VARIATION IN FITNESS AND MOLECULAR EVOLUTION

W. F. BODMER
UNIVERSITY OF OXFORD
and
L. L. CAVALLI-SFORZA
STANFORD UNIVERSITY

1. Introduction

Molecular studies, especially of proteins and nucleic acids have added important new insights into evolutionary processes by providing new ways of investigating and measuring evolutionary rates over long periods of time. In particular, the estimation of the mean time necessary for an amino acid substitution (Zuckerkandl and Pauling [30]) has rightly generated much interest and has given considerable stimulus to further investigation into the mechanisms of evolution.

There seems, at the present time, to be substantial disagreement as to the meaning of the quantities observed and their interpretation in evolutionary terms (see, for example, Kimura and Ohta [16] who give citations to the relevant literature). Specifically, the analysis of data on molecular evolution has led to a revival of the old controversy concerning the relative roles in evolution of random genetic drift and selection.

In this paper, we shall extend some considerations that were made in a book that appeared recently. We shall also review some experiments on computer simulation of molecular evolution that were done some two years ago, and also review the molecular evidence from a variety of sources and organisms concerning the roles of random genetic drift and selection in evolution. The model of molecular evolution which we have used for computer simulation was designed to evaluate mean evolutionary time, both for neutral mutations and also for mutations which have an effect on fitness. It also provides an estimate of the extent of polymorphism for a given locus at any given time.

2. The computer model

Since the number of possible changes in a protein molecule is very large, we have used, as have others, a model in which every allele of a gene that can be

These experiments were supported in part by grant number GM10452-09 of the NIH and grant number GB 7785 of the NSF and AT(04-3)326PA33 of the USAEC.

produced by mutation is a new one, so that in practice there is an infinite number of alleles. This is very close to what is observed in molecular evolution, since with a protein of 100 amino acids and the possibility of twenty amino acids at each site, there are 20^{100} possible types, plus all other changes which do not involve a simple amino acid substitution. Many of these will, of course, be nonviable, but the number which are viable may still be very large.

A haploid population is used for our model, as is usually the case for genetic drift theories. The extension to diploids is easy as long as fitness is considered to be additive with respect to genotypes. The population is kept at a constant size N and mutation is allowed to occur with a constant rate μ per generation. Every new mutant is different. When fitness is allowed to vary, the mutant will have a fitness which may be different from that of the allele in which the mutant arises. The fitness of the mutant is assigned according to a chosen distribution of fitness values.

In our experiments, the fitness distribution was taken to be normal with arbitrary standard deviation σ_w and with a mean equal to that of the allele in which the mutation took place, plus a constant quantity Δw, which is zero if the average fitness of mutants is equal to that of the parental type. Checks were imposed to avoid negative fitness values. In such a system, one can, therefore, produce advantageous deleterious, neutral, or quasi neutral mutations in the desired proportions. All individuals present in the population were allowed to reproduce according to a Poisson distribution. The next generation was thus formed by giving to each type represented in the former generation an expectation of progeny equal to the number of individuals of that type times its fitness, and letting a Poisson variate represent its number of progeny. When the expectation computed in this way was above 20, then the computation of the number of descendants was simplified by replacing the Poisson distribution by a normal distribution having mean and variance equal to the expected number of progeny of that type. Under these conditions, the total number of individuals in the next generation also varies approximately according to a Poisson distribution with expectation N. In order to keep N constant, the realized population size was adjusted to its constant value by adding or eliminating individuals of the various types at random, that is, taking into account only the proportions of the various types. The constancy of N is a requirement which nearly always creates difficulties when setting up mathematical models. The program was adapted for the exact treatment (with a multinomial distribution) by Harry Guess and found to give undistinguishable results from those obtained with the above procedure. For N not very small the multinomial simulation requires more computer time than the Poisson approximation. In fact, it makes the computer time proportional to N (times the number of generations) while with the Poisson approximation the computer time is proportional to the number of alleles present, which is a function of the product $N\mu$ (times the number of generations).

We are grateful to Harry Guess who pointed out an error in the computer program used in the simulation.

