RANK-ORDER SELECTION IS CAPABLE OF MAINTAINING ALL GENETIC POLYMORPHISMS

CHRISTOPHER WILLS

Dept. of Biology, University of California at San Diego,
La Jolla, California 92037

Manuscript received June 9, 1977
Revised copy received December 27, 1977

ABSTRACT

The fitness of organisms may be due chiefly to a fitness curve imposed on their ranking in the population with respect to heterozygosity. If this is so, then the number of polymorphisms that can be retained at a particular selective equilibrium increases as the square of the population size. All of the genetic variation that we currently observe and infer to exist can probably be maintained by selection in a population of about \(10^5\) individuals. Selection acting in this way is so strong that these polymorphisms can be expected to behave very differently from neutral ones.

It is generally agreed that selection based on strictly multiplicative fitnesses is inadequate to explain the amount of polymorphism observed in natural populations. In the multiplicative model, the fitness of an individual homozygous at \(n\) loci can be represented by \((1 - s)^n\), where \(s\) is the selection coefficient associated with homozygosity at each locus. As Lewontin and Hubby (1966) pointed out, if \(s\) is sufficiently large that selection, rather than drift, has the predominant effect on allele frequencies at each locus, then the load quickly becomes so great as to threaten extinction of the population as the number of segregating loci becomes large. Indeed, this was one of the arguments employed by Kimura (1968) to explain why he considered that the great majority of observed genetic polymorphisms must be selectively neutral.

Several alternative forms of selection have been proposed (Sved, Reed and Bodmer 1967; King 1967; Milkman 1967; Tobari and Kojima 1967) in order to circumvent this problem. I wish in this paper to present some consequences of the selection model proposed by Milkman (1967), following a suggestion of Mayr (1959). A less quantitative version of this model, termed "soft selection," was discussed by Wallace (1968, 1975). This model, or a modification of it, can provide for selection of all observed genetic variability. I will give an analytical model showing the numbers of polymorphisms that can be maintained with this type of selection, which I would like to term rank-order selection. I will also discuss the magnitudes of the selection coefficients that can be expected to be associated with these polymorphisms, and show that rank-order selection permits them to be so large that the behavior of these polymorphisms can be
expected to be very different from that of neutral polymorphisms. In a later paper, I will deal with the effect of linkage disequilibrium on multigenic systems subject to rank-order selection.

RANK-ORDER SELECTION

Let me begin by formalizing the concepts that underlie MILKMAN’s (1967) type of selection. In his view, it is the relative ranking of organisms with respect to some variable that in conjunction with the environment determines their fitnesses. This immediately circumvents the difficulty inherent in the multiplicative model in which fitness is given by an absolute genotypic value, since fitness becomes relative rather than absolute. What is this underlying variable that determines the rank order of the organisms in the population? MILKMAN (1973) termed it the organism’s “fitness potential,” since it is only indirectly translated into fitness; the final fitness value is determined by the relative ranking of the organism’s fitness potential among those of the rest of the population, and by the severity of the environment. The environment may impose a distribution of fitnesses on this ranking in such a way that certain organisms may not reproduce at all, even though they may be perfectly capable of doing so under less severe circumstances. This is a truncation, of the sort familiar to quantitative geneticists, and it may take many forms. Modifications of this simple model will be considered later in the paper.

It is worth noting that under the multiplicative model all organisms not homozygous for a lethal have a non-zero chance of reproduction. The range of fitnesses in a population subject to multiplicative selection is not great. For example, if 1000 loci are segregating in such a population, and at each locus \( s_1 = s_2 = 0.001 \), then the mean fitness of the population is \( 0.999^{1000} = 0.606 \), and 95% of the population will have fitnesses lying between \( 0.999^{950} = 0.587 \) and \( 0.999^{48} = 0.626 \). A truncation model immediately introduces the possibility of a fitness of zero, so that the range of fitnesses in the population can be enormously greater. Further, and this is an important difference, in the multiplicative model the effective size of the population is immediately affected by the distribution of genotypes and, therefore, (since the two are perfectly correlated on a log-linear scale) by the distribution of fitnesses. In a rank-order or other truncation model, the effective size is a function of the distribution of fitnesses, but not a direct function of the distribution of genotypes in the population. Truncation permits large differences in fitness even if the effective population size remains large. Because of this divorcing of effective population size from the distribution of genotypes, it is possible under the rank-order model to impose approximately equal amounts of selection on a per organism basis in populations with very different ranges of fitness potential. This will occur if the distribution of fitnesses imposed on these different populations by the environment is approximately the same in each case.

