the gametes will have the frequencies

$$\begin{aligned} g_{11} &= \hat{g}_{11}+D, & g_{10} &= \hat{g}_{10}-D, \\ g_{01} &= \hat{g}_{01}-D, & g_{00} &= \hat{g}_{00}+D, \end{aligned} \quad (2)$$

where $D = g_{11}g_{00} - g_{10}g_{01}$. The equations in (2) imply that any time the gametic frequencies will deviate from the equilibrium frequencies by an amount D which is the product of the coupling gametic frequencies minus that of the repulsion gametic frequencies. D, thus defined, may be considered as a measure of *linkage disequilibrium*. Obviously, at equilibrium

$$\hat{D}=\hat{g}_{11}\hat{g}_{00}-\hat{g}_{10}\hat{g}_{01}$$
$$=pr(1-p)(1-r)-p(1-r)r(1-p)=0.$$

Finally it should be noted that in the absence of any evolutionary pressure such as selection, the value of D will approach zero with succeeding generations. The relation between the initial value, D_0, and the value in the n-th generation through random mating is

$$D_n = (1-R)^n D_0 \quad (3)$$

where R is the recombination frequency between the loci. It is important to stress that the relation (3) is true only when no other evolutionary forces are assumed, because, as will be shown, natural selection may upset the approach to eventual linkage equilibrium.

Populations in Linkage Equilibrium

The approach used in this paper for investigating the interactions of recombination and gene effects is to first assume, incorrectly, that the evolving population is in linkage equilibrium in every generation, i.e., that $D_n = 0$ for all n, and then this inexact model will be compared with the exact treatment in which no such assumption is made.

With two loci and two alleles at each locus all possible genotypes can be tabulated in a two-way table of nine cells, in each of which are found the frequency and adaptive value of the individual genotypes (table 1). The frequency of each genotype is denoted by Z_{ij}, and the adaptive value by W_{ij}. The latter is the relative probability that a zygote of this genotype will leave a zygote in the next generation. The subscript, i, stands for the number of alleles, A, and j for the number of alleles, B, in each genotype. The mean adaptive value of the population is obtained from Table 1 as

$$\bar{W} = \sum_{i,j=0}^{2} Z_{ij} \cdot W_{ij}. \quad (4)$$

TABLE 1. *The frequencies (the upper entries) and adaptive values (the lower entries) of all possible genotypic arrays with two loci (see text)*

	AA	Aa	aa	Marginal mean
BB	Z_{22}	Z_{12}	Z_{02}	$\bar{W}_{BB}$
	W_{22}	W_{12}	W_{02}	
Bb	Z_{21}	Z_{11}	Z_{01}	$\bar{W}_{Bb}$
	W_{21}	W_{11}	W_{01}	
bb	Z_{20}	Z_{10}	Z_{00}	$\bar{W}_{bb}$
	W_{20}	W_{10}	W_{00}	
Marginal mean	$\bar{W}_{AA}$	$\bar{W}_{Aa}$	$\bar{W}_{aa}$	$\bar{W}$

Under the assumptions of random mating and of linkage equilibrium at every generation, the Z_{ij}'s are the products of the appropriate gametic frequencies in formulae (1). For example, $Z_{22} = p^2r^2$, $Z_{21} = 2p^2r(1 - r)$ and so on. From these and the set of W_{ij} given, it is possible to calculate the mean adaptive value, $\bar{W}$, for any combination of gene frequencies, p and r. These mean adaptive values may then be put in the form of *adaptive topography*, a concept introduced by Wright (1932). Figure 1 shows a topography taken from Lewontin and White (1960). The two axes represent the two gene frequencies p and r. The lines in the plane (*isodapts*) join points of equal $\bar{W}$ in the manner of lines of equal altitude in a topographic map. Two adaptive peaks marked by P, two valleys marked by V and a saddle point, S, are seen in figure 1.

It has been shown by Wright (1942) and others that in the simple case of linkage equilibrium the peaks represent points of stable gene frequency equilibrium, while the saddle or minimax point is one of unstable equilibrium. That is, in a population the gene frequencies will change under the influence of natural selection until they arrive at a peak, and, although the frequencies may remain in equipoise at a saddle point, any slight disturbance will cause them to change toward a peak. In the case of two peaks, as in figure 1, the particular peak to which the frequencies move, depends upon the initial frequencies. The genetic conditions for a stable gene frequency equilibrium when linkage disequilibrium is unimportant are given by Kojima (1959).

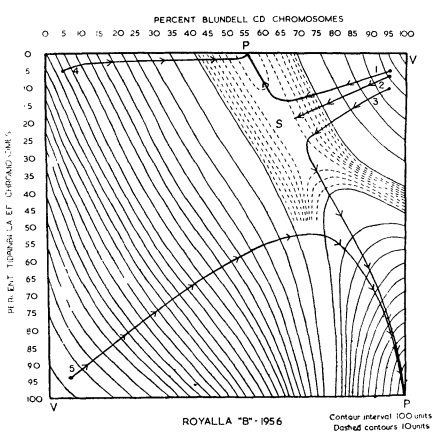

FIG. 1. Adaptive topography for Royalla population of *Moraba scurra* taken from Lewontin and White (1960). The abscissa is the frequency of the Blundell inversion on chromosome *CD*, the ordinate is the frequency of the Tidbinbilla inversion on Chromosome *EF*. Lines are those of equal adaptive value (isodapts). The numbered trajectories are the paths of gene frequency change computed from equations 24 *a-c*. *P*, *V*, and *S* show the locations of adaptive peaks, valleys and a saddle respectively.

What is to be examined further here is the trajectory of the population as it approaches equilibrium. This will then be compared with the trajectory and equilibrium situation when linkage equilibrium is no longer assumed.

