

fruits on the forest floor are coarse-grained differences for *Drosophila* larvae, which usually develop within a single fruit, but are fine-grained for adults. Since it is in coarse-grained selection that inbreeding may be advantageous, we expect that for the small, relatively immobile invertebrates there will be a maximum of formation of partly isolated "quasiraces" living more or less sympatrically over wide geographic ranges.

More mobile small invertebrates experience more of the environmental variation as fine-grained. Here polymorphism requires at least average heterosis, concavity of the fitness set leads to fixation of genes and specialization, and alternative adaptive peaks are less common. Where different parts of the life cycle differ in mobility and hence in environmental grain, we expect polymorphism to be controlled more by the immobile stages. Thus, in *Drosophila*, we expect larval viabilities to be responsible for more polymorphism than are adult phenotypes.

If the approach developed in this chapter is valid, it can be applied as follows:

1. The notion of genetic coherence, deliberately left ambiguous at this stage of the investigation, should be explored systematically and studied as a sufficient parameter.
2. The varying steepness of clines and the occurrence of discontinuities can be studied by fitness set analysis.
3. The taxonomic significance of regions of increased variability (usually regarded as zones of secondary contact) and of partial reproductive isolation (often interpreted as incipient speciation) requires re-examination.
4. Predictions can be made as to the different geographic patterns of groups differing greatly either in their environmental patterns (grain) or flexibility.

## CHAPTER FIVE

## The Genetic System

Population genetics began with the study of changes in gene frequency at a single locus in a random mating population under the joint action of selection, mutation, random drift, and migration. The major results were usually the establishment of conditions for equilibrium. Subsequent work has made the models more complex by allowing many loci, arbitrary epistatic interactions among them, effects of linkage, diverse systems of mating, fluctuations in selection, dependence of selective values on gene frequencies, etc.

The growing availability of high-speed computers has made it possible to simulate situations that cannot be handled analytically, so that great masses of data are accumulating. But it is not obvious what models to consider, how to specify the gene interactions, or even what results to look for. In some cases, the model is determined by a specific biological problem—a species to be analyzed. But for general theoretical work, the problem is, what are the sufficient parameters of complex genetic systems which are relevant to ecological evolutionary studies?

In the purely ecological analysis of adaptive strategies we looked at optimum phenotypes or optimal polymorphic populations. Where the optimum strategy is an individual phenotypic trait such as size, degree of habitat selectivity, or photoperiod threshold for the initiation of diapause it can be selected for like any other phenotypic trait. One of the restrictions is that with a continuum of possible phenotypic values but discrete genes there may not be any single genotype which corresponds to the optimum, or if it does exist it may be heterozygous and therefore segregate non-optimal types. This difficulty becomes less important as

THE GENETIC SYSTEM

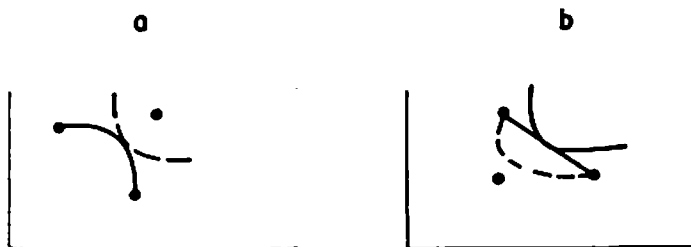


FIGURE 5.1. The genetic constraints on optimality. (a) The phenotypic optimum is a heterozygote, so that polymorphism is imposed by segregation. (b) The optimum would be a mixture of homozygotes, but this creates heterozygotes.

the number of loci increases. But it must be borne in mind that not all stable polymorphisms corresponds to an optimum strategy. Conversely, a mixed strategy may be optimal, but if it involves a mixture of homozygotes it cannot be achieved. Figure 5.1 illustrates both of these situations.

