THE GENETICS OF COMPLEX SYSTEMS

R. C. LEWONTIN
UNIVERSITY OF CHICAGO

1. Introduction

In recent years, among theoretical geneticists, there has been a flourishing of interest in the exact treatment of complex genetic systems. By complex, I mean systems in which more than one genetic locus is varying and in which the intensity of natural selection operating on a genotype is a function of the interaction of physiological forces among the several loci. Obviously, if we can specify the relative fitnesses of different genotypes at each locus separately without reference to the state of other loci, then the usual single locus formulations are adequate. If the relative fitnesses are not so unambiguously defined, however, the dimensionality of our problem increases and with it the complexity of the various solutions.

When very large populations are considered, and if environments are assumed to be essentially constant over reasonably long periods of generations, then the genetic change of a population, caused by differential fitness of various genotypes, can be analogized to the movement of a point on a potential surface. At any time the population state is given by a point in $(n + 1)$ dimensional space, in which n dimensions are the n variables needed to specify the genetic composition of the population and the $(n + 1)$st variable is the *mean fitness* or potential function of the population. Looked at in this way, there are three problems to be attacked. First, what is the minimum number of dimensions sufficient to describe and predict population change and composition? For example, if we consider two gene loci, each with two alternate alleles, will two dimensions, the frequency of one allele at each locus, be sufficient (in addition to the extra dimension of mean fitness)? Or will it be necessary to have three dimensions to describe the genetic composition, corresponding to the gametic types AB, Ab and aB (the frequency of the fourth type ab is fixed by the other three)?

That is, is the genetic composition of a population sufficiently specified by knowing the frequencies of alleles at each locus separately or do we need to know the joint distribution of three alleles at different loci? In the former case, the population composition can be represented by a point in a hypercube of $\ell(a - 1)$ dimensions, while in the latter case we require a hypertetrahedron of $(a^\ell - 1)$ dimensions, where ℓ is the number of gene loci and a is the number of allele per locus, assuming a to be the same for all loci.

The second question is what are the kinetics of the process of genetic change in time. What path on the n dimensional hypersurface, the fitness surface, will

the population take? Third, we must inquire into the static or equilibrium configuration of the population. What kinds of singularities does the fitness surface have, what are the maxima, minima, minimaxes and at which, if any of these, does the population have stable equilibrium state of genetic composition?

These three questions are obviously closely related, and it is the purpose of this paper to summarize our present understanding of these problems at the theoretical level.

2. The basic model

Let us assume that a given genetic locus has n alleles a_i, each with frequency x_i in some generation t. Further, let us assume an infinite population mating at random so that the frequency of any diploid genotype a_ia_j among fertilized eggs is x_ix_j ($a_ia_j \equiv a_ja_i$). Associated with each genotype a_ia_j is a fitness W_{ij} ($=W_{ji}$) which is the probability that an egg of that genotype produces a progeny egg in the next generation. Defining the mean fitness of the population $\overline{W}$, by

$$\overline{W} = \sum_{ij} x_ix_jW_{ij}, \tag{2.1}$$

it can be shown that the change in gene frequency in one generation as a result of differential fitness is given by [16]

$$\Delta x_i = \frac{x_i(1 - x_i)}{2\overline{W}} \frac{\partial \overline{W}}{\partial x_i}. \tag{2.2}$$

Equation (2.2) shows clearly that $\overline{W}$ is a potential function and that Δx_i will be zero for all i if

$$x_i = 0 \text{ for some subset of } i, \tag{2.3}$$

and

$$\frac{\partial \overline{W}}{\partial x_i} = 0 \text{ for all } i \text{ not in } I. \tag{2.4}$$

Condition (2.4) simply describes singularities in the $\overline{W}$ surface. Because of the definition of $\overline{W}$ given by (2.1), the set of equations (2.4) has at most one solution strictly inside the unit tetrahedron. This means that there is at most one equilibrium value at which any particular subset of alleles a_i is present in the population. If such an equilibrium value exists, it may not be stable; but if it is stable it corresponds to a maximum on the $\overline{W}$ hypersurface [4], [16].

For any particular set of fitnesses W_{ij}, there may be many stable equilibrium points, each one corresponding to the presence in the population of a particular subset of the alleles so that stability in this sense is not global. Let us suppose that three alleles at a locus are at equilibrium at $\hat{x}_1 = \hat{x}_2 = \hat{x}_3 = 0.33$. Any perturbation of the frequencies may result in a return to this equilibrium so that it is stable, by definition. Nevertheless, the introduction of two new alleles may result in a completely new stable equilibrium configuration with all five in stable

equilibrium. A well known case of this is the equilibrium for self-sterility alleles [15].

3. Multiple loci

There is no loss of generality when we consider multiple loci, if each locus is assumed to have only two alleles. If we assume as a first approximation that the distribution of allelic frequencies at the ℓ different loci are independent, the description of the genetic composition is a point in a unit hypercube of ℓ dimensions since the allelic frequency at each locus separately is a sufficient description. The mean fitness function $\overline{W}$ is defined by

$$\overline{W} = \sum Z_{ij\cdots t} W_{ij\cdots t}, \tag{3.1}$$

where $Z_{ij\cdots t}$ is the frequency of a diploid genotype and is equal to the product of the separate frequencies of the partial genotypes at each locus. Thus, the frequency of the genotype $AaBBcc$ is equal to the product of the marginal frequency of Aa by the marginal frequency of BB by the marginal frequency of cc.