It has been shown by Wright (1942) that for linkage equilibrium the changes in gene frequency per generation, Δp and Δr, may be written as

$$\Delta p = \frac{1}{2\bar{W}} p(1 - p) \frac{\delta \bar{W}}{\delta p} \qquad (5.a)$$

$$\Delta r = \frac{1}{2\bar{W}} r(1 - r) \frac{\delta \bar{W}}{\delta r} \qquad (5.b)$$

or in the case of a continuous change of gene frequency with time,

$$\frac{dp}{dt} = \tfrac{1}{2}p(1 - p) \frac{\delta \bar{W}}{\delta p} \qquad (6.a)$$

$$\frac{dr}{dt} = \tfrac{1}{2}r(1 - r) \frac{\delta \bar{W}}{\delta r}. \qquad (6.b)$$

Combining (6.a) and (6.b) to eliminate dt, yields,

$$r(1 - r) \frac{\delta \bar{W}}{\delta r} dp = p(1 - p) \frac{\delta \bar{W}}{\delta p} dr. \qquad (7)$$

By integrating both sides of (7), an equation for the trajectory of p and r in the gene frequency plane can be obtained. This equation will contain the initial gene frequencies, p_0 and r_0, which

determine the specific paths of gene frequency changes. To see the significance of the equation (11), it is fruitful to analogize the path of genetic evolution with the path of a particle moving in a two-dimensional potential field. The coordinates of the particle are the p- and r-axes, and the isodapts are equipotential lines. This type of analogy has been explored to some extent by Wright (1955) and Lewontin (1958) for a single locus case. In such a potential field a particle will move so as to make the maximum change of potential per unit length of path. This can be represented in the symbolism used in (7), by the differential equation

$$\frac{\delta \bar{W}}{\delta r}\, dp = \frac{\delta \bar{W}}{\delta p}\, dr, \tag{8}$$

and the rate of change of potential along this path is

$$\sqrt{\left[\frac{\delta \bar{W}}{\delta r}\right]^2 + \left[\frac{\delta \bar{W}}{\delta p}\right]^2}. \tag{9}$$

(Proofs of (8) and (9) may be found in any elementary text of calculus in discussions of directional derivatives and gradients.) Now a comparison of (7) and (8) shows that gene frequencies do *not* change in such a way that the rate of increase of $\bar{W}$ is maximized, although eventually the population does reach a point of maximum mean adaptive value (Wright, 1949). In a sense the population takes an "indirect" path to the peak. It is to be emphasized that the actual position of a population in the gene frequency plane as well as the gradient in fitness determines the vectors of gene frequency displacement and the change of mean adaptive value is only a resultant of gene frequency changes.

To illustrate the difference between the "steepest path" predicted by (8) and the "genetic path" given by (7) we have solved these equations for two cases. In Case I, the adaptive values which are given in table 2, show no epistasis, while for Case II given in table 3 there is a strong epistasis or non-allelic interaction between the two loci. In figure 2 the steepest path (solid curve) and the genetic path (dashed curve) are given for Case I. Several sets of paths are shown, each corresponding to different initial values of p and r. Because of the symmetry of the adaptive topography with one central peak at $p = r = .50$, only one quadrant of the

TABLE 2. *Hypothetical fitness values of nine genotypes of locus A and locus B. No epistasis*

	AA	*Aa*	*aa*
BB	1	3	1
Bb	2	4	2
bb	1	3	1

TABLE 3. *Hypothetical fitness values showing epistasis between locus A and locus B*

	AA	*Aa*	*aa*
BB	1	1.5	1
Bb	2.5	4	2.5
bb	1	1.5	1

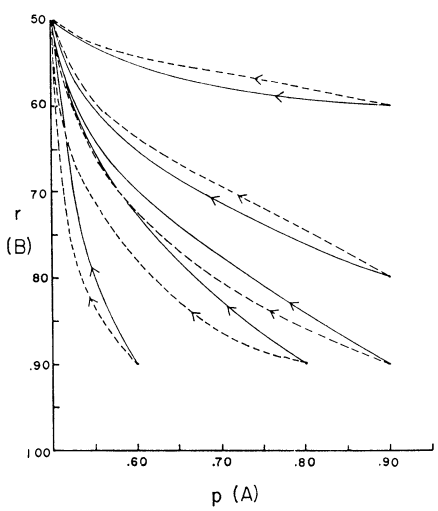

FIG. 2. Steepest path (solid lines) and path taken by gene frequencies with constant linkage equilibrium (broken lines) for adaptive values in table 2 (no epistasis). Ordinate and abscissa are frequencies of alleles at two loci.

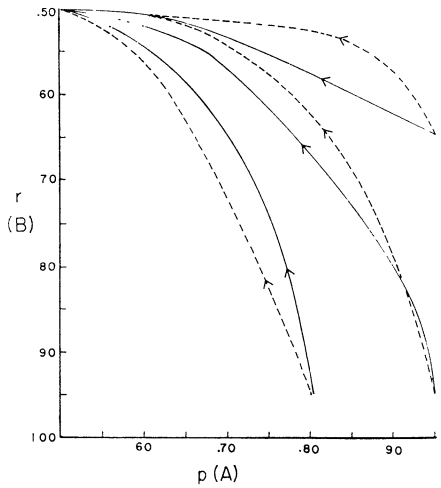

FIG. 3. Steepest path (solid lines) and path taken by gene frequencies with constant linkage equilibrium (broken lines) for the adaptive values in table 3 (epistasis). Ordinate and abscissa are frequencies of alleles at two loci.

p—r plane is shown. The results of calculations for Case II are shown in figure 3. The deviations of the genetic paths from the steepest paths are in the direction of a larger vector component for that gene frequency closer to .50. The paths are more divergent for the epistatic case than for the non-epistatic, these deviations being in some cases quite large.

One of the most interesting concomitants of this effect arises when there is more than one adaptive peak in the topography as in figure 1. In such cases the population may actually go to a different peak than that which has the *steepest gradient* on the adaptive topography. This complication is in addition to the fact that the population does not necessarily ascend the *highest* peak, since gene frequencies only move to points of *local* maxima of mean adaptive values. This latter point can be seen in the trajectory of figure 1. Depending upon initial conditions, the populations may go to either peak or toward the saddle.

THE COMPLICATION OF LINKAGE

When the assumption of constant linkage equilibrium is relaxed, which is necessary for an accurate treatment of the problem, it is gametic rather than gene frequencies that must be examined. Using the notations developed in the previous sections, let $g_{kl}^{(t)}$, $k = 1$ or 0 and $l = 1$ or 0, be the frequency of a particular gametic type in the t-th generation immediately after meiosis. Then $Z_{ij}^{(t)}$ will be the frequency of a zygote formed from the gametes $g_{kl}^{(t)}$'s, while $g_{kl}^{(t+1)}$ denotes the frequency of a particular gamete formed by meiosis of the zygotes in the t-th generation. In table 4 the relations between the $Z_{ij}^{(t)}$ and the $g_{kl}^{(t)}$ are given in the second and third columns for a random mating population. The relative frequencies of these ten genotypes after selection are given in the fifth column. After selection, meiosis occurs again to produce the gametes of the $(t + 1)$-th generation The gametic frequencies, $g_{kl}^{(t+1)}$, are easily computed with the aid of the last column of table 4. The gametic frequencies in the $(t + 1)$-th generation are then:

$$g_{11}^{(t+1)} = \frac{g_{11}^t g_{11}^t W_{22} + \frac{1}{2}(2g_{11}^t g_{10}^t W_{21} + 2g_{11}^t g_{01}^t W_{12}) + [(1-R)g_{11}^t g_{00}^t + Rg_{10}^t g_{01}^t]W_{11}}{\bar{W}} \tag{10.a}$$

$$g_{10}^{(t+1)} = \frac{g_{10}^t g_{10}^t W_{20} + \frac{1}{2}(2g_{11}^t g_{10}^t W_{21} + 2g_{10}^t g_{00}^t W_{10}) + [Rg_{11}^t g_{00}^t + (1-R)g_{10}^t g_{01}^t]W_{1.}}{\bar{W}} \tag{10.b}$$

$$g_{01}^{(t+1)} = \frac{g_{01}^t g_{01}^t W_{02} + \frac{1}{2}(2g_{01}^t g_{11}^t W_{12} + 2g_{01}^t g_{00}^t W_{01}) + [Rg_{11}^t g_{00}^t + (1-R)g_{10}^t g_{01}^t]W_{11}}{\bar{W}} \tag{10.c}$$

$$g_{00}^{(t+1)} = \frac{g_{00}^t g_{00}^t W_{00} + \frac{1}{2}(2g_{00}^t g_{10}^t W_{10} + 2g_{00}^t g_{01}^t W_{01}) + [(1-R)g_{11}^t g_{00}^t + Rg_{10}^t g_{01}^t]W_{11}}{\bar{W}}. \tag{10.d}$$

TABLE 4. *Relations between gametic and genotypic frequencies in a random mating population*

Genotype	Zygotic frequency	Composition	Adaptive value	Frequency after selection	Gametes produced
$\frac{AB}{AB}$	$Z_{22}^{(t)}$	$g_{11}^{(t)}g_{11}^{(t)}$	W_{22}	$g_{11}^{(t)}g_{11}^{(t)}W_{22}/\bar{W}$	$1AB$
$\frac{AB}{Ab}$	$Z_{21}^{(t)}$	$2g_{11}^{(t)}g_{10}^{(t)}$	W_{21}	$2g_{11}^{(t)}g_{10}^{(t)}W_{21}/\bar{W}$	$\frac{1}{2}AB$ $\frac{1}{2}Ab$
$\frac{Ab}{Ab}$	$Z_{20}^{(t)}$	$g_{10}^{(t)}g_{10}^{(t)}$	W_{20}	$g_{10}^{(t)}g_{10}^{(t)}W_{20}/\bar{W}$	$1Ab$
$\frac{AB}{aB}$	$Z_{12}^{(t)}$	$2g_{11}^{(t)}g_{01}^{(t)}$	W_{12}	$2g_{11}^{(t)}g_{01}^{(t)}W_{12}/\bar{W}$	$\frac{1}{2}AB$ $\frac{1}{2}aB$
$\frac{AB}{ab}$	$Z_{11}^{(t)}$	$2g_{11}^{(t)}g_{00}^{(t)}$	W_{11}	$2g_{11}^{(t)}g_{10}^{(t)}W_{11}/\bar{W}$	$\frac{1}{2}(1-R)AB$, $\frac{1}{2}R\ Ab$ $\frac{1}{2}R\ aB$, $\frac{1}{2}(1-R)ab$
$\frac{Ab}{bB}$		$2g_{10}^{(t)}g_{01}^{(t)}$	W_{11}	$2g_{10}^{(t)}g_{01}^{(t)}W_{11}/\bar{W}$	$\frac{1}{2}R\ AB$, $\frac{1}{2}(1-R)Ab$ $\frac{1}{2}(1-R)aB$, $\frac{1}{2}R\ ab$
$\frac{Ab}{ab}$	$Z_{10}^{(t)}$	$2g_{10}^{(t)}g_{00}^{(t)}$	W_{10}	$2g_{10}^{(t)}g_{00}^{(t)}W_{10}/\bar{W}$	$\frac{1}{2}Ab$ $\frac{1}{2}ab$
$\frac{aB}{aB}$	$Z_{02}^{(t)}$	$g_{01}^{(t)}g_{01}^{(t)}$	W_{02}	$g_{01}^{(t)}g_{01}^{(t)}W_{02}/\bar{W}$	$1aB$
$\frac{aB}{ab}$	$Z_{01}^{(t)}$	$2g_{01}^{(t)}g_{00}^{(t)}$	W_{01}	$2g_{01}^{(t)}g_{00}^{(t)}W_{01}/\bar{W}$	$\frac{1}{2}aB$ $\frac{1}{2}ab$
$\frac{ab}{ab}$	$Z_{00}^{(t)}$	$g_{00}^{(t)}g_{00}^{(t)}$	W_{00}	$g_{00}^{(t)}g_{00}^{(t)}W_{00}/\bar{W}$	$1ab$

Making the convention that W_{AB} is the average adaptive value of gametic phase AB in all its combinations, i.e., *the the marginal mean* of the gamete AB, and similarly W_{aB}, W_{Ab} and W_{ab} for AB, Ab and ab, respectively, the four equations from (10.a) to (10.d) can be written in the form

$$g_{11}^{(t+1)} = \frac{g_{11}^{t}W_{AB} - RW_{11}D^{t}}{\bar{W}} \quad (11.a)$$

$$g_{10}^{(t+1)} = \frac{g_{10}^{t}W_{Ab} + RW_{11}D^{t}}{\bar{W}} \quad (11.b)$$

$$g_{01}^{(t+1)} = \frac{g_{01}^{t}W_{aB} + RW_{11}D^{t}}{\bar{W}} \quad (11.c)$$

$$g_{00}^{(t+1)} = \frac{g_{00}^{t}W_{ab} - RW_{11}D^{t}}{\bar{W}}. \quad (11.d)$$

Finally, the changes in gametic frequencies in each generation, Δg_{kl}, are expressed as follows (dropping the superscripts),

$$\Delta g_{11} = \frac{g_{11}(W_{AB} - \bar{W}) - RW_{11}D}{\bar{W}} \quad (12.a)$$

$$\Delta g_{10} = \frac{g_{10}(W_{Ab} - \bar{W}) + RW_{11}D}{\bar{W}} \quad (12.b)$$

$$\Delta g_{01} = \frac{g_{01}(W_{aB} - \bar{W}) + RW_{11}D}{\bar{W}} \quad (12.c)$$

$$\Delta g_{00} = \frac{g_{00}(W_{ab} - \bar{W}) - RW_{11}D}{\bar{W}}. \quad (12.d)$$

Equations (12.a)–(12.d) are equivalent to those derived by Kimura (1956) in his treatment of a special case of inter-

action between two genes. His derivation was based on a time-continuous model (instead of generation-time) with instantaneous rates of birth and death.