A second type of constraint involves the adaptive value of genetic change. In an engineer's adaptive system information about the environment is collected and analyzed. The record of previous experience can be designated the state of the system. On the basis of this information an optimum strategy is computed which becomes the "output" of the system. But in population genetics the output and state of the system are confounded. The information about past environments is stored as gene frequencies and gametic frequencies, but the strategy itself is also an array of genotypes. This means that a mixed strategy which requires genetic heterogeneity imposes the possibility of changing frequencies in response to selection even if the optimum strategy should always be the same (when there is no correlation between the environments of successive generations). Conversely, the capacity to respond to selection requires genetic variance even if the optimum strategy is always monomorphic.

The array of gene frequencies constitutes the "memory" of the system. Different genetic systems may have longer or shorter memories. For instance, consider additive selection as follows:

THE GENETIC SYSTEM

Genotype	Frequency	Fitness
AA	$x^2$	1
AA'	$2x(1-x)$	$1-s$
A'A'	$(1-x)^2$	$1-2s$

$$\bar{W} = 1 - 2s(1-x) \quad (5.1)$$

so that the rate of change of gene frequency is

$$\frac{dx}{dt} = s(t)x(1-x) \quad (5.2)$$

where  $s(t)$  is a random variable.

Here we can integrate to directly obtain

$$\log\left(\frac{x}{1-x}\right) - \log\left(\frac{x_0}{1-x_0}\right) = \int s(t) dt. \quad (5.3)$$

Thus the present gene frequency depends on the environments of the past, all with equal weight. This is the longest memory system we can have. But by virtue of its long memory it cannot track the recent environment.

The simplest way to reduce memory is by mutation. Let there be mutation in both directions with equal frequency  $u$ . Then equation 5.2 becomes

$$\frac{dx}{dt} = sx(1-x) + u(1-2x). \quad (5.4)$$

Let the dependence of  $x(t)$  on  $s(t-\gamma)$ , some environment of the past, be designated  $Z(t,\gamma) = E[x(t)S(t-\gamma)]$ , where the symbol  $E$  means expected value. Then

$$Z(t+h,\gamma) = E\{x(t)S(t-\gamma) + h[S(t-\gamma)S(t)x(t)[1-x(t)] + US(t)[1-2x(t)]]\}. \quad (5.5)$$

For situations in which  $E(s) = 0$  and there is no autocorrelation of the  $s(t)$  with  $s(t-\gamma)$  we get

$$Z(t+h,\gamma) = Z(t,\gamma) - 2uhZ(t,\gamma). \quad (5.6)$$

This leads to the differential equation

$$\frac{dz}{d\gamma} = -2uZ \quad (5.7)$$

## THE GENETIC SYSTEM

which means that the dependence of  $x(t)$  on the environment at  $t - \gamma$  falls exponentially. The factor  $2u$  is the rate at which memory is destroyed. Different genetic systems have different, usually non-linear, memory functions. Even the case we treated as a differential equation loses memory if it is replaced by a generation model.

The paradox which now emerges is that only a system with short memory can follow the environment. But the optimum parameters of the tracking system depend on the mean, variance, and autocorrelation of the environment. These can only be estimated accurately by a system with a long enough memory so that the law of large numbers operates. Since the statistics of the environment are also subject to change, the calibrating system cannot have infinite memory. There is some optimum level of memory for it, which can only be established by systems with longer memory, etc. Lewontin (1966) has used the term "capricious" to describe a random process with limited memory.

Thus the memory of different genetic systems, especially those involving multiple loci, becomes an important sufficient parameter.

The time/response pattern of multiple genetic systems has other aspects. Unpublished computer simulation studies by Lewontin and his collaborators showed that the response to directional selection may be delayed at the phenotypic level while internal changes occur in linkage relations. The delay seems to depend on tightness of linkage, and kinds of dominance, and epistatic relations. A system which waits five generations before responding can be resistant to ephemeral fluctuations of the environment and yet respond well to long-term changes. It would serve as a "filter," discarding the short-period changes of the environment as "noise."