The equation for gene frequency change at each locus separately is

$$\Delta p_i = \frac{p_i(1 - p_i)}{2\overline{W}} \frac{d\overline{W}}{dp_i}, \tag{3.2}$$

where p_i is the frequency of one of the two alleles at the ith locus and again the problem can be seen as a potential problem with the point representing the gene frequency composition of the population going to a value in the hypercube representing a maximum of $\overline{W}$. Two important points emerge from this analysis.

First, because of the definition of $\overline{W}$ in expression (3.1), there is not necessarily a unique maximum in the hypercube. Unlike the single locus case, where the population always goes to the same configuration of allelic frequencies given the subset of alleles segregating in the population, there may be more than one local maximum in the multiple locus case and two different populations will tend toward different equilibria depending upon their instantaneous position. Figure 1 shows a case of a two locus, two allele fitness surface derived from data on natural populations of a grasshopper, *Moraba scurra* [12]. There are two stable points, represented by two peaks at the margins of the square, as well as a low point (V) and a maximum (S). The figure shows different paths of gene frequency change from different initial positions and it is particularly noteworthy that two initial conditions very close to each other in the upper right corner may lead to totally different final results. It is difficult to say what is the maximum number of singularities on the $\overline{W}$ surface and what their nature may be.

Since $\overline{W}$, defined by (3.1), is the product of ℓ quadratics, it would seem that there would be 3^ℓ possible singularities including the boundaries and that 2^ℓ of them would be on a boundary. This is obviously true for $\ell = 1$ and Moran [12] has shown that it is true for $\ell = 2$, where, of the nine possible equilibria, five

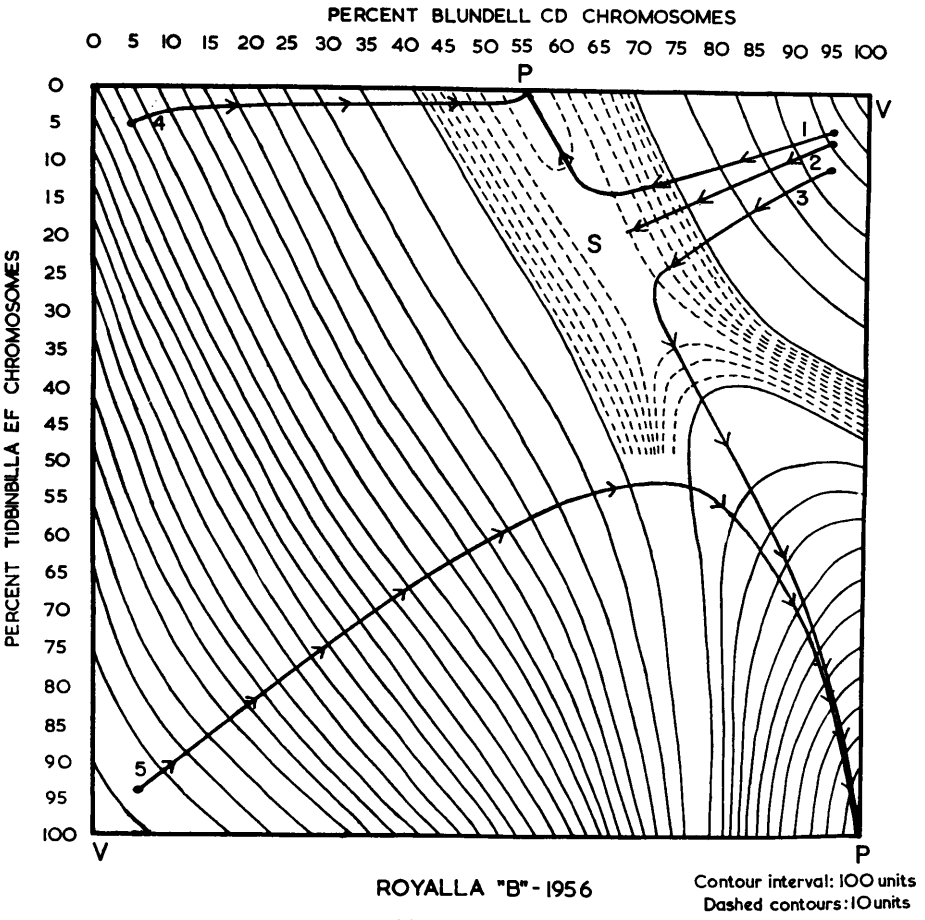


FIGURE 1

Fitness surface for two locus system estimated from field data on *Moraba scurra* [12].
Arrows show the direction of change in the gene frequencies from given starting points.
P are fitness maxima, V are fitness minima, and S is a minimax.

can lie entirely *inside* the unit square. Moran has produced a case where three of these are maxima and therefore stable.