The Equilibrium Condition

First, to be examined is the question of whether the complication of linkage disturbs the simple equilibrium picture expected with the assumption of constant linkage equilibrium. Specifically, do the gene frequencies come to the same equilibrium values, irrespective of linkage, and does the population reach linkage equilibrium as it approaches equilibrium?

An equilibrium of the population in the present investigation is defined by

$$\Delta g_{11} = \Delta g_{10} = \Delta g_{01} = \Delta g_{00} = 0. \quad (13)$$

Under this definition not only do the frequencies of gametes and zygotes not change but also the value of D must be a constant. Substituting (13) in equations (12.a)–(12.d), yields

$$g_{11}(W_{AB} - \bar{W}) - RDW_{11} = 0 \quad (14.a)$$

$$g_{10}(W_{ab} - \bar{W}) + RDW_{11} = 0 \quad (14.b)$$

$$g_{01}(W_{ab} - \bar{W}) + RDW_{11} = 0 \quad (14.c)$$

$$g_{00}(W_{ab} - \bar{W}) - RDW_{11} = 0. \quad (14.d)$$

These deceptively simple equations are not easy to solve in general since they are cubic equations in the g_{kl}'s. By examining some symmetrical cases of adaptive values, however, the equations can be rendered simple enough to solve. Specifically, we have used adaptive values given in table 5 where a, b, c, and d can, in general, take any value. This model allows epistasis when

$$a - b - c + d \neq 0.$$

TABLE 5. *General scheme of adaptive values for the symmetrical case*

	AA	*Aa*	*aa*
BB	*a*	*b*	*a*
Bb	*c*	*d*	*c*
bb	*a*	*b*	*a*

In this model the generality is restricted, but not to a great extent when one quadrat in the gene frequency plane, say p and r both being from 1.00 to .50, is considered. Substituting the adaptive values in table 5 into the equations, (14.a)–(14.d), and writing them *in extenso* gives:

$$g_{11}[ag_{11}+cg_{10}+bg_{01}+dg_{00}-\bar{W}] - Rd[g_{11}g_{00}-g_{10}g_{01}]=0 \quad (15.a)$$

$$g_{10}[cg_{11}+ag_{10}+dg_{01}+bg_{00}-\bar{W}] + Rd[g_{11}g_{00}-g_{10}g_{01}]=0 \quad (15.b)$$

$$g_{01}[bg_{11}+dg_{10}+ag_{01}+cg_{00}-\bar{W}] + Rd[g_{11}g_{00}-g_{10}g_{01}]=0 \quad (15.c)$$

$$g_{00}[dg_{11}+bg_{10}+cg_{01}+ag_{00}-\bar{W}] - Rd[g_{11}g_{00}-g_{10}g_{01}]=0. \quad (15.d)$$

Because of the symmetry of the equations, (15.a) with (15.d) and (15.b) with (15.c), the solutions require that $g_{11} = g_{00}$ and $g_{10} = g_{01} = \frac{1}{2} - g_{11}$. Making the appropriate substitutions in the equations, (15.a), and noting that

$$\bar{W}=4(a+d-b-c)g_{11}^2 - 2(a+d-b-c)g_{11}+\frac{a+d}{2} \quad (16)$$

it is found that

$$g_{11}(b+c-d-a)(4g_{11}^2-3g_{11}+\tfrac{1}{2}) - Rd(g_{11}-\tfrac{1}{4})=0. \quad (17)$$

In the following, two cases, epistatic and non-epistatic, are to be examined.

Case I: *Additivity between loci*

When $a + d - c - b = 0$, there is no epistasis (see, e.g. Kojima, 1959). If the adaptive values satisfy this condition, then the solution is

$$g_{11} = g_{10} = g_{01} = g_{00} = \tfrac{1}{4}.$$

That is, there is an equilibrium at $p = r = \frac{1}{2}$ and there is no linkage disequilibrium since

$$D = (g_{11}g_{00} - g_{10}g_{01}) = 0$$

The stability of the equilibrium now depends simply on whether the value of

d is the largest among the four adaptive values or not. If it is, then the point, $p = r = \frac{1}{2}$, is a simple overdominant equilibrium.

Case II: *Interaction between loci—epistasis*

If $a + d - b - c \neq 0$, then the equation, (17), is a cubic equation with three distinct solutions. They are:

$$g_{11} = \tfrac{1}{4} + \tfrac{1}{4}\sqrt{1 + \frac{4Rd}{b + c - a - d}} \quad (18.a)$$

$$g_{11} = \tfrac{1}{4} - \tfrac{1}{4}\sqrt{1 + \frac{4Rd}{b + c - a - d}} \quad (18.b)$$

$$g_{11} = \tfrac{1}{4}. \quad (18.c)$$

Since $g_{11} = g_{00} = \frac{1}{2} - g_{01} = \frac{1}{2} - g_{10}$, the gene frequencies of both loci, p and r, are again $\frac{1}{2}$ at equilibrium. But if (18.a) or (18.b) should be the solution, there would *no longer be linkage equilibrium*, because

$$D = g_{11}g_{00} - g_{10}g_{01}$$
$$= \pm \tfrac{1}{4}\sqrt{1 + \frac{4Rd}{b + c - d - a}}. \quad (19)$$

In order for (18.a) and (18.b) to hold true, the quantity under the radical sign must be between zero and one, inclusive. This requires that $b + c - a - d$ be negative and that

$$R \leq \frac{a + d - b - c}{4d}. \quad (20)$$

That is, linkage must be tighter than the value on the right-hand side of (20), or else the system will go to linkage equilibrium given in (18.c). For example, using values in Table 3, $a = 1$, $b = 1.5$, $c = 2.5$, $d = 4$, R must then be less than .0625 for permanent linkage disequilibrium. When $R = .0625$, three solutions, (18.a)–(18.c) are equal.

Finally, there is the possibility that (18.a) and/or (18.b) represent unstable equilibrium and only (18.c) is stable, or the reverse. The stability of equilibria in this case can be tested by regarding the four gametic types as four alleles of a single super-locus. The rules for testing stability of this type of equilibria are given by Kimura (1956). The algebra is extensive but straightforward and will not be reproduced here in any detail.