Another aspect of optimization which is poorly understood is second-order selection. We define second-order selection as selection for a trait which does not appear in the equation for  $\bar{W}$  directly, such as mutation rate. Levins (1967) showed that at a single locus with a heterotic lethal, when the fitness of the viable homozygote fluctuates the

## THE GENETIC SYSTEM

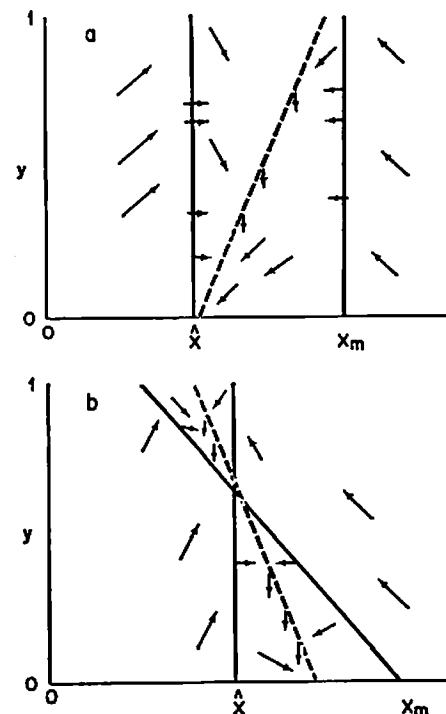


FIGURE 5.2. Selection for mutation rate. The abscissa is the frequency  $x$  of the principal gene, the ordinate is the frequency  $y$  of the gene that increases mutation. The dotted line is the equilibrium value of  $x$  for each  $y$ ,  $\hat{x}$  is the equilibrium of  $x$  under selection alone, and  $x_m$  is the equilibrium of  $x$  under mutation alone. Selection and mutation together move  $x$  toward the dotted line while  $y$  increases outside the interval  $(\hat{x}, x_m)$  and decreases within the interval. (From Levins, 1967.)

average frequency of the lethal is below optimum. Therefore, some mutation toward the lethal would be advantageous. However, this does not prove that the mutation rate would increase. Figure 5.2 explores the direction of selection. For the principle locus, let  $\hat{x}$  be the equilibrium value for a given environment under selection alone, and let  $x_m$  be the equilibrium value under mutation alone. It can readily be shown that, when  $x$  lies between  $\hat{x}$  and  $x_m$ , an allele increasing mutation rate will be selected against, while if  $x$

lies outside this interval the mutation rate gene will be favored. It can further be seen that, in a constant environment,  $x$  always ends up at an equilibrium between  $\hat{x}$  and  $x_m$ . Only if there is enough fluctuation to keep  $x$  outside this interval most of the time can we expect mutation rate to be increased. But it is not yet clear how to define the conditions more precisely, since the extent of selection for mutation depends on linkage. Similar problems arise for genes affecting linkage.

If we approach the genetic system from the point of view of evolution and ecology, the complexities of multiple systems with epistasis can be reduced to relatively few sufficient parameters: (1) memory, (2) delay, (3) ridginess, and (4) multiplicity of peaks.

The emphasis on genetic studies has usually been on equilibrium situations. Experiments with laboratory populations, computer simulation, and the analytical study of simple situations support the view that populations approach the neighborhood of equilibrium rather rapidly (say in 10–30 generations). However, there are other circumstances in which the approach is greatly slowed down. In Figure 5.3 we show the fitness/gene frequency curve for a single locus with simple heterosis. The rate of change depends on  $x(1-x)$  and the slope of the fitness curve. This can be flattened when fitness depends on gene frequency.

Frequency-dependent selection has scarcely been studied systematically, partly because we don't know what models may be biologically meaningful. However, Landahl (unpublished) has considered a situation that may be of general interest. In addition to ordinary selection acting on zygotic



FIGURE 5.3. Fitness as a function of gene frequency for ordinary heterosis.