3. Some results of the model

Table I shows an example of a simulation with $N = 30{,}000$ and $\mu = 10^{-5}$. The fitness w of the original type at the beginning of the experiment was 1. The variation in fitness had a standard deviation σ_w of 0.01 and the average decrease in viability of new mutants Δw was = 0.01. Newly produced mutants were thus mostly deleterious, having on average a fitness which was one standard deviation below the fitness of the type in which they were produced. But because of the normal distribution of fitness values, about 15 per cent of new mutants had fitnesses which were higher than that of the parental type. The table shows the composition of the population at various times. Each mutant is identified by its fitness as well as its birth date, which is the generation in which it arose. Each mutant is also associated with a count of the *number of mutational transitions* which it has undergone since the beginning of the experiment. Thus if, for example, a new mutation arises in an allele produced by a mutation from the allele which was present in all individuals in the original population, this has undergone two mutational transitions, and so on. This quantity, the number of

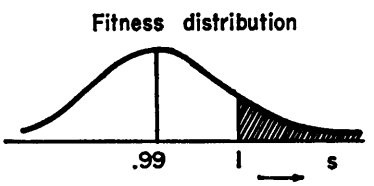
TABLE I

AN EXPERIMENT OF EVOLUTION BY COMPUTER SIMULATION

$N = 20{,}000$, $\mu = 10^{-5}$

Each column refers to one of the mutant alleles present in the population at the time given. There are as many columns as alleles.

Mutant born at generation 1,320 was fixed by generation 3,750.



Generation 1,000							
No. individuals	19,833	5	162				
Fitness	1	0.9962	0.9987				
Birth date	1	610	809				
No. mutational transitions	0	1	1				
Generation 2,000							
No. individuals	16,794	3,206					
Fitness	1	1.0036					
Birth date	1	1,320					
No. mutational transitions	0	1					
Generation 10,000							
No. individuals	10,521	8,151	1,024	274	23	5	2
Fitness	1.0244	1.0221	1.0176	1.0147	1.0147	1.0140	1.0412
Birth date	8,859	7,887	9,783	9,965	9,965	9,982	9,997
No. mutational transitions	4	3	4	4	5	4	5

mutational transitions, had to be introduced in order to deal with the statistics of evolutionary rates. The original purpose of the simulations was to estimate the time taken to fix new mutations. It soon became evident, however, that unless the mutation rate was much lower than the reciprocal of the population size, no mutant, or at least very few mutants, ever really became fixed.

The general consequences of this model, which seem quite close to reality, were rather that there are usually several alleles present in a population which may have undergone different numbers of mutational transitions from the original allele, which was assumed to have a frequency of 1 at time zero. The mean number of mutational transitions for the alleles present in a population can be calculated at each time point. The time taken for this mean number to increase by 1 is the reciprocal of the rate of gene substitution. The mean evolutionary time estimated from amino acid substitutions should correspond to this number. In fact, the number of amino acid differences between two proteins is, assuming an almost infinite number of alleles, proportional over a wide range to the number of mutational transitions. The proportionality constant is somewhat less than one because of reverse mutation (a rare event), the complications arising from the degeneracy of the genetic code, and other sources.

Part of the experiment shown in Table I is illustrated in Figure 1. Here the mutation rate is less than $1/N$ and the *effective* mutation rate, that is, the rate of production of mutants that have a fitness above neutrality, is very low, being about $1/30$ of $1/N$. In this example, a few mutants do get fixed. Two were actually fixed during the first 10,000 generations (see Figure 1). A third mutant was not fixed because at the time its frequency was approaching 100 per cent, it was supplanted by a new mutant with a higher fitness that had meanwhile developed from it.

That very few mutants ever get fixed, is more clearly illustrated in Figure 2, which gives an experiment with the same values of N and μ as before, but with only neutral mutations. Because there are no fitness differences, there are considerable short term fluctuations in the frequencies. Only one of the many mutants indicated in the figure became fixed. The frequency of this particular mutant, which underwent two mutational transitions, is also indicated in its descent phase to emphasize its long persistence in the population. It should also be noticed that around generation 7,500, for instance, mutants that differ by more than one mutational substitution may be present with appreciable frequencies at the same time, in one population.

This suggests that the variance of the number of mutational transitions undergone by mutants present in a given population at the given time may be an indication of the evolutionary forces at work. Our simulation experiments are still inconclusive on this point, but it may be worth remembering that Prager and Wilson [23] reported the coexistence in a population of two alleles differing by at least six mutational transitions.