If this condition holds, an organism possessing a particular fitness potential may be very fit in one population in which most of the organisms fall below it
in this property, but be much less fit if it is transported to a population in which most of the organisms have a higher fitness potential.

The fitness potential undoubtedly has many components. The one with which we will be concerned here, however, is that component which leads to the maintenance of polymorphisms by balancing selection. This paper deals with simple heterotic selection, but regardless of the mode of balancing selection at each locus—single-gene heterosis, marginal overdominance, geometric mean overdominance (Felsenstein 1976) or frequency-dependent selection—the upshot is that for polymorphisms to be maintained the heterozygote must have or must appear to have the highest average fitness. In actuality, the relative contribution to this component of fitness potential by different loci and by different genotypes at a locus must vary greatly, but I will use here Milkman's original simplified model in which fitness potential is directly proportional to average heterozygosity. As will be developed later, I do not consider that small deviations from this simple model will have a large effect on the consequences for the maintenance of polymorphism.

There are several limitations associated with the rank-order model. I will mention two here and others at the end of the paper. The first is that only that component of fitness potential which leads to the maintenance of balanced polymorphisms will be considered in this discussion. Transient selected polymorphisms will also contribute to the fitness potential; there will, of course, be a large environmental component, and there will be a component of uncertain size contributed by genes exhibiting a very low genotype-environment interaction. These latter genes, chiefly lethals and severe subvitals, make up the mutational load and are not usually polymorphic in the accepted sense. I have pointed out elsewhere (Wills 1973) that genes which are not of the mutational load type, but rather are of the sort that should be subject to rank-order selection, will tend to accumulate in natural populations if they arise at all, and that such genes are simply polymorphisms in which all the alleles are fully functional. An organism homozygous for such "selected functional polymorphisms" at all loci should be quite capable of surviving and reproducing in the absence of competition, but may not be capable of doing so in the presence of organisms that are more heterozygous, and therefore have a higher fitness potential.

A second limitation is that in this paper I will assume that all polymorphisms at the same frequency contribute equally to fitness potential. This is obviously not correct in detail, but if a large number of loci in the population are in fact selected functional polymorphisms, then they will tend to have approximately equal effect.

Consider a locus that is not a selected functional polymorphism by the present definition, the polymorphism for the sickle cell allele in man. The polymorphism is maintained by single-gene heterosis (and indeed forms the best example of such maintenance), but the two alleles in question are not equally functional. A homozygote for hemoglobin S suffers a considerable deficit in fitness with only a limited interaction with the environment, since a reliable cure has yet to be found. The rest of the genotype also appears to have relatively little effect.
It would not take many such loci for an insupportably large genetic burden to be placed on the population. But polymorphisms of the selected functional type should, if they exist, have no mutational load component. One can imagine such a polymorphism in which the heterozygote has a greater survival in the presence of falciparum malaria than does either homozygote. This may turn out to be the case, for example, for the widespread polymorphism of Hemoglobin E. In the absence of malaria, all three genotypes seem to survive equally well (FLATZ, PIK and SRINGAM 1965). But if malaria is present, homozygosity or heterozygosity for such a polymorphism would make a very large contribution to the organism’s fitness potential. Adding another such polymorphism, and assuming that the double heterozygote is more resistant to malaria than is either of the single heterozygotes, will result in an increased load if malaria is the only selective factor. But if malaria is only one of a large number of factors contributing to the final determination of fitness, the relative importance of these two polymorphisms to that determination may be less than that of either polymorphism by itself. As more and more selected functional polymorphisms are added in our hypothetical scheme, the relative importance of each to the survival of the organism diminishes—provided that the set of fitnesses imposed by the environment on the population does not change greatly as it would with the multiplicative model. Thus, the stepwise addition of selected functional polymorphisms will result in a very large number of polymorphisms, none of which contributes disproportionately to fitness potential or to the organism’s final fitness. The number of such polymorphisms that can be supported by the population is limited, as we will see, not by the load, but only by the overall intensity of selection and the population size, which determine the rate of loss of these polymorphisms due to drift.