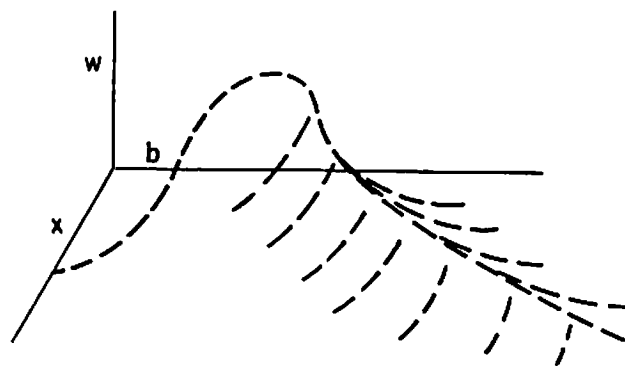


FIGURE 5.4. An adaptive surface where a ridge would show selection.

genotypes, there is an effect of the maternal genotype which acts by determining the sites at which eggs are laid. He found that, when the selective values of the genotypes for choice of egg site and for zygotic viability act in opposite directions, selection can be slowed down greatly, and gene frequencies can be quasi-stable for hundreds of generations.

Multiple-locus situations also permit quasi-equilibrium. Lewontin (1964) found this to happen in simulation studies where there was selection for some intermediate optimum. Here it was the result of selection eliminating the available variance in phenotypic effect rapidly (through elimination of gametic types) while tight linkage prevented fixation of individual loci.

A third type of quasi-equilibrium has been described by William Bossert (1967). Consider an adaptive landscape for two loci, two alleles at each. Suppose that there is a major peak which tapers out into a long ridge as in Figure 5.4.

A population which has reached the lower levels of the ridge should move upward along it. But any random perturbation due to environmental fluctuation, sampling in small populations, or occasional migration will displace the population from the ridge. Much of the response to selection will be expended, returning to or keeping on the ridge rather than moving along it. This effect would increase as

we add loci and thus increase the number of dimensions. At eight or nine dimensions the time needed to get near equilibrium may become effectively infinite.

It is clear that the degree of ridginess in the adaptive landscape is a major sufficient parameter of the genetic system. It would be defined mathematically as the ratio of the largest to the smallest eigenvalue of the set of differential equations describing the simultaneous changes of gene frequency. Biologically it will depend on the system of epistatic (non-additive) interactions among loci.

Besides its importance for the study of quasi-equilibrium, the ridginess of an adaptive landscape is one of the points of contact with developmental biology. A developing system depends on genetic interactions. But if these give rise to an adaptive landscape which is too ridgy, selection could not establish that system. This may impose restrictions on the kinds of complexity that can arise in evolution.

Although variable environments have been studied, these have usually been of the simplest kind—pure white noise (environments of successive periods independent) or simple autocorrelation. But there may be more complicated environmental patterns, with combinations of short-term and long-term fluctuations. However, it would be futile to attempt a comprehensive study of possible patterns of environment without more knowledge of the actual patterns of environment.

If in fact we have identified the relevant parameters, the approach of this chapter can be applied as follows:

1. These parameters become objects of study, measurement, and comparison in the description of different groups.
2. A knowledge of the parameters of the genetic system leads to predictions as to the kinds of adaptation which will be possible. For instance, if our argument about epistasis and ridginess is correct, the amount of epistatic interaction should be greater in species which are mostly parthenogenic.
3. Insofar as the genetic system is itself evolving, we should be able to relate the direction of its evolution to the

statistical pattern of the environment. For instance, mutation rates should be higher in species with variable environments and for those loci whose adaptive values fluctuate with the environment.

4. Laboratory experiments could attempt to produce different kinds of changes in the genetic system by natural selection, using only the pattern of environmental change.

5. Theoretical work is needed on the relation of chromosome number, tightness of linkage, and similar genetic parameters to delay, memory, ridginess, and multiple equilibria.

6. The possibilities and limits on second-order selection must be explored systematically.