We simply point out that for (18.c) to be a stable equilibrium it is necessary for the matrix

$$\begin{vmatrix} 2(a-d) & (a+c-b-d) & (a+b-c-d) \\ (a+c-b-d) & 2(a-b)-4Rd & (a+d-b-c)-4Rd \\ (a+b-c-d) & (a+d-b-c) & 2(a-c)-4Rd \end{vmatrix}$$

to be negative definite. This turns out to be equivalent to requiring that

$$R > \frac{a + d - b - c}{4d} \quad (21.a)$$

and

$$|a - d| > |b - c| \quad \text{and} \quad d > a. \quad (21.b)$$

That is to say, these conditions are necessary and sufficient for an equilibrium at $g_{11} = g_{10} = g_{01} = g_{00} = \frac{1}{4}$ to be stable. Since the condition (21.a) is the converse of the inequality, (20), linkage equilibrium at $p = r = \frac{1}{2}$ for the model given in Table 5 will represent a stable equilibrium of the population if and only if (21.a) and (21.b) are true. The condition, (21.a) assures that the population tends to linkage equilibrium as it approaches gene frequency equilibrium. The stability of such equilibria was studied by Kojima (1959): the necessary and sufficient conditions for such equilibria to be stable are (1) overdominance on the marginal means of three zygotic phases at each locus and (2) additive × additive epistatic variance being smaller than dominance variance. In the present model, the condition, (21.b), turns out to be the same as overdomi-

nance, i.e., $\bar{W}_{Aa} > \bar{W}_{AA}$, $\bar{W}_{Aa} > \bar{W}_{aa}$, $\bar{W}_{Bb} > \bar{W}_{BB}$ and $\bar{W}_{Bb} > \bar{W}_{bb}$ (see Table 1). The second condition on the magnitude of additive × additive epistasis is also satisfied in this case, since the additive × additive variance is zero at $g_{11} = g_{10} = g_{01} = g_{00} = \frac{1}{4}$.

When the inequality, (20), is true, linkage may be so tight that linkage equilibrium would not be achieved in the population. If there is an equilibrium under this condition, the solutions of gametic frequencies in (18.a) and/or (18.b) represent those at equilibrium. The amount of linkage disequilibrium in such a case is given in (19). It is extremely difficult to obtain the *critical upper bound of R* with which populations tend to a stable equilibrium without linkage equilibrium, although, when the values of fitness and recombination fractions are given, the stability of such equilibria is readily tested.

Consider the adaptive values of table 3. It has already been found that with recombination fractions greater than .0625, the population tends to a stable equilibrium at $g_{11} = g_{10} = g_{01} = g_{00} = \frac{1}{4}$. If $R = .05$, then

$$\hat{g}_{11} = \hat{g}_{00} = \tfrac{1}{4} \pm \tfrac{1}{4}\sqrt{1-16R} = .362 \text{ or } .138$$

$$\hat{g}_{01} = \hat{g}_{10} = \tfrac{1}{4} \mp \tfrac{1}{4}\sqrt{1-16R} = .138 \text{ or } .362$$

$$D = \pm\tfrac{1}{4}\sqrt{1-16R} = \pm .112.$$

Both of these equilibria turn out to be stable. When R tends to zero, the disequilibrium tends increasingly to $\pm\frac{1}{4}$ which is the maximum value for D. Using Table 1 it can be shown that at equilibrium

$$\bar{W}_{Aa} > \bar{W}_{AA} \quad \text{and} \quad \bar{W}_{aa}$$

and

$$\bar{W}_{Bb} > \bar{W}_{BB} \quad \text{and} \quad \bar{W}_{bb}.$$

The next example is a model with $a = 3$, $b = 4$, $c = 1$, and $d = 5$ which shows epistasis. In this case there is *no* stable equilibrium with linkage equilibrium since these adaptive values violate the condition (21.b). The converse of the condition, (21.a), i.e.,

$$R \leq \frac{a + d - b - c}{4d}$$

becomes

$$R \leq .15.$$

However, the numerical computation shows that equilibria are not stable unless R is less than approximately .116. In other words, the model under present consideration does not lead to any stable equilibrium unless linkage is tighter than .116. Computing $\bar{W}_{AA}$, $\bar{W}_{Aa}$, etc. appearing in table 1, it is found that these marginal means satisfy overdominance at such stable equilibria. Findings from this and several other examples suggest that over-dominance at equilibrium is necessary for populations to be maintained in a stable equilibrium, just as in the case of the epistatic model in linkage equilibrium (Kojima, 1959).

The effect of the linkage disequilibrium on the mean adaptive value can be judged from the expression, (16), which can be rewritten as

$$\bar{W} = 4D^2(a+d-b-c) + \frac{a+b+c+d}{4}. \quad (22)$$

At equilibrium the difference in $\bar{W}$ with and without linkage disequilibrium is

$$4D^2(a + d - b - c)$$

which is not large unless the epistatic deviation measured by

$$(a + d - b - c)$$

is extremely large. Moreover, the size of $\bar{W}$ does not depend on the sign of D. That is, the excess of coupling or repulsion phase in the case of symmetrical adaptive values given in table 5 is immaterial.

It might be supposed that results so far presented are restricted to the relatively simple symmetrical set of adaptive values shown in table 5. This set was chosen as an illustration because analytic

TABLE 6. *Hypothetical fitness values for a partially asymmetrical case showing the same degree of epistasis as the symmetrical case in Table 3*

	AA	*Aa*	*aa*
BB	1	1.5	1
Bb	2.5	4	2.5
bb	2	2.5	2

solutions to the equilibrium equations are possible for a symmetrical case. For an asymmetrical case although analytic solutions are not possible, numerical solutions can be found, and we have included a case of this sort to show the generality of our result. Table 6 gives the adaptive values for a case in which there is symmetry with respect to the A locus but asymmetry with respect to the B locus. The epistatic deviations are identical with these in table 3. The results of numerical calculation for this new case are given in table 7. As the table shows, not only are there profound effects of linkage on the gametic frequencies but in this case there are also marked effects or *gene frequencies*. This latter phenomenon did not appear in the symmetrical case. Again, however, rather tight linkage must be assumed before there is any departure from the expected result with no linkage complication. The critical upper bound for R cannot be given analytically but it appears to be close to if not equal to .0625 as in the previous case. For R as large as .08 the stable equilibrium condition is that predicted if linkage is ignored. The smaller the value of recombination the greater is the departure from this expectation both in gametic and gene frequencies and as in the previous examples two stable, conjugate, equilibria exist for each condition. This occurrence of paired solutions is a result of the symmetry of the A locus since in a sense, A and a are indistinguishable.