Table II gives data from another experiment in which the variation of fitness was so small that most mutations can be thought of as almost neutral ("quasi

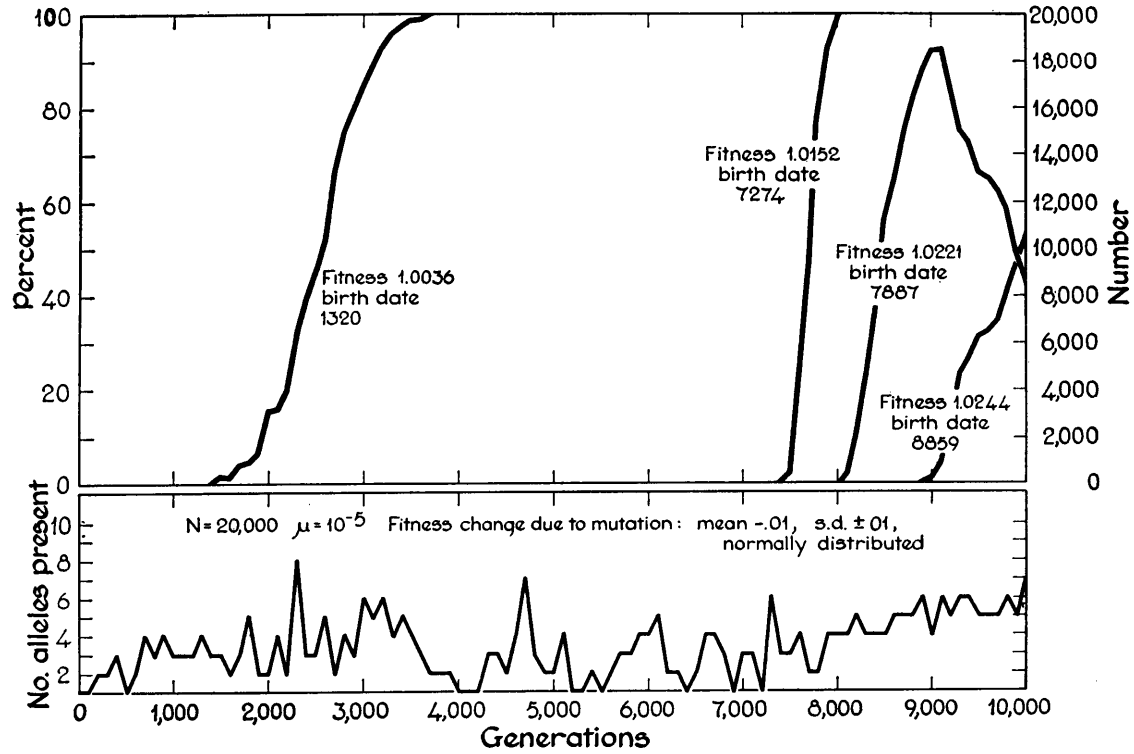


FIGURE 1

The experiment of Table I plotted.