I will give here the number of polymorphisms that can be maintained for long periods of time by different intensities of rank-order selection in finite populations. It will be shown that even when the selective equilibrium of the polymorphisms is near zero or one, enough polymorphisms can be maintained to account for essentially all observed and inferred variation in both the translated and the untranslated parts of the genome. Values for loci near fixation are important to obtain, because polymorphisms of this type are very common in natural populations (GILPIN et al. 1976).

Recently, modifications of electrophoretic techniques (JOHNSON 1975; COYNE 1976; SINGH, LEWONTIN and FELTON 1976) and thermal denaturation studies (BERNSTEIN, THROCKMORTON and HUBBY 1973; SINGH, HUBBY and THROCKMORTON 1975) have lent support to the view that current estimates of allelic variation may be too low. The exact extent of this most recent increase cannot yet be determined, since few loci have been examined and little precise mapping of the new variants has been carried out (but see SAMPSELL, in press). But, as will be seen, rank-order selection can in fact account for levels of heterozygosity that may be much higher than those currently observed.
How many polymorphisms can be maintained by rank-order selection?

If selected functional polymorphisms are being introduced into a population at a high rate through mutation and migration, they may be lost through drift at a high rate, and the amount of polymorphism will remain constant. But it seems probable that such polymorphisms will be introduced at a very low rate, particularly since they are likely to be less common than strictly deleterious mutations. In the present calculations, a low rate of loss is ensured by making selection so strong that the variance of the frequencies of the polymorphic alleles around their selective equilibria is very small. This also has the advantage that this variance may be treated as if it were binomial, with no great loss in precision if it is not. The rate of loss of polymorphisms will be a function of the height of the distribution at frequencies 0 and 1. Values employed in the present model have been chosen to permit reduction of this height to a very small value. This was done by making selection sufficiently strong to reduce the standard deviation of the allele frequencies to a size one-fifth of the distance to the nearer fixation point. For example, if the selective equilibrium frequency is 0.1, then the standard deviation should be no greater than $0.2 \times (0.1 - 0) = 0.02$. The area under the normal curve to the left of a point five standard deviations from the mean is about $3 \times 10^{-2}$, so that only about three in each ten million polymorphisms with this selective equilibrium will be lost during the first generation. Doctor Milkman (personal communication) points out that a very low rate of introduction of new selected functional polymorphisms will be sufficient to retain a balance between introduction and loss with this selection intensity. Haldane (1927) first showed that the probability of fixation of an advantageous mutation is approximately $2s$, where $s$ is the selection coefficient. Thus, suppose that new selected functional alleles arise at a rate of $10^{-9}$/locus/generation and that there are 100 monomorphic loci capable of becoming polymorphic at any time. Then a population of $3 \times 10^{-7}/10^{-9} \times 2s \times 100 = 1.5/s$ individuals would be sufficient to maintain the balance. Even lower mutation rates could be accommodated if the pool of monomorphic loci were larger.

Assumptions of the model

1. The population is diploid and panmictic.
2. There are two alleles at each polymorphic locus.
3. All the polymorphisms considered in a given calculation have the same selective equilibrium frequency.
4. Fitness potential is determined by the amount of heterozygosity possessed by each organism. Penetrance is complete.
5. Fitness is determined by the application of an externally generated fitness distribution to the ranked organisms in the population, so that their relative rather than their absolute fitness potentials will govern their survival.
(6) There is no linkage or linkage disequilibrium.
(7) Generations are discrete.