THE APPROACH TO EQUILIBRIUM

Although a fairly marked epistasis associated with a tight linkage is required to disturb linkage equilibrium and gene frequency equilibrium, the approach to the equilibrium will be affected by any values of linkage and of fitness. To investigate these effects, the equations, (12.a)–(12.d), are put in their approximate differential forms:

$$\frac{dg_{11}}{dt} = g_{11}(W_{AB} - \bar{W}) - RDW_{11} \quad (23.a)$$

$$\frac{dg_{10}}{dt} = g_{10}(W_{Ab} - \bar{W}) + RDW_{11} \quad (23.b)$$

$$\frac{dg_{01}}{dt} = g_{01}(W_{aB} - \bar{W}) + RDW_{11} \quad (23.c)$$

TABLE 7. *Equilibrium values of gametic frequencies, gene frequencies, adaptive value and linkage disequilibrium for the fitnesses shown in table 6. R is the value of recombination*

R	g_{11}	g_{10}	g_{01}	g_{00}	p	r	$\bar{W}$	D
.01	.014	.576	.376	.034	.590	.390	2.762	−.216
	.376	.034	.014	.576	.410	.390	2.762	+.216
.02	.030	.549	.351	.071	.579	.381	2.724	−.190
	.351	.071	.030	.549	.421	.381	2.724	+.190
.04	.068	.482	.289	.161	.550	.357	2.652	−.128
	.289	.161	.068	.482	.450	.357	2.652	+.128
.062	.156	.348	.179	.317	.504	.335	2.584	−.013
	.179	.317	.156	.348	.496	.335	2.584	+.013
.08	.167	.333	.167	.333	.500	.333	2.582	.000

$$\frac{dg_{00}}{dt} = g_{00}(W_{ab} - \bar{W}) - RDW_{11}. \quad (23.d)$$

The factor of time can then be eliminated by dividing (23.a), (23.b) and (23.c) by (23.d), yielding

$$\frac{dg_{11}}{dg_{00}} = \frac{g_{11}(W_{AB} - \bar{W}) + RDW_{11}}{g_{00}(W_{ab} - \bar{W}) + RDW_{11}} \quad (24.a)$$

$$\frac{dg_{10}}{dg_{01}} = \frac{g_{10}(W_{Ab} - \bar{W}) - RDW_{11}}{g_{00}(W_{ab} - \bar{W}) + RDW_{11}} \quad (24.b)$$

$$\frac{dg_{01}}{dg_{00}} = \frac{g_{01}(W_{aB} - \bar{W}) - RDW_{11}}{g_{00}(W_{ab} - \bar{W}) + RDW_{11}}. \quad (24.c)$$

When these equations are written in their extensive forms in terms of adaptive values of genotypes and solved simultaneously, the trajectories of gametic frequency changes can be plotted, and these, in turn, can be put in terms of gene frequencies by adding the appropriate gametic frequencies. The simultaneous equations, (24.a), (24.b) and (24.c), are not easy to solve analytically, so numerical solutions for several cases have been obtained by using a combination of the Runge-Kutta and Milne methods adapted to digital computers. Specifically, we have used the two sets of adaptive values given in tables 2 and 3 in combination with different recombination fractions. The results are shown in figures 4 and 5. The trajectories under the assumption of linkage equilibrium are included for comparison. In each case the initial gametic frequencies were such that the populations started in linkage equilibrium.

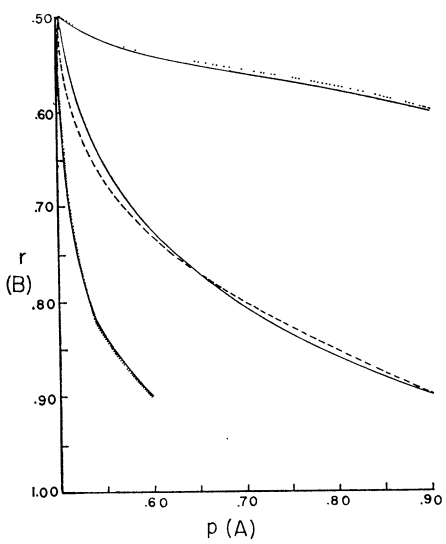

FIG. 4. Trajectories of gene frequencies under the assumption of constant linkage equilibrium (solid lines) compared with true trajectories for $R = .1250$ (dashed lines) and $R = .0000$ (dotted lines). Non-epistatic case of table 2.

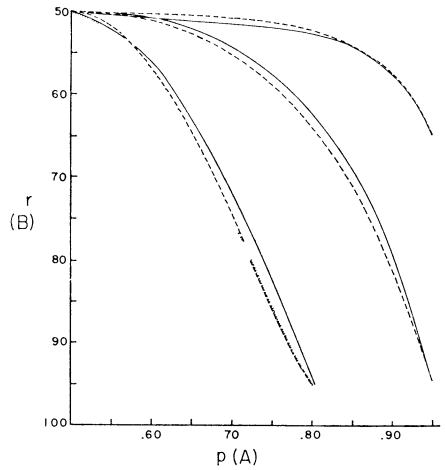

FIG. 5. Trajectories of gene frequency under the assumption of constant linkage equilibrium (solid lines) compared with true trajectories for $R = .1250$ (dashed lines) and $R = .0000$ (dotted lines). Epistatic case of table 3.

Any deviation from the simpler case is then due to an accumulation of linkages during the process of selection. As the curves show, there is a real, but small, difference between the curves computed with and without the linkage equilibrium assumption. Again it is borne out that there may not be a great effect on gene frequencies due to linkage.

There is, however, a profound effect on gametic frequencies. Figure 6 shows this effect clearly. It is a plot of the values of D against the changing frequency of the allele, A, for the epistatic case with initial conditions $p = .80$ and $r = .95$. The largest value that D can

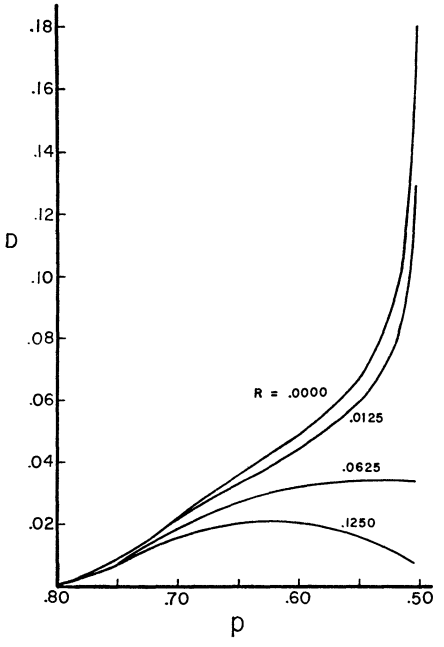

FIG. 6. Changes in the value of D, the linkage disequilibrium, during the process of gene frequency change. The trajectory is the lowermost one in figure 5 (epistatic case). Ordinate shows values of D corresponding to gene frequency, p, of allele A on the abscissa.

ever attain is $\pm.25$ which would correspond to only coupling or only repulsion gametes. As the figure shows, very tight linkages result in a constantly increasing value of D up to the limit of .25. When $R = .0625$ which separates equilibria with and without linkage disequilibrium, D goes as high as .036. For $R = .125$ a maximum is reached at $D = .02$ and then D declines toward zero at gene frequency equilibrium. No points are given at $p = .50$, because the method of solution becomes unstable at the point where the derivatives do not exist. For this reason the curves for $R = .0625$ and $R = .125$ are not shown decreasing to zero although they do come down sharply toward zero near $p = .50$.