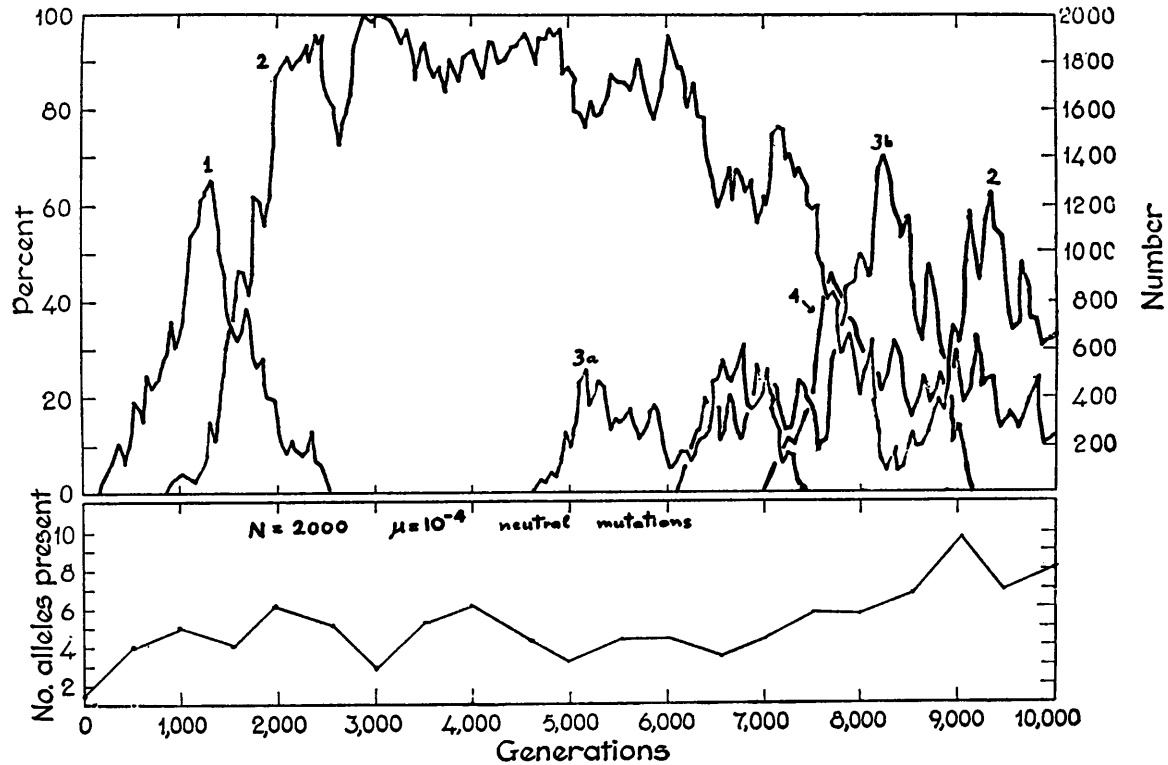


FIGURE 2

An experiment with neutral mutations.

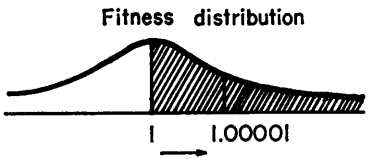
The numbers 1, 2, 3a, 3b, and 4 indicate mutations, sampled because they reached substantial frequencies whose histories are traced throughout all generations. One of them (2) was fixed for a short time and was the only one to reach fixation.

TABLE II

COMPUTER SIMULATION OF EVOLUTION WITH QUASI NEUTRAL MUTATIONS

$N = 2{,}000$, $\mu = 10^{-4}$

Each column refers to one of the mutant alleles present in the population at the time given. There are as many columns as alleles.



Generation 10,000				
No. individuals	1,959	38	3	
Fitness (minus 1, % s.d.)	0.0	2.9	1.14	
Birth date	5,617	9,694	9,949	
No. mutational transitions	1	2	2	
Generation 20,000				
No. individuals	1,108	724	134	4
Fitness	0.0	0.67	1.72	1.14
Birth date	5,617	18,360	19,764	19,990
No. mutational transitions	1	2	2	3
Generation 30,000				
No. individuals	1,908	72	16	4
Fitness	2.19	2.0	3.24	2.38
Birth date	29,866	27,844	29,991	29,998
No. mutational transitions	5	6	6	6
Generation 40,000				
No. individuals	1,676	241	42	41
Fitness	2.19	2.67	1.43	2.48
Birth date	36,469	38,609	39,021	38,249
No. mutational transitions	8	9	9	9

neutral" following Kimura's definition). In such experiments, the mean evolutionary time is close to that expected for neutral mutations, but a small increase in fitness is observed and "positive" mutants are eventually preferred; thus, even if the advantages are very small, they cannot be neglected. From observations obtained from a number of similar experiments, it appears that the mean fitness increases by an amount that tends to be smaller than that expected, the smaller the variation in fitness is with respect to $1/N$. In other words, an increase in the relative importance of drift decreases the expectation of the rate of increase in fitness. One might thus visualize a possible generalization of Fisher's fundamental theorem of natural selection which included terms that represent a reduction in the expected rate of increase of fitness due to drift.

Some data on the mean observed number of substitutions and other quantities of interest obtained in various experiments are given in Table III. The mean substitution time was computed by dividing the number of generations the experiment was run by the mean observed number of mutational transitions (NMT). The first 1,000 generations were not included to avoid possible effects of initial conditions. All populations are started at time zero with only one type. Standard errors of NMT and other quantities are computed on the basis of the

TABLE III

RESULTS OF SOME COMPUTER EXPERIMENTS SIMULATING MOLECULAR EVOLUTION
The number of mutational transitions (NMT) is given per 1,000 generations, and its expectation for neutral changes ($\sigma_w = 0$) is $1{,}000\mu$. The mean substitution time is 1,000/NMT.

Population size N (haploid)	Mutation rate μ	Variation of fitness	NMT ($\times$ 1,000 gen.) obs.	exp.	Mean substitution time (generations) obs.	exp.	Mean F obs.	exp.	Average no. of alleles
100	0.01	$\sigma_w = 0$ (neutral)	10.01 ± 1.04	10	99.9	100	.368	.333	6.7
100	0.003	$\sigma_w = 0$	$3.36 \pm .29$	3	297.6	333	.656	.769	3.2
100	0.001	$\sigma_w = 0$	$0.85 \pm .16$	1	1,176.5	1,000	.879	.833	1.55
500	0.01	$\sigma_w = 0$	$11.44 \pm .86$	10	87.4	100	.102	.091	33.67
100	0.01	$\sigma_w = 0.05$	13.38 ± 1.77		74.7		.389		7.1
100	0.01	$\sigma_w = 0.02$	14.10 ± 1.29		70.9		.366		7.4

variation of estimates of NMT obtained every 1,000 generations (from 9 to 22 such observations for each mean). In general, the number of mutational transitions is found to be equal to expectation; that is, equal to $1/\mu$ and independent of N for neutral mutations (Kimura [14], Cavalli-Sforza and Bodmer [6]). It is higher when selection is involved ($\sigma_w > 0$, the last two lines of Table III) even though in the experiments presented in Table III ($\Delta w = 0$) half of all the mutations have fitness lower than the parental type and are constantly discarded.