Definitions

$N =$ population size.

$N_s =$ population size after selection. Calculation of effective population size with truncation selection is extremely difficult. I will therefore make the approximation that $N_s = N$, the effective population size.

$L =$ number of polymorphic loci.

$p^* = (1 - q) =$ the selective equilibrium frequency of the less frequent alleles at the polymorphic loci.

$p_H = (1 - q_H) =$ that frequency of the less frequent alleles which would give the observed average proportions of homozygotes if they were all at the same frequency.

$H =$ individual heterozygosity.

$ar{H} =$ the average heterozygosity of the population (all individuals over all loci).

$\sigma^2_H =$ variance of $H$.

$\sigma^2_p =$ variance of $p$.

If all alleles are at their equilibrium, then $\bar{H} = \hat{p}(1 - \hat{p})$. Drift will reduce this value by approximately $\bar{H}/2N_s$ each generation, and selection will tend to raise it. We seek a number of polymorphic loci and an intensity of selection such that an increase in $\bar{H}$ due to selection is just balanced by the decrease due to drift, and to add the further constraint that at the balance point $\sigma_p = \hat{p}/5$.

Rank-order selection, by its very nature, presupposes no constraints on the shape of the fitness distribution imposed on the population by the environment. Various types of fitness distribution can be imagined: selection in which a fixed proportion of the population, with the lowest $H$ values, cannot reproduce; selection in which the rate of reproduction increases with the rank of $H$, and so on. A sharp truncation raises $\bar{H}$ by the largest amount for a given proportion of genetic deaths, but it is of course unrealistic to expect a sharp truncation to occur in nature. Yet other shapes of selective curve giving rise to the same proportion of genetic deaths do not markedly affect the efficiency of selection. MILKMAN (personal communication) has shown, for example, that the efficiency of selection varies by a factor of less than two for a wide variety of shapes of selective curve.

Selection under natural conditions is unlikely to involve a sharp truncation, since the environment, penetrance and population densities vary greatly in most cases. A truncation involving two thresholds will be employed in these calculations, but with the clear understanding that it is not necessary to the argument and is employed only because it makes the calculation straightforward. Some of the calculations will involve a very small truncation (5% of the popula-
tion), so that it will be seen that many other less efficient fitness distributions having the same effect on $\bar{H}$ could be imposed without involving more than a small fraction of the population.

The selection is accomplished by preventing a fixed proportion of the organisms at the low end of the heterozygosity (fitness potential) distribution from reproducing, and allowing an equal proportion at the high end of the distribution to produce twice as many offspring as the average organism. Thus, the number of offspring is equal to the number of individuals in the parental generation.

Figure 1 shows the fitness distribution of the organisms in such a selected population. This distribution may be compared in the figure with two normalized fitness distributions of populations segregating for 100 polymorphisms under the multiplicative model with $s_1 = s_2 = 0.01$ and 0.1 at each locus. The truncation is obviously much more effective because of the great range of possible fitnesses, but it must be remembered that this fitness distribution is extremely efficient at increasing $\bar{H}$.

With this type of truncation and replacement, it is straightforward to calculate the increase in $\bar{H}$ when $H$ is normally distributed. In a normal distribution, the average variate value, $\bar{x}$, of a truncated sample is equal to $f(x_i)/F(x_i)$, where $f(x_i)$ is the probability density of the variate value at the truncation threshold $x_i$, and $F(x_i)$ is the proportion that is truncated. In the present model, the truncation thresholds (for no reproduction, $H_a$, and for double reproduction, $H_b$) are placed symmetrically. Thus $f(H_a) = f(H_b)$. In units of standard deviation, the average heterozygosity is now

$$\Delta H = 2\sqrt{2\pi} \cdot \sqrt{\bar{H}(1 - \bar{H})/L} \cdot \exp \left(-\frac{H_a^2}{2}\right).$$

(1)

After random mating, $\bar{H}'$, the average heterozygosity in the next generation, will be greater than $\bar{H}$. At the same time, drift will be reducing $\bar{H}'$ by a factor $1/2N_s$.