A qualitatively similar picture is obtained for the additive case except that even with extremely tight linkage the value of D never exceeds 10^{-7}. This suggests that linkage among non-interacting genes does not build up disequilibrium of linkage.

Whether D will be positive or nega-

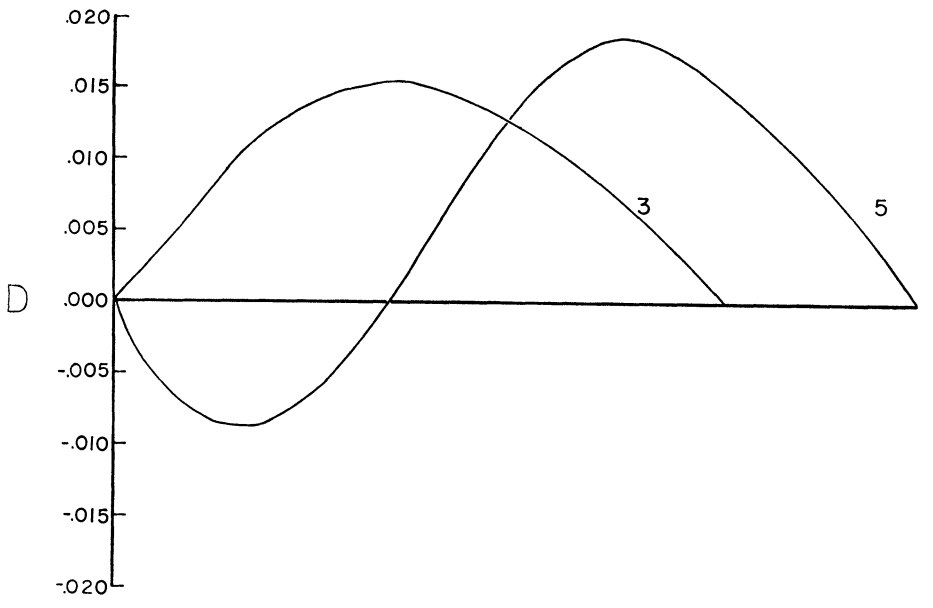

FIG. 7. Values of the linkage disequilibrium, D, on the ordinate, plotted against distance along the trajectory in arbitrary units for trajectories 3 and 5 of figure 1. $R = .50$.

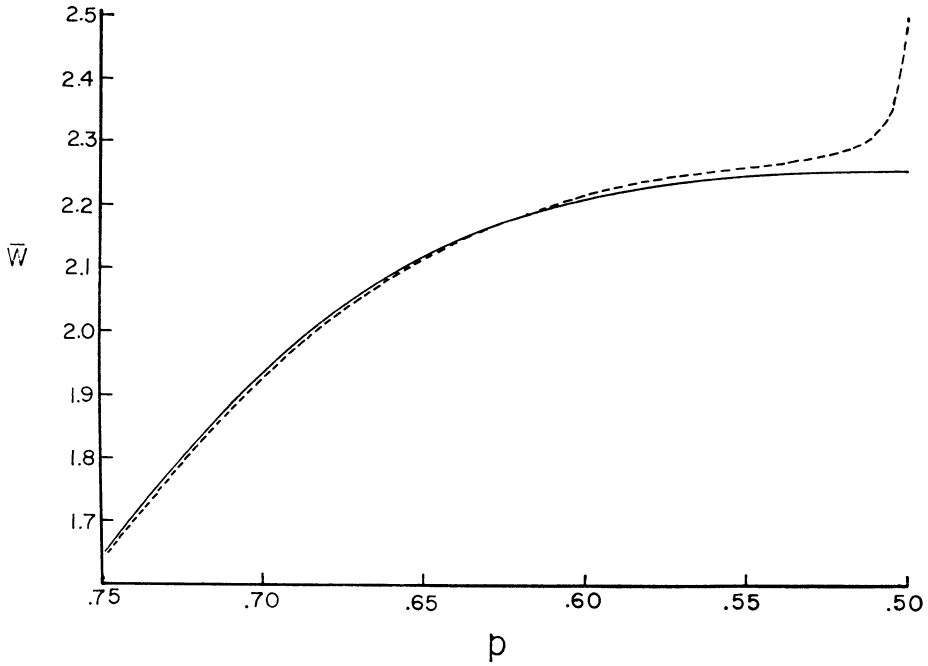

FIG. 8. Values of the mean adaptive value, $\bar{W}$, (ordinate) plotted against successive gene frequencies of allele A for the lowermost trajectory of figure 5 (epistatic case). Solid line is $\bar{W}$ for linkage equilibrium, broken line is $\bar{W}$ for R = .0000.

tive, and whether increasing or decreasing in any generation will depend upon the instantaneous value of D, the recombination fraction, the adaptive values of genotypes and the particular path on the adaptive surface which is being traversed by the gene frequencies. As an example of the complexity of changes possible in D, figure 7, has been drawn. Figure 1 shows a variety of evolutionary trajectories computed for the complex adaptive surface. Corresponding to trajectory 3 and 5 in figure 1, a plot of changes in D has been made in figure 7. On the ordinate of figure 7 are the values of D and on the abscissa are values of the distance along the trajectory from the starting point. The most interesting of these graphs is that for trajectory 5 in which D becomes progressively negative, reaches a minimum value, then rises, becoming positive, to a maximum value and finally declines to zero.

Lastly the values of $\bar{W}$ may be compared between different linkage values. As figure 8 shows, there is virtually no difference in $\bar{W}$ along the trajectory between tight and loose linkages except very near to equilibrium. At this point there is a sudden rise in mean adaptive value in the tightly linked case, but again this amounts to no more than 10%.