The second point is that $\overline{W}$ is not exactly a potential function in that the gene frequency does not take the path of least action on the surface. Equation (3.2) shows that Δp_i is proportional not only to the gradient of the surface but also to the absolute position, $p(1 - p)$. Thus, the direction of the vector of gene frequency change which will be given by $\Delta p_i / \Delta p_2$ in the two locus case is

$$\frac{\Delta p_1}{\Delta p_2} = \frac{p_1(1 - p_1)\dfrac{d\overline{W}}{dp_1}}{p_2(1 - p_2)\dfrac{d\overline{W}}{dp_2}}, \tag{3.3}$$

instead of only by the ratio of the gradients. This effect is shown graphically in figure 2 from Lewontin and Kojima [11]. One quadrant of the unit square is shown with the stable equilibrium point corresponding to the maximum of the $\overline{W}$ surface in the upper left corner. The solid line represents the path the population composition would take if changed according to the principle of least action, while the broken line shows the actual course of the change under the assumption of independence of the two loci.

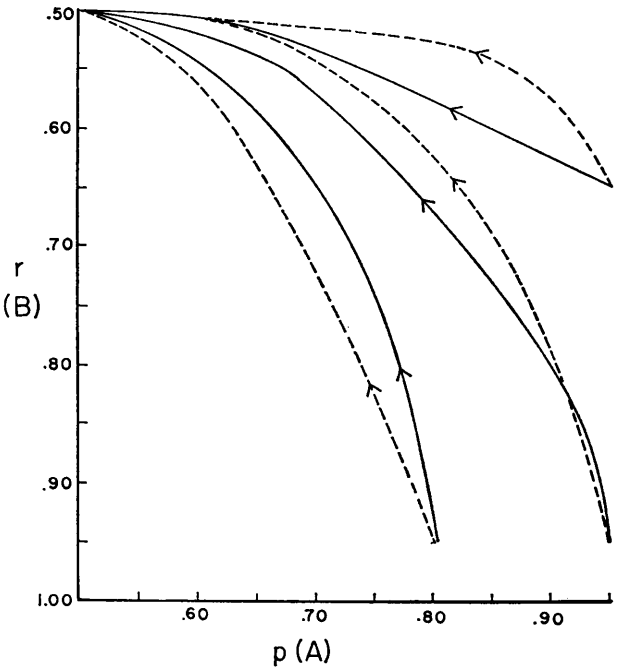


FIGURE 2

Contrast between the actual trajectories of gene frequency change on a fitness surface (broken lines) and the path of steepest ascent of "minimum action" (solid lines) [11].

4. The complication of linkage

When two or more loci are considered, it is not at all obvious under what conditions the joint distribution of genotypic frequencies at all loci can be rep-

resented as the product of independent distribution at each locus separately. If we relax the assumption of independence, the frequencies of the alleles at each locus are insufficient to specify the populations since there are only ℓ such allelic frequencies while there are $a^\ell - 1$ independent gametic frequencies.

In order to handle this problem, we return to the one locus, multiple allele model but now regard each gametic type as an allele. In such a model, the "alleles" change frequency for two reasons. First, natural selection as summarized in the fitness values W causes some gametic types to increase at the expense of others. Second, *recombination* in heterozygotes produces "alleles" that were not present in the parents. Thus, the double heterozygote AB/ab also produces the alleles Ab and aB and the proportion of these each generation is a function of the frequency of heterozygotes from which they can be produced, but also of the linkage distance between the A and B loci. In general [10], the change in frequency x_i of a gametic type can be written as

$$\Delta x_i = \frac{x_i(1 - x_i)}{2\overline{W}} \frac{\partial \overline{W}}{\partial x_i} - \rho(x, R), \tag{4.1}$$

where $\rho(x, R)$ is a function of the vector of gametic frequencies x and the set of recombination fractions among the loci R. In general, $\rho(x, R)$ is a messy function with large numbers of terms that make the actual evaluation of Δx_i tedious, and completely general statements about rate of change and equilibrium values difficult. For the case of two loci and two alleles at each locus Lewontin and Kojima [11] have given a simple form of these equations (erroneously transcribed in [8]),

$$\Delta x_i = \frac{x_i(W_{i\cdot} - \overline{W}) - k(i)RW_{11}D}{\overline{W}}, \tag{4.2}$$

where x_0, x_1, x_2, and x_3 are, respectively, the frequencies of ab, aB, Ab, and AB gametes;

$$\begin{aligned} W_{i\cdot} &= \sum_j W_{ij}x_j; \\ \overline{W} &= \sum_i W_{i\cdot}x_i; \\ D &= x_0x_3 - x_1x_2; \\ R &= \text{recombination fraction between locus } A \text{ and } B; \end{aligned} \tag{4.3}$$

and $k(i) = 1$ for $i = 0, 3$; $k(i) = -1$ for $i = 1, 2$. A closely analogous set of equations was given by Kimura [4] for continuous generations and is the first published exact treatment, to my knowledge, of the problem of linkage and selection. Further analogous equations for haploid models are given by Felsenstein [2]. A completely general but not very useful form of $\rho(x, R)$ is given by Lewontin [8] for arbitrary numbers of loci. In general, resort must be had to numerical solutions, to simplified symmetrical situations, or to approximations for small values of selection coefficients and recombination fractions, since even the simple equations given by (4.2) do not have simple equilibrium solutions.