The relationship between $\sigma^2_p$ and $\bar{H}$ is given by:

$$\frac{\sigma^2_p}{\bar{H}q} = 1 - \frac{\bar{H}}{2\bar{H}q}.\quad (2)$$

Therefore, for a given $\bar{H}$, $\sigma^2_p$ can be calculated.

Selection of this type “replaces” some of the two types of homozygotes with heterozygotes. The proportions of homozygotes present in the population at a given time can be calculated by determining $p_H$ and $q_H$, the average allele frequencies that would give the observed $\bar{H}$ if all the alleles were at these frequencies. Of course they are not; drift has ensured that there is a spread of the indi-
ORGANISMS ORDERED BY INCREASING HETEROZYGOSITY

Figure 1.—The fitness curve of organisms subject to the type of rank-order selection considered in this paper (solid line). A fixed percentage (20% in the figure) of the organisms with the greatest average heterozygosity is assigned a fitness of two, while the same percentage of those with the lowest average heterozygosity is assigned a fitness of zero. For purposes of comparison, the dotted lines show the normalized fitness curves (mean fitness adjusted to one) expected with the multiplicative model, given a population of 100 organisms segregating for 100 loci, each with selection coefficients against the homozygotes of 0.01 or 0.1. It is assumed that the heterozygosities in these latter populations are normally distributed. Selection coefficients of 0.1 per locus are required in the multiplicative model to obtain a selection curve that begins to approximate in intensity that obtained by rank-order selection. In the multiplicative model, this results in unrealistically low average fitnesses ($0.950 = 0.005$ for the $s = 0.1$ case).

If selection is towards an equilibrium frequency, $\hat{p}$, of 0.5, the ratio of the proportions of the two types of homozygotes replaced by heterozygotes is simply the individual $p$ values in the population. But for our purposes, we do not need to know the actual distribution of $p$ values, but only the average values for the homozygotes, $p^a_H$ and $q^a_H$. 
ratio of the two types of homozygotes in the population. This is because each will contribute in proportion to its frequency. If the equilibrium is not 0.5, and the ratio is \( R \), then:

\[
R = \frac{p_H^2 \hat{q}^2}{q_H^2 \hat{p}^2},
\]

which gives a ratio of 1 when \( p_H = \hat{p} \), so that the gene frequencies will not change. If \( p_H \) deviates in either direction, selection will tend to move it back towards \( \hat{p} \) and an \( R \) value of 1. Of course, this ratio is likely to apply only near the equilibrium point, since rank-order selection will result in variable selection coefficients over wide ranges of frequency. Further, it will not apply if the probability that an organism is homozygous for the rarer allele at one or more loci is less than the truncation value, since then some organisms will not be rankable and the loci will not be driven to their assigned equilibrium values. In the present examples, however, we are dealing with large numbers of loci. Further, we are imposing conditions such that selection is sufficiently strong to ensure only small deviations of \( p_H \) from \( \hat{p} \), so that selection coefficients should remain relatively constant.

Given these conditions, selection will decrease \( p_H^2 \) by

\[
\frac{p_H^2 \hat{q}^2}{p_H^2 \hat{q}^2 + q_H^2 \hat{p}^2} \Delta \bar{H} \quad \text{and} \quad q_H^2 \text{ by } \frac{q_H^2 \hat{p}^2}{p_H^2 \hat{q}^2 + q_H^2 \hat{p}^2} \Delta \bar{H}.
\]

The new value of \( p_H \) after selection, \( p_H' \), is given by

\[
p_H' = p_H - \frac{p_H^2 \hat{q}^2}{p_H^2 \hat{q}^2 + q_H^2 \hat{p}^2} \cdot \Delta \bar{H} \quad + \frac{\bar{H} + \Delta \bar{H}}{2}.
\]

The new value of \( \bar{H} \) after random mating, \( \bar{H}' \), is then calculated by \( p_H' (1 - p_H') \). At equilibrium between selection and drift, the difference between \( \bar{H} \) and \( \bar{H}' \) will be equal in magnitude, but opposite in sign, to the loss of heterozygosity due to drift, \( \bar{H}/2N_s \). Once this equilibrium value for \( \bar{H}' \) is found, \( \sigma^2 \) can be calculated from equation (2).