The mean F value ($\sum p_i^2$, where p_i is the frequency of each existing mutant) corresponds well to its expectation $1/(1 + 2N\mu)$ (see Kimura and Crow, [15]), where we have $2N\mu$ instead of $4N\mu$, the population being haploid. It was observed, however, that F values have an extremely high variance. This corresponds to expectation according to theoretical work (unpublished) by Ewens. Also the average number of alleles observed is given in Table III.

4. Form of the fitness distribution

Two examples of approximate distributions illustrating the variation in fitness of new alleles, assumed in our computer model are shown in Figure 3. In both cases the majority of mutations are deleterious. Such mutations practically never get fixed unless the population is extremely small, and so can safely be neglected. Thus, the mutation rate that must be considered is that to advantageous and neutral mutations. The latter are shown in the figure as corresponding to the approximate range $1 \pm 1/2N$. Our experiments confirm the prediction by Kimura that, when the variation in fitness is of this order of magnitude, the mean number of transitions is practically the same as that observed with strictly neutral mutations. In the upper distribution the fraction of advantageous mutations which cannot be considered neutral is relatively large, while in the lower distribution it is small. The lower distribution, therefore, corresponds more

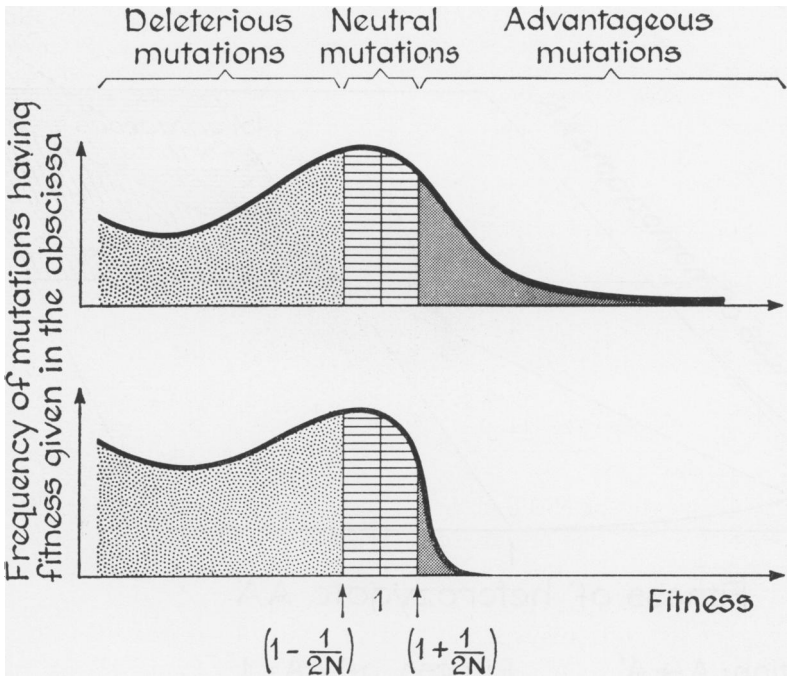


FIGURE 3

Approximate distributions illustrating the variation in fitness of new alleles.

closely to the model suggested by Kimura for molecular evolution, in which most mutations are neutral or almost so.

The picture suggested in Figure 3 is of course an over simplification which can at best be valid for haploids. In diploid organisms, the situation is further complicated by the fact that we must consider the fitnesses both of the homozygote and of the heterozygote. Figure 4 shows a suggested distribution of fitness values in mutant homozygotes and heterozygotes, taking the fitness of the normal homozygote as equal to 1. It is perhaps reasonable to assume that most mutations will be distributed around the line indicating additive fitnesses, that is, the situation in which the homozygote has a fitness $1 + 2s$ when the heterozygote has fitness $1 + s$. The distribution indicated in the figure has, for illustrative reasons, a variance which is much larger than appropriate for the actual distribution. There is actually little, if any, data for individual mutations that can be used to give this distribution.