In recent years, a number of authors (Bodmer and Parsons [1], Felsenstein [2], Haldane [3], Kimura [4], [5], Kojima and Kelleher [6], Lewontin [8], [9], [10], Lewontin and Kojima [11]) have studied a number of models and situations by a variety of methods. All together, these studies give a fairly broad picture of the equilibrium and kinetic properties of multilocus systems in large populations. As yet, the importance of linkage in stochastic models is poorly understood. The remainder of this paper will be devoted to the general principles that have been established for linked gene systems.

5. Equilibrium properties

(1) *If linkage is tight enough, gametic frequencies are affected by linkage even though gene frequencies may not be.* Table I shows a general set of symmetrical

TABLE I

FITNESSES FOR A SIMPLE SYMMETRICAL TWO LOCUS MODEL ALLOWING A LITERAL SOLUTION

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

fitnesses for the nine genotypes in a two locus system. In such a case, Lewontin and Kojima [11] show that stable equilibrium of gametic frequencies is given by

$$x_i = \frac{1}{4}, \tag{5.1}$$

or

$$x_i = \frac{1}{4} \pm \frac{1}{4}\left(1 + \frac{4Rd}{b + c - a - d}\right)^{1/2}. \tag{5.2}$$

At both of these equilibria, the allelic frequencies at both loci are one half since by symmetry $x_0 = x_3 = 1/2 - x_1 = 1/2 - x_2$, and the allelic frequencies at the two loci, p and r, are given by $x_0 + x_1$, and $x_0 + x_2$, respectively. Equilibrium of the form (5.1) are also seen to be in *linkage equilibrium* since the gametic frequencies are the product of the respective allelic frequencies. Equilibrium of the form (5.2) are in *linkage disequilibrium* and this can be measured by

$$D = x_0x_3 - x_1x_2 = \pm\frac{1}{4}\left(1 + \frac{4Rd}{b + c - d - a}\right)^{1/2}. \tag{5.3}$$

Lewontin and Kojima show that the condition for (5.2) to be the stable equilibrium as opposed to (5.1) is that

$$R \leqq \frac{a + d - b - c}{4d}, \qquad b + c - a - d < 0. \tag{5.4}$$

That is, linkage must be tighter than a value determined by the relationship

among the fitnesses. Moreover, if $a + d - b - c = 0$, no value of linkage is tight enough to cause stable linkage disequilibrium. But $a + d - b - c$ is a measure of deviation from additivity of the fitnesses. This leads to the second rule, one which has been found to be general even for nonsymmetrical cases.

(2) *Linkage has an effect on equilibrium gametic frequencies only if there is non-additivity among the fitness effects at different loci.*

(3) *Tight linkage may produce stable equilibrium of intermediate gene frequency where none would exist without linkage.* Lewontin and Kojima [11] show that for the condition (5.1) to be stable not only must linkage be looser than that given by relation (5.4), but also $|a - d| > |b - c|$ and $d > a$. Consider the fitnesses $a = 3$, $b = 4$, $c = 1$, and $d = 5$. These do not satisfy the condition just stated so that stable equilibrium at $x_i = 1/4$ is not possible. However, if linkage is tighter than approximately $R = 0.116$, the equilibrium given by (5.2) is stable.

(4) *Linkage, if it is too tight, may destroy the stability of a gene frequency equilibrium.*

Table II from Lewontin [8] shows the equilibria that result from assignments of $a = 0.9$, $b = 0.2$, $c = 0.2$, and $d = 1.00$ in table I. The values of D' given in the table are obtained by dividing the measure of linkage disequilibrium D by

TABLE II

RESULTS OF THE APPLICATION OF THE FITNESSES IN TABLE I WITH $a = 0.9$, $b = 0.2$, $c = 0.2$, AND $d = 1.00$
g_{00}, g_{01}, g_{10}, AND g_{11} REFER TO THE GAMETES ab, aB, Ab, AND AB
(Other symbols are explained in the text).

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

its maximum value given the allelic frequencies. The symbols p and r stand for gene frequencies at the two loci. As the table shows, when linkage is quite tight there are stable equilibria with gene frequency at one half and the gametic frequencies very much distorted from independence. When linkage is between 0.10 and 0.375, however, there are no stable equilibria at all and the gene frequencies at both loci fix at 0 or 1. Then, with yet looser linkage, there is a stable equilibrium, but one corresponding to linkage equilibrium.

Table II also shows in striking form a phenomenon that will be apparent in all results to be given.

(5) *Linkage increases the mean fitness of the equilibrium population.* Table II shows that very tight linkage results in a 65 per cent increase in fitness in this model. For the general model of table I, Lewontin and Kojima [11] show that

$$\overline{W} = \overline{W}_{LE} + 4D^2(a + d - b - c), \tag{5.5}$$

where $\overline{W}_{LE}$ is the value of mean fitness at linkage equilibrium ($x_i = 1/4$). The increase will not be large unless the epistatic deviation ($a + d - b - c$) is large since D^2 is always less than or equal to 0.0625.