An iterative computer program employing these equations was written to find the value of \( \bar{H} \), for a given \( N \) and strength of selection, that would satisfy the condition \( \sigma^2 = \bar{p}/5 \). Figure 2 shows the numbers of loci that fit this criterion at each of several strengths of selection and selective equilibria.

For equilibria near 0.5, the ability of rank-order selection to maintain polymorphisms is very high. By extrapolating from Figure 2, it may be seen that for \( \bar{p} = 0.5 \), 48 million polymorphisms may be maintained by truncation and replacement of 20% of a population of \( 10^5 \) individuals, and 10 million if truncation and replacement is reduced to 5%. These numbers far exceed the maximum possible number of genes in man based on total DNA per gamete (OHNO 1972).

It is also apparent that weak selection and an equilibrium close to fixation result in a considerable loss of the ability to maintain polymorphisms, as expected.
Figure 2.—The numbers of loci that can be maintained under the condition that $\sigma_p = \hat{\rho}/5$ for different numbers of organisms, different strengths of truncation and replacement, and different selective equilibria. Note that this is a log-log plot.

Nevertheless, even with 5% truncation and replacement and $\hat{\rho} = 0.05$, as many as 66,000 polymorphic loci can be retained in a population of $10^4$ individuals. If selection is increased to 20%, this number reaches 330,000. These values can be compared with the estimate that there are 35,000 different poly-A-containing messages in a HeLa cell (Bishop et al. 1974). If this estimate corresponds roughly to the number of translated genes in a higher eukaryote, then even if all of these are polymorphic with equilibrium frequencies near 0.05, rank-order selection still has the capacity to maintain additional untranslated polymorphism. An example of such polymorphism is found in the spacer regions of Xenopus rRNA genes (Reeder et al. 1976), a function for which has not yet been determined.

Note that the moderate selection represented by 20% truncation and replacement is more effective than 50% selection, since the reduced population size in
the latter case increases the role of drift. It will be worthwhile, however, to explore the possibility that this effect is due to the nature of the truncation, and might be avoided by other fitness distributions leading to identical changes in $H$.

The chief element of uncertainty in these calculations involves the equating of $N_s$ with $N_e$. $N_s$ is, however, the minimum population size obtained each generation, and mating is at random, so that $N_e$ will exceed it, and these calculations will have overestimated the effect of drift.

MAGNITUDE OF SELECTION COEFFICIENTS

It will be seen from Figure 2 that the slopes of the lines are all two. The properties of truncation selection are such that the number of polymorphisms that can be maintained goes up as the square of the population size. Falconer (1960) showed that selection coefficients in truncation selection vary as an inverse function of the standard deviation of the distribution. This is because truncation selection increases mean heterozygosity as a direct function of the standard deviation of the distribution.

This results in an enormous gain as the number of organisms increases, and is a general property of models in which selection acts on the standard deviation. Such models have been considered by Sved (1968) and Maynard Smith (1968) from the standpoint of the substitution load. They showed that this type of selection reduces the substitution load to levels far lower than those calculated by Haldane (1957) on the basis of the multiplicative model. Maynard Smith demonstrated that with 50% truncation of an infinite population in which a frequency of 0.9 was assigned to the less favorable allele at each locus, 71,000 loci could be acted upon by a selection differential of 0.01, compared to only 77 if selection were multiplicative. If the selection coefficients are reduced tenfold to 0.001, then substitution in his formulas reveals that the permitted number of loci rises to over seven million in the truncation case, but increases only to 770 in the multiplicative case. Similarly, King (1967) showed, for an infinite-population truncation model with a rigid threshold, that reducing the selection coefficient per locus by a factor of 10 reduced the proportion of variance due to overdominance by a factor of 100.