Perhaps the best indication from published data is given by observations of Dobzhansky, Holz, and Spassky (see Hadorn [11] and Figure 5). These data, however, refer only to homozygotes and not to heterozygotes and so give only one marginal distribution that would be obtained from the surface given in

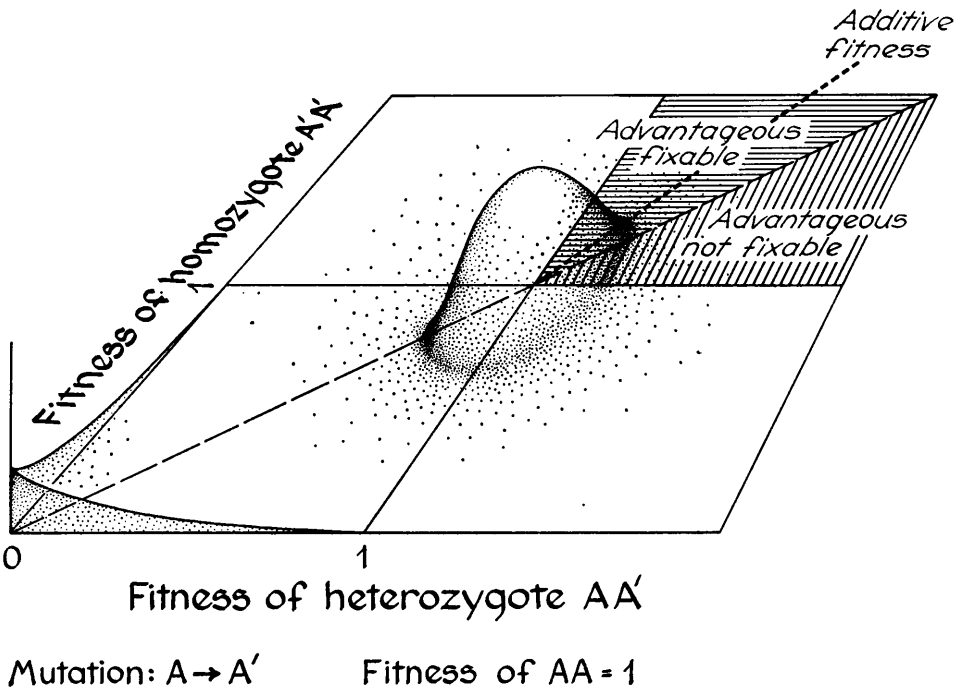


FIGURE 4

Distribution of fitness values in mutant homozygotes and heterozygotes, taking the fitness of the normal homozygote as equal to one.

Figure 4. The viabilities computed by these workers were for entire chromosomes. Therefore, they refer to the sum of a large, unknown number of different mutations located on these chromosomes. The standard deviation for fitness in the part of the distribution which peaks around normal fitness is approximately 0.05. This may correspond to the sum of hundreds, possibly thousands or more, of different mutations that were heterozygous in the population that was analyzed. It may be, therefore, that the average fitness of each of the individual mutations is exceedingly small so that a large fraction of them lie within the range $\pm 1/2N$ of quasi neutral mutations. These are not, however, *new* mutations, but a sample of mutations that has already been tested by natural selection, because they have been found in wild populations. A distribution which may be closer to that appropriate for new mutations was given by Käfer (see [11]) who studied X-ray induced mutations. The fraction of deleterious mutations is then increased, but the general shape of the distribution remains the same as that shown in Figure 5. This is, perhaps, surprising because in the irradiation experiment only a relatively small number of mutations should be induced on each chromosome. This type of observation is, however, subject to a large experimental error which may