(6) *The equilibrium allelic frequencies are generally altered by linkage.* The previous results based on the fitnesses in table I show no sensitivity of the allelic frequencies at equilibrium to different values of R. This is an artifact of the symmetry of the fitnesses in the table. Table III shows an asymmetrical

TABLE III

FITNESSES FOR AN ASYMMETRICAL TWO LOCUS MODEL

	AA	*Aa*	*aa*
BB	.5000	.5000	.3750
Bb	.5625	1.0000	.3125
bb	.3750	.4375	.3750

model and table IV gives the stable equilibria found by numerical methods. Not only are the gametic frequencies dependent upon the value of R but there

TABLE IV

RESULTS OF THE MODEL WHOSE FITNESSES ARE SHOWN IN TABLE III
(Symbols are as in Table II.)

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	.51860	−.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

is a considerable change in allelic frequencies at each locus with increasingly tight linkage. This table also illustrates two other points.

(7) *There may be alternative stable equilibrium for a given linkage value.* Condition (5.2) allows for reciprocal equilibria with either the coupling gametes in excess (x_0 and x_3 have the positive sign before the radical while x_1 and x_2 have the negative) or repulsion in excess. These alternative and totally equivalent equilibrium have oppositely signed D values but the same mean fitness. In the asymmetrical case of table IV, however, there are two nonequivalent equilibria for each of the tighter linkage values. The coupling equilibria have a slightly higher $\overline{W}$ in each case, and it is these that persist at looser linkages; but in other models it is the repulsion equilibrium that gives the higher fitness.

(8) *Even genes on different chromosomes may be out of random combination at equilibrium.* The last line in table IV shows that a linkage disequilibrium of 2.6 per cent of maximum persists at gene frequency equilibrium for this model even when the genes recombine with a frequency of 50 per cent. Thus, genes on different linkage groups can nevertheless be correlated in their allelic frequencies. What is required is that the epistatic interaction be quite strong. In the model of table III, the heterozygote Bb changes from underdominance to overdominance as the genotype at the A locus changes from aa to Aa. A similar change takes place for the Aa heterozygote for changes at the B locus. This is an extreme form of interaction and may be considered as unrealistic. As we shall show, there are natural models of natural selection that generate even greater amounts of epistatic interaction.

(9) *Multiple locus systems show cumulative effects of linkage along the chromosome.* The effects of linkage on the equilibrium configuration of gametes is small unless linkage distances are short. Table IV shows that the changes of gene frequency and fitness are not very great for recombination values in excess of 0.075. Since most genes controlling a character are sprinkled throughout the genome, it does not seem that tight linkage is very important. However, it turns out that effects are cumulative along the chromosome as shown in table V. This is a sample of results from a five locus model discussed by Lewontin [8]. The body of the table shows stable equilibrium frequencies of 16 gametic types out of the 32 possible for five loci with two alleles each. The complementary gametic types have frequencies equal to those shown because of symmetry of the model. The bottom half of the table shows all possible pairwise linkage disequilibrium measures D'_{ij}. What is important about this result is that when $R = 0.05$ between *adjacent* loci, the outside loci are 20 recombination units apart yet they are out of random combination with a $D' = 0.23$. Thus, even though the linkages between adjacent genes must be small for there to be a significant linkage effect, the entire chromosome can be strongly correlated.

The selection model on which table V is based is one of cumulative heterosis where fitness depends only upon the number of loci heterozygous. Individuals with zero, one, two, three, four and all five loci heterozygous have fitness 0.03, 0.06, 0.12, 0.24, 0,48, and 0.96, respectively. A multiplication model is a reason-

TABLE V

THE RESULTS OF A FIVE LOCUS CUMULATIVE HETEROSIS MODEL
(Symbols are explained in the text.)

	R between Adjacent Loci									
Gametes	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

able one for fitnesses although the actual fitness loss per locus is unusually strong for this model. One result of the big fitness difference is that there is more than a doubling of $\overline{W}$ from the loosest to the tightest linkage (last line of table V).

(10) *Genes closer to the center of a linkage group are more strongly correlated than those at the ends.* This can be seen in table V by comparing D'_{23} and D'_{34} which are adjacent intervals at the center of the group, with D'_{12} and D'_{45} which are intervals of the same length but at the ends. In all cases

$$D'_{12} = D'_{45} < D'_{23} = D'_{34}. \tag{5.6}$$

(11) *Optimum deviation models generate enough epistasis to cause profound effects of linkage.*

We define epistatic interactions as deviations from additivity between loci. As we have seen, it is necessary to postulate very intensive selection to produce strong epistasis. However, in any model in which an intermediate phenotype is

the most fit and in which that phenotype is determined additively by many loci, there will be extreme epistatic effects, even when selection is weak. This is because the relation between gene dose and fitness is not monotone.