Another advantage of truncation models is that the selection coefficients associated with each locus can be relatively large. In the present model these coefficients are given by $\Delta p_{ii}^s/p_{ii}^s$ and $\Delta q_{ii}^s/q_{ii}^s$, where the numerator is the change in $p_{ii}^s$ and $q_{ii}^s$ due to selection over one generation. For a given criterion governing the number of loci maintained, for instance our current criterion of $\sigma_p = \bar{p}/5$, these values remain constant regardless of the value of $N$ and the strength of selection. They may be compared with Kimura's limit of $1/4N_e$ (Kimura 1971), a value for selection coefficients below which alleles may be expected to behave in a way indistinguishable from neutrality. For $\sigma_p = \bar{p}/5$ and $\bar{p} = 0.5$, both selection coefficients remain constant at a level 18 times greater than Kimura's limit. When $\bar{p} = 0.1$, the coefficients associated with the rarer homozygotes are 2,000.
times the limit, and those associated with the commoner homozygotes are 45 times the limit.

These values remain constant for a given criterion for $c_\rho$, since varying the selection intensity simply permits the maintenance of proportionately more or fewer polymorphisms with the same strength of selection per polymorphism.

It is interesting to compare the magnitudes of the selection coefficients to be expected for rank-order selection and for strict multiplicative selection. In the former case, the population size after selection, $N_s$, is a function of the shape of the selection curve; in the latter, $N_s/N = W$. Since when $\hat{\rho} = 0.5$, $N_s/N = (1 - s)/(1 - H)$, the size of the selection coefficients in the multiplicative case resulting in a given $N_s$ can easily be calculated.

Consider a population of size $10^5$, subject to 5% truncation and replacement selection, which as we have seen can maintain about $10^7$ polymorphisms at $p^* = \frac{1}{5}$. Here $N_s$ is $9.5 \times 10^4$ after truncation, and $H$ is approximately 0.5. Then under the multiplicative model, the average $s$ will be $10^{-4}$, which is 250 times smaller than Kimura's lower limit. In this case, the selection coefficients obtained from rank-order selection are 4,500 times as great as those obtained from the same amount of multiplicative selection. The differential becomes even greater for larger $N$.

**DISCUSSION**

An important limitation of the model considered in this paper is that the fitness curve imposed by the environment on the population stays approximately the same regardless of the values of the fitness potentials of the organisms making up the population. This is "pure" rank-order selection, and it assumes that environmental pressures are such that only a certain proportion of organisms survives, regardless of the value of the underlying fitness potentials. It is also pure "soft" selection as envisaged by Wallace (1968). It seems likely, however, that most selection of functional polymorphisms falls between pure rank-order selection and multiplicative selection (the "hard" selection of Wallace). In such modified rank-order selection, the relative differences between fitness potentials determine to some degree the shape of the fitness curve imposed on the population. It seems reasonable to suppose, for example, that if a very great absolute difference exists between the most homozygous and the most heterozygous organism in a population, they will exhibit a larger fitness difference than they would if their heterozygosites were very similar. The extent of such an influence may be difficult to determine in practice, but in neither type of rank-order selection does the absolute fitness potential value determine fitness directly, as it does in the multiplicative model.