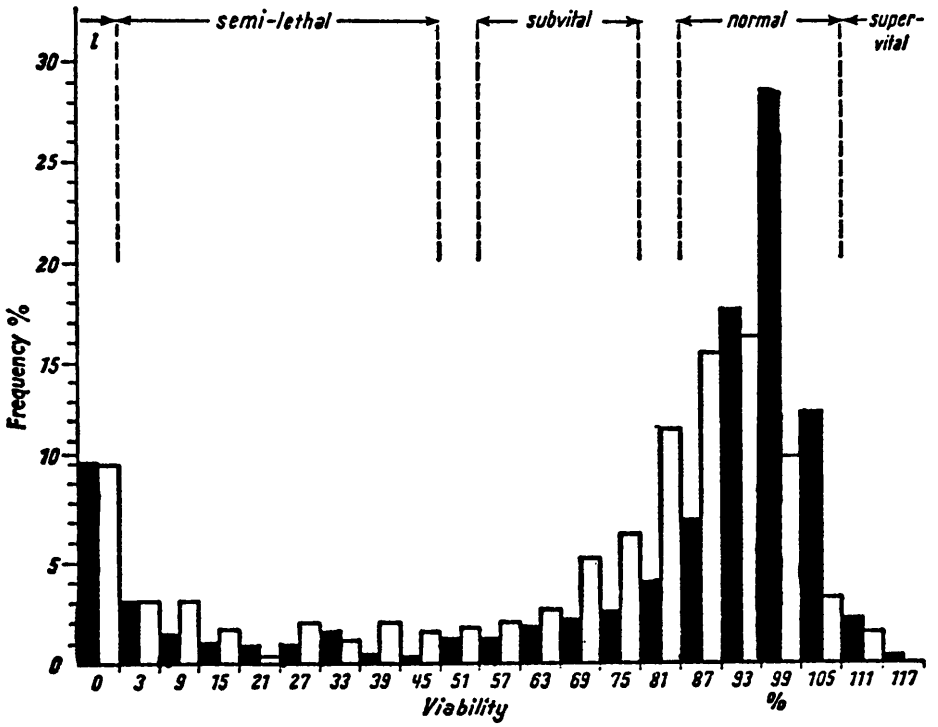


FIGURE 5

Distribution (possible) for individual mutations (from [11], page 118). Viability spectra for factors from wild populations of *Drosophila pseudoobscura*. Black = distribution of relative viabilities of homozygotes for 326 second chromosomes, white = the same for 352 fourth chromosomes, *l* = lethal. (Recalculated and illustrated after data by Dobzhansky, Holz, and Spassky, 1942.)

obscure the actual variation in fitness values. Small fitness differences are extremely difficult to measure, especially in higher organisms, and it is very difficult to measure satisfactorily fitness differences that are less than 0.01 (see below). If many mutations have fitness differences less than 0.01, the problem of estimating the distribution of fitnesses associated with new mutations, especially the part that matters for the present discussion, may be exceedingly difficult.

Even if it were possible to obtain actual data giving the distribution surface illustrated in Figure 4, it would still have to be remembered that this surface would refer to a specific environment. The variety of environments with which an organism might be confronted would complicate the interpretation of such surfaces still further. Most organisms of course live in a great variety of environments that are heterogeneous in time as well as space, even perhaps over quite

small distances. Fitnesses estimated in natural populations, however, for example in man, do generally represent average values that may be valid over a wide range of different environments.

5. The fitting of theory to observation

There are three major observable evolutionary quantities which have to be explained by our models. The first, which we have already discussed extensively, is the mean rate of gene substitution or the mean time taken for the average number of evolutionary transitions to increase by one. This is estimated from data on amino acid substitution. The second is the observed degree of polymorphism. This can be expressed in a variety of ways such as the overall fraction of time during which a gene is polymorphic, or $1 - F$, where F is the overall frequency of homozygotes for the gene in question, or also the mean number of alleles present at a given time. The mean number of alleles and the F value can be estimated from data on electrophoretic variation for enzymes which can be stained or otherwise identified on gels following electrophoresis. This procedure permits us to study unselected loci but has the disadvantage that it underestimates the number of existing alleles by a factor which may be one third and possibly higher. In fact, only one third of amino acid substitutions give rise to observable electrophoretic changes. It is also possible that changes detectable by electrophoretic techniques may be more usually subject to selective pressure than mutational changes which do not determine a charge difference and are, therefore, usually not detectable by electrophoresis. The third observable is the variation between different populations in different environments in the level of polymorphism for a given locus. This is usually expressed as the variance of the gene frequencies from the various populations. For existing theories to be applicable to the data, effective migration rates between the populations must be neglibible, or at least their intensity should be known.

Our computer model is based on four main parameters: N the population size, μ the mean mutation rate per locus, Δw the mean difference in fitness between a new mutant and its immediate ancestor, and σ_w the variance of the distribution (assumed normal) of the fitnesses of new mutants. The problem, in principle, is the estimation of these four parameters, if possible, from data on the three major observable evolutionary quantities. The issue, for example, that has been raised by Kimura [14], by King and Jukes [18], and by others is whether the observables are compatible with values of Δw and σ_w inside the range $\pm 1/2N$. Since the population size and mutation rate can in principle be estimated using quite different sorts of information from that we are considering, there should be adequate scope for estimating Δw and σ_w and even for testing the goodness of fit of the model using the third degree of freedom in the observables.

There are, however, at least two major complicating factors in this apparently simple approach. The first is that there is no universal agreement on what are the appropriate values for N and especially for μ. The second, and perhaps more

important, is that a single normal distribution with parameters Δw and σ_w is not enough to describe adequately the distribution of fitness values for new mutants. Apart from anything else, as already pointed out, this model can only apply to diploid organisms on the assumption of additive fitness values.