A commonly studied form of optimum model is the squared deviation model of Wright [15]. If

P_i = phenotype of the ith genotype,

$\bar{O}$ = optimum phenotype, and

K = an arbitrary scaling constant,

then the fitness of the ith phenotype is

$$W_i = K - (P_i - \bar{O})^2. \tag{5.7}$$

Table VI shows the result of calculations on a two locus quadratic optimum model [8]. The model assumes that the phenotype P_i is simply the sum of contributions from the two loci. Each locus has phenotypes corresponding to genotypes as follows:

TABLE VI

RESULTS OF A TWO LOCUS OPTIMUM QUADRATIC DEVIATIONS MODEL
(Symbols are explained in the text.)

R	g_{11}	g_{10}	g_{01}	g_{00}	$p = r$	D'	$\bar{W}$	$\bar{P}$	V
.00	.19913	.40043	.40043	.00001	.59956	−1.0000	.95138	5.84670	17.5465
.01	.30506	.34188	.34188	.01118	.64694	−.91038	.94814	6.81564	17.9847
.02	.36381	.30850	.30850	.01919	.67231	−.82129	.94546	7.30789	17.8294
.03	.40313	.28600	.28600	.02487	.68913	−.74265	.94345	7.62661	17.5641
.05	.45275	.25738	.25738	.03250	.71013	−.61327	.94065	8.00729	17.1108
.07	.48238	.24025	.24025	.03712	.72263	−.51746	.93885	8.22940	16.7698
.10	.50913	.22471	.22471	.04144	.73384	−.41501	.93710	8.42475	16.4406
.15	.53288	.21090	.21090	.04532	.74378	−.30968	.93557	8.59495	16.0963
.20	.54575	.20340	.20340	.04744	.74915	−.24600	.93467	8.68589	15.9160
.25	.55400	.19867	.19867	.04866	.75267	−.20483	.93416	8.74475	15.7663
.30	.55950	.19547	.19547	.04956	.75497	−.19121	.93376	8.78303	15.6847
.35	.56350	.19314	.19314	.05022	.75664	−.15213	.93349	8.81086	15.6191
.40	.56650	.19139	.19139	.05072	.75789	−.13477	.93328	8.83166	15.5717
.45	.56900	.18997	.18997	.05106	.75897	−.12101	.93310	8.84954	15.5301
.50	.57050	.18898	.18898	.05153	.75948	−.11098	.93295	8.85796	15.5231

genotype	AA	Aa	aa
phenotype	6	3.6	−6,

so that there is considerable dominance of the A allele. The optimum value is $\bar{O} = 6$, so that no genotype in particular corresponds to the optimum. The double heterozygote $Aa\ Bb$ is closer than any other, with a score 7.2. Table VI shows that all values of recombination including $R = 0.50$ have a significant effect on gametic frequencies and that tighter linkages alter the gene frequencies as well. Two general points are illustrated that apply to all quadratic deviation models investigated.

(12) *Quadratic optimum models always generate repulsion equilibria.* All the

D' values in table VI are negative indicating an excess of repulsion classes. This is to be expected since an intermediate phenotype is achieved by a mixture of alleles operating in opposite directions.

(13) *The effect of linkage in quadratic optimum models is to increase the genetic variance and decrease the deviation of the population mean from the optimum.*

If we take the expectation of both sides in expression (5.7), we get

$$E(W_i) = \overline{W} = K - [(\overline{P} - \overline{O})^2 + \sigma_P^2]. \tag{5.8}$$

In table VI it can be seen that a very small change in $\overline{W}$ takes place over the entire range of recombination values because of the compensatory changes in the two terms of the right side of (5.8). As linkage tightens the mean phenotype $\overline{P}$ changes from 8.86 to 5.85 so that its deviation from $\overline{O}$ changes from 2.85 to -0.15. At the same time the genetic variance of the population increases from 15.52 to 17.54. Therefore, the sum of squared deviation of genotypes from the *optimum* has changed very little.

6. Kinetics of gene frequency change

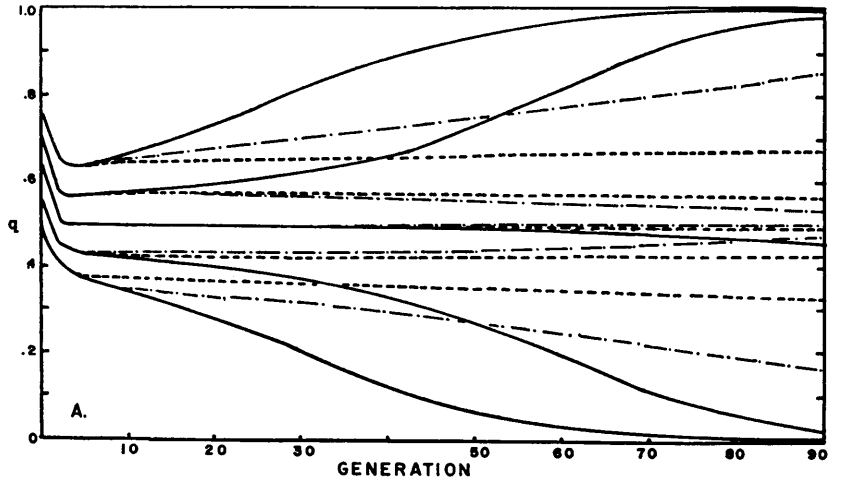
If we consider selection problems in which no stable equilibrium of allelic frequencies at intermediate values will occur, or in which the approach to an equilibrium is of interest, we can ask about the rate of gene frequency change and fitness change with time as well as under what conditions linkage disequilibrium will be generated or in what direction it will change if it is already present. General results of this nature have been derived by Felsenstein [2] for both continuous and discontinuous generation models. In contrast to the equilibrium studies where epistasis is measured on an *additive* scale, it should be measured on a *multiplicative* scale for kinetic studies. Redefining epistasis as deviation of fitnesses from the product of the fitnesses at each locus separately, Felsenstein has shown:

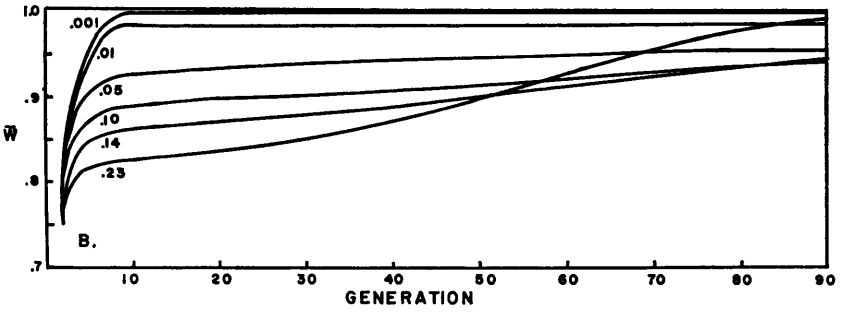
(14) *if $D = 0$ it will remain zero in the absence of epistasis;*

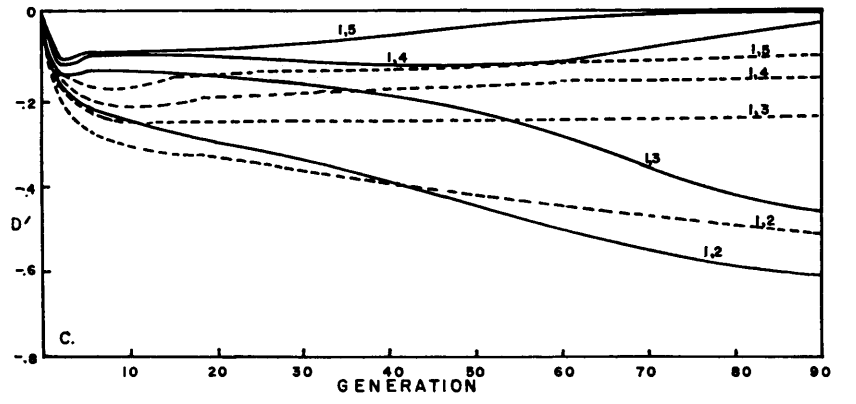
(15) *if there is epistasis, linkage disequilibrium will be generated of a type predicted by the direction of the epistasis;*

(16) *tight linkage will accelerate the rate of gene frequency change if coupling linkages are generated by selection and retard the rate if repulsion linkages are generated.*

The general principle given in property (16) can have quite remarkable results. Lewontin [9] considered several intermediate optimum models in which no stable intermediate equilibrium of gene frequency is expected. In these models the effect of genotype on the phenotype was completely additive both between loci and between alleles within loci, so that a multiple balanced homozygote of the form $AA\ bb\ CC\ dd\ \cdots$, and so forth, could have the optimum phenotype. In fact, in such a model the gene frequencies go to fixation at 0 or 1 to produce such a balanced genotype. Figure 3 shows the results for this model with five loci and the optimum value exactly halfway between the extremes.


1.0
.8
.6
q
.4
.2
0
A.
10
20
30
50
60
70
80
90
GENERATION


1.0
.9
.8
.7
$\bar{W}$
.001
.01
.05
.10
.14
.23
B.
10
20
30
40
50
60
70
80
90
GENERATION


0
-.2
-.4
-.6
-.8
D'
1,5
1,4
1,3
1,2
C.
10
20
30
40
50
60
70
80
90
GENERATION

In the top graph the allelic frequencies of the five loci are shown. The solid lines give the results when the adjacent loci are 23.4 linkage units apart. By 90 generations, two loci are virtually fixed at $q = 0$ and two at $q = 1.0$. The fifth locus is beginning to fix at $q = 0$. The dashed lines show the contrast when the linkage between adjacent loci is reduced to $R = 0.05$ and $R = 0.01$. For this tightest linkage there is essentially *no change* in allelic frequencies after the initial adjustment of the population mean to be at the optimum. The tight linkage has produced a quasiequilibrium of allelic frequencies. As the bottom graph of the figure shows, all D' are negative so that the retardation of gene frequency change is to be expected. The virtual stasis of the frequencies, however, is unexpected. Linkage has produced what is effectively an equilibrium of gene frequencies.