The differences between the pure and modified rank-order selection models are shown in diagrammatic form in Figure 3. Part A of the figure shows pure rank-order selection. The fitness curve imposed by the environment (dotted line) is the same, regardless of the absolute value of the heterozygosity. In part B, modified rank-order selection, an increase in the variance of heterozygosity may re-
Figure 3.—Diagrams contrasting pure and modified rank-order selection. The solid lines show the distribution of heterozygosities in the selected populations, ranked from lowest to highest heterozygosity (which we take to be equivalent to "fitness potential"). The upper diagram in each set shows a population with a small variance of heterozygosity, and the lower diagram shows a population with a larger variance. The ordinate, in units of fitness, refers to the selection curve imposed on each population, which is shown by a dotted line and which is dependent on the stringency of the environment. In "pure" rank-order selection, represented in A, the shape of the fitness curve imposed by a particular environment does not alter as the variance of heterozygosity (fitness potential) changes. In modified rank-order selection, shown in B, as the variance of heterozygosity increases the intensity of selection also increases, giving a higher fitness to organisms with an extremely high fitness potential. Thus, the differences between organisms on the fitness potential scale play a role in modified rank-order selection.

result, for example, in a displacement of the fitness curve to the right. Modified rank-order selection is undoubtedly much more realistic, but much more difficult to handle analytically. It will be noted, however, that modified rank-order selection is influenced by the variance of heterozygosity. Factors such as linkage disequilibrium that change this variance will be considered in another paper.

The estimates derived for the numbers of polymorphisms that can be retained should be clearly understood to apply only to selected functional polymorphisms, since other types of genetic variability will not have these properties. I have not yet succeeded in extending the model to more than two alleles at a locus, but preliminary simulations suggest that the number of polymorphisms maintained will not change markedly as the numbers of alleles increase. The model is also limited
at the moment by the requirement that in each calculation all the loci be driven
to the same selective equilibrium.

Finally, the calculations employed here are based on the assumption of un-
conditional heterosis at every locus. An equilibrium population subject to such
selection should have a distribution of allele frequencies very different from
those actually observed in natural populations. Rank-order selection can main-
tain 150 times as many heterotic polymorphisms at an equilibrium frequency of
0.5 as can be maintained at a frequency of 0.05. Thus, if polymorphisms with a
variety of internal equilibria are constantly arising, the overwhelming majority of
those that are retained should be those with internal equilibria near 0.5.

In fact, the distribution of allele frequencies in natural populations is U-
shaped, and therefore the mirror image of what these calculations might lead us
to expect. Indeed, this distribution of alleles forms the basis for powerful neutral-
ist arguments (YAMAZAKI and MARUYAMA 1972). If, on the contrary, the ma-
jority of polymorphisms consist of functional alleles that are subject to selection,
then the selection must be of a type that leads to an allele distribution mimicking
the U-shaped one expected for neutrality. GILLESPIE (1977) has recently shown
that this expectation is fulfilled for polymorphisms subject to randomly fluctu-
ating selection pressures in which the heterozygote has approximately intermedi-
ate fitness. Such polymorphisms are not being driven towards a single equi-
librium. It may be that a U-shaped distribution can also be derived from other
selective models, such as for example models in which different types of selection
are mixed. More information is therefore needed on the type of selection acting
to maintain the majority of polymorphisms in natural populations, particularly
with regard to the strength of genotype-environment interactions and the extent
to which polymorphisms exhibit partial dominance. Despite these difficulties,
there appears to be no insuperable obstacle to the maintenance of most genetic
variability by selection.

I thank ROBERT A. ARMSTRONG, MICHAEL GILPIN and MICHAEL SOULé for discussions.
Special thanks go to ROGER MILKMAN for endless patience and help. Supported by grant
GM 19967 from the Public Health Service.

LITERATURE CITED

BERNSTEIN, S., L. H. THROCKMORTON and J. L. HUBBY, 1973 Still more genetic variability in

BISHOP, J. P., J. G. MORTON, M. ROSBASH and M. RICHARDSON, 1974 Three abundance classes
in HeLa cell messenger RNA. Nature 250: 199-204.

COYNE, J., 1976 Lack of genic similarity between two sibling species of Drosophila as revealed
by varied techniques. Genetics 84: 593-607.


FELSENSTEIN, J., 1976 The theoretical population genetics of variable selection and migration.

FLATZ, G., C. PIK and S. SINGAM, 1965 Hemoglobin E and the β-thalassemias: their distribu-


Corresponding editor: R. W. Allard