GENERAL CONCLUSIONS

It must be said that as a general rule joint effects of linkage and epistasis do not produce serious changes of population structure except under special circumstances. These circumstances are the simultaneous existence of marked epistasis and tight linkage. For example, in order for there to be any important effect of linkage and epistasis among genes on different chromosomes (R = .50); the epistatic deviation, $a + d - b - c$, must be greater than twice the fitness of the double heterozygotes, d. This is equivalent to demanding that the fitness, a, of the four double homo-

zygotes be slightly greater than that of the double heterozygote, while the fitness of the four single homozygotes be virtually zero, provided that fitnesses, a, b, c and d are all positive. This would result in four steep adaptive peaks at the four corners of the adaptive landscape and one rather shallow peak at the center.

There does, in fact, seem to be a case of this extreme type of epistasis in nature. Levitan and Salzano (1959) have described two linked systems of inversions, E and H, on the fourth chromosome of *D. guaramunú*. Although there are some ambiguities in classification (HH and hh homozygotes are indistinguishable), the data show a great excess of double heterozygotes and double homozygotes with a virtual absence of the four single homozygotes. Data on crossing-over between the E and H systems are not available directly, but linkage seems to be tight. This is then a combination of moderately tight or tight linkage and marked epistasis which produces a most aberrant zygotic array.

When there is tight linkage with the more moderate degree of epistasis, the main effect is on gametic rather than gene frequencies. However, the importance of this effect should not be underestimated because it is the frequencies of the gametes that determine the zygotic frequencies and thus the phenotypic composition of the population. Thus, as in the two epistatic models discussed previously, one can expect a considerable change in population structure when tight linkage is introduced. In the first epistatic model, the frequency of the genotype, $AABB$, would be .0625 at equilibrium if there were no tight linkage, while it is .131 or more than twice as great if recombination fraction is .05. In the second epistatic model, there would *not be any stable equilibrium* and the population would tend to fixation if there were no linkage complication. Gene frequencies alone do not give an adequate picture of the genotypic constitution of the population in such cases and attention must be paid to the gametic frequencies.

An example of permanent linkage disequilibrium in a naturally occurring polymorphism comes from the work of Levitan (1958) on the X chromosome inversions in *Drosophila robusta*. Each arm of the chromosome contains an inverted sequence, I and a normal sequence S. Crossing over between sequences on the left arm and those on the right arm is infrequent, about one percent. Data on gametic frequencies in males and females are as follows:

	♂	♀
S_LS_R	178	266
S_LI_R	369	587
I_LS_R	176	259
I_LI_R	538	912

These data give values of D♂ = .019 and D♀ = .022. Such values would not require any extraordinary epistasis to maintain them considering the tightness of linkage coupled with the lack of crossing-over in males. It is difficult to see how such an explanation could be invoked for Levitan's second-chromosome data, however, in which the association observed is not consistently in favor of coupling or repulsion gametes.

To what extent linkage and epistasis are important in nature remains a question, but there is abundant evidence that restriction of crossing-over is a common evolutionary mode. Inversions, translocations, localization of chiasmata and restriction of recombination to one sex are common features in natural populations. When these mechanisms reduce recombination to a low degree, a consideration of the linkage-epistatic interaction becomes important in the study of the evolution of these populations.

Summary

The joint effects of linkage and epistasis (interaction between non-allelic genes in determining fitness) have been

examined for two-locus polymorphisms. The general results are of the following nature:

(1) Gene frequencies change toward a stable equilibrium condition which corresponds to a local maximum in mean adaptive value. This change occurs in such a way that the rate of increase of mean adaptive value is not maximum. That is, the trajectory of gene frequency changes is not the "steepest path" on the adaptive surface.

(2) In the absence of epistasis, linkage does not affect the final equilibrium of the population.

(3) When epistasis is present, linkage must be fairly tight in order for there to be any effect on the final equilibrium. The amount of recombination allowed for such cases is, in general, a function of epistatic deviations.

(4) If linkage is tighter than the value demanded by the magnitude of epistatic deviations, there may be *permanent linkage disequilibrium* of considerable magnitude, and the *gene frequencies* may also be affected.

(5) There are some cases where a stable equilibrium is possible only with a tight linkage. Without such linkage complication, there will be no intermediate gene frequency stable equilibrium.

(6) The approach to the equilibrium condition is affected by linkage irrespective of whether there is epistasis although the effect is greater when epistasis is present.

Literature Cited

Cain, A. S. and P. M. Sheppard. 1952. The effects of natural selection on body color in the land snail *Cepaea nemoralis*. Heredity, **6**: 217–231.

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

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

Kojima, K. 1959. Role of epistasis and overdominance in stability of equilibria with selection. Proc. Nat. Acad. Sci., **45**: 984–989.

Lamotte, M. 1951. Recherches sur la structure génétique des populations naturelles de *Cepaea nemoralis* L. Bull. Biol. France et Belg., Suppl., **35**: 1–239.

Levitan, M. 1955. Studies in linkage in populations. I. Association of second chromosome linkages in *Drosophila robusta*. Evolution, **9**: 27–41.

——. 1958. Non-random association of inversions. Cold Spring Harb. Symp. Quant. Biol., **23**: 251–268.

—— and F. M. Salzano. 1959. Studies of linkage in populations. III. An association of linked inversions in *Drosophila guarumunu*. Heredity **13**: 243–248.

Lewontin, R. C. 1958. A general method for investigating the equilibrium of gene frequency in a population. Genetics, **43**: 419–434.

—— and M. J. D. White. 1960. Interaction between inversion polymorphisms of two chromosome pairs in the grasshopper *Moraba scurra*. Evolution, **14**: 116–129.

Sheppard, P. M. 1959. The evolution of mimicry; a problem in ecology and genetics. Cold Spring Harb. Symp. Quant. Biol., **24**: 131–140.

White, M. J. D. 1957. Cytogenetics of the grasshopper, *Moraba scurra*. II. Heterotic systems and their interaction. Austral. J. Zool., **5**: 305–337.

Wright, S. 1932. The roles of mutation, inbreeding and selection in evolution. Proc. Sixth Int. Cong. Genetics, pp. 356–366.

——. 1942. Statistical genetics and evolution. Bull. Am. Math. Soc., **48**: 223–246.

——. 1949. Adaptation and selection pp. 365–389 in "Genetics, Paleontology and Evolution." Ed. by G. L. Jepson. Princeton Univ. Press. N. J.

——. 1955. Classification of the factors of evolution. Cold Spring Harb. Symp. Quant. Biol., **20**: 16–24D.