Changes in fitness are more complex. The most tightly linked cases rise in fitness very rapidly while loose linkage is slower in its rate of improvement but finally catches up and surpasses the fitness of the tightly linked case. The course of events is as follows. In the loosely linked case, gene frequencies are going to fixation, and fitness slowly improves as the alleles are fixed. When all genes are fixed, the fitness goes to unity. In the tightly linked case, however, there is no change in allelic frequency but a very rapid change in gametic frequency so that essentially only alternate repulsion gametes of the type *AbCdE* and *aBcDe* are left. As a result, nearly every individual is at the phenotypic optimum and fitness is very high. However, since the genes continue to segregate, no further advance in fitness can be achieved. Loosely linked systems achieve high fitness by changing allelic frequencies, while tightly linked systems become highly fit through elimination of poor gametic combination.

7. Work to be done

Our most complete understanding is in the realm of equilibrium models, but even for these, general literal solutions are difficult to find. What is required to complete our understanding of the equilibria in complex systems is, first, a solu-

---

FIGURE 3

Results of optimum model with five loci.
Part a shows the frequency of the *O* allele (ordinate)
at various times (abscissa) for five loci.
Dashed lines are for $R = 0.01$, mixed line for $R = 0.05$,
solid lines for $R = 0.234$.
Part b shows the mean fitness of the population (ordinate) at
various times (abscissa) for different degrees of linkage.
Part c shows linkage disequilibrium parameters D' (ordinate) at
various times (abscissa).
Dashed lines are for $R = 0.01$, and solid lines for $R = 0.234$.

tion to the problem of how many stable equilibria are possible with a given set of fitnesses in multilocus models, and, second, an exact general relationship between the fitnesses and the degree of departure of gametic frequencies at equilibrium from random combination.

For kinetic studies, Felsenstein has achieved some important general results, but these are mostly in terms of the direction in which D will change. The magnitude of changes in D can be specified as complex functions of the fitnesses and the gene frequencies, but at the moment we have few formulations that give insight into the relation between fitnesses and their magnitudes [5]. We know even less about how drastic are retardations or increases in the speed of selection as a result of linkage.

Another important problem in kinetics arises when different groups of genes have different linkage relations. We have some preliminary information that a tightly linked system will accelerate the change in gene frequencies in another more loosely linked system of genes controlling the same character.

Some studies have been made on changes in gene frequencies when these are near fixation, by Bodmer and Parsons [1] and by Kojima and Schaeffer [7]. These results have important implication for the role of linkage in helping or preventing the incorporation of new variation from mutation. A general attack on the rate of fixation of genes as influenced by linkage has not yet been made. A union of linkage studies with stochastic theory would be a long step toward a realistic theory of population genetics.

REFERENCES

[1] W. F. Bodmer and P. A. Parsons, "Linkage and recombination in evolution," *Adv. Genet.*, Vol. 11 (1962), pp. 1–99.

[2] J. Felsenstein, "The effect of linkage on directional selection," *Genetics*, Vol. 52 (1965), pp. 349–363.

[3] J. B. S. Haldane, "The selection of double heterozygotes," *J. Genet.*, Vol. 58 (1962), pp. 125–128.

[4] M. Kimura, "A model of a genetic system which leads to closer linkage by natural selection," *Evolution*, Vol. 10 (1956), pp. 278–287.

[5] ———, "Attainment of a quasi-linkage equilibrium when gene frequencies are changing by natural selection," *Genetics*, Vol. 52 (1965), pp. 875–890.

[6] M. Kojima and T. Kelleher, "Changes of mean fitness in random mating populations when epistasis and linkage are present," *Genetics*, Vol. 46 (1961), pp. 527–540.

[7] K. Kojima and H. E. Schaffer, "Accumulation of epistatic gene complexes," *Evolution*, Vol. 18 (1964), pp. 127–129.

[8] R. C. Lewontin, "The interaction of selection and linkage. I. General consideration; heterotic models," *Genetics*, Vol. 49 (1964), pp. 49–67.

[9] ———, "The interaction of selection and linkage. II. Optimum models," *Genetics*, Vol. 50 (1964), pp. 757–782.

[10] ———, "The role of linkage in natural selection," *Genetics Today*, New York, Macmillan, 1964, Vol. 2, pp. 517–525.

[11] R. C. Lewontin and K. Kojima, "The evolutionary dynamics of complex polymorphisms," *Evolution*, Vol. 14 (1960), pp. 458–472.

[12] R. C. Lewontin and M. J. D. White, "Interaction between inversion polymorphisms

of two chromosome pairs in the grasshopper *Moraba scurra*," *Evolution*, Vol. 14 (1960), pp. 16–24.

[13] P. A. P. MORAN, "On the measurement of natural selection dependent on several loci," *Evolution*, Vol. 17 (1963), pp. 182–186.

[14] S. WRIGHT, "Evolution in populations in approximate equilibrium," *J. Genet.*, Vol. 30 (1935), pp. 257–266.

[15] ———, "The distribution of self-sterility alleles in populations," *Genetics*, Vol. 24 (1939), pp. 538–552.

[16] ———, "Factors of evolution," *Cold Spring Harbor Symp. Quant. Biol.*, Vol. 20 (1955), pp. 16–24.