The important features of the distributions illustrated by Figures 3 and 4 are the proportions of deleterious, neutral, heterotic, and fixable alleles. The heterotic and fixable parts of the distribution can be further subdivided according to whether they apply to all environments or only to some environments. This distinction is especially important in the consideration of observed variations in the level of polymorphism, when different populations are compared (our third observable above). If we characterize the distribution of fitness values of new mutants by these six subdivisions (equivalent to considering six different mutation rates according to the fitnesses of the newly derived genotypes), we have, with N, seven rather than only four parameters for our theoretical model. We may not, however, even with six parameters, have adequately catered for variations in the environment changing the shape of the fitness distribution. Even accepting independent estimates of N and μ (the overall average mutation rate), we are now no longer in a position to be able to estimate from observed data, all the parameters of the model, let alone test the goodness of fit. The best that can now be done is to see whether the observed data rule out any significant regions of the parameter space defined by the values of N, μ, and the describers of the fitness distribution. A schematic summary of the effects of increases in the seven parameters defined above on the three major observable evolutionary quantities is shown in Table IV.

We shall now review briefly published data on three major observable evolutionary quantities starting with variations in the level of polymorphism between different populations. Apart from man, the best studied mammal is the mouse. A paper by Petras, Reimer, Biddle, Martin, and Linton [22] has shown that relatively unrelated populations of *Mus musculus* can show quite similar distributions of polymorphisms. This is more in agreement with selectively balanced polymorphism than with neutrality of the mutants present in a population. Little is, however, known about migration in the mouse so that populations that seem to be widely isolated geographically may in fact be more interconnected by migration than one might expect *a priori*. If this were true, the similarity of polymorphism found at a great distance might also be compatible with the theory of neutral mutation. The authors of this study also mention the possibility that the observed similarity of polymorphisms in widely separated geographical isolates may represent transient polymorphism due to selection following the introduction of new pesticides.

Prakash, Lewontin, and Hubby [24] have found even more extensive similarities in the polymorphism exhibited by many loci in *Drosophila pseudobscura* from quite different geographical origins. Here, again, population sizes, mutation rates, and migration rates are generally not well known, though the similarity in the distribution of polymorphisms encountered in widely separated localities is

TABLE IV

EFFECTS OF INCREASES IN SEVEN PARAMETERS ON THREE OBSERVABLE EVOLUTIONARY QUANTITIES

See text for further explanation.
Parentheses indicate effects are limited to some environments.

Parameters (which increase)	Observable evolutionary quantity		
	Mean evolutionary time	Average level of polymorphism	Variation in level of polymorphism
N	increase (only in presence of selection)	increase	no effect
Mutation rate to deleterious alleles	no effect	no effect	no effect
Mutation rate to neutral alleles	decrease	increase	no effect
Mutation rate to heterotic alleles:			
In some environments	(some contribution)	(increase)	increase
In all environments	small contribution	increase	decrease
Mutation rate to fixable alleles:			
In some environments	(decrease)	(increase)	increase
In all environments	decrease	increase	no effect

certainly surprising. It would be difficult not to conclude with the authors that the simplest explanation is that polymorphisms showing such a remarkable similarity in the frequency of the various genes in different populations represent the consequence of balancing selection. The identification of an allele purely on the basis of electrophoretic mobility is not, however, generally sufficient, and identity should be shown by further molecular analysis. A number of hemoglobins previously believed to be identical on the basis of identical electrophoretic mobility were later shown to be different alleles when fingerprinting and sequencing were carried out. It should also be emphasized that it may be very hard to distinguish the direct selective effects of an identifiably polymorphic locus from those of other so far unidentified but closely linked loci. Weak selective interaction between closely linked genes may make an important contribution to the overall maintenance of polymorphism (see, for example, Bodmer and Parsons [4], Bodmer and Felsenstein [3], and Franklin and Lewontin [10]). Even in the absence of selection, close linkage to a selectively maintained polymorphic locus can also in finite populations contribute to the overall level of polymorphism (see, for example, Sved [26], [28]). The results presented by Ayala at this conference extend considerably the range of the original observations by Prakash, Lewontin, and Hubby [24], but do not alter the conclusions above.

In man, the average frequency of polymorphisms is similar to that so far observed in other species. Population sizes and migration rates are, on the whole,

more easily ascertained in man than in other species. This makes it possible to compare the observed level of geographic variation of polymorphisms with that expected on the basis of relevant demographic quantities. The migration matrix method (Bodmer and Cavalli-Sforza, [2]) has been used in various studies of rural populations from various parts of the world (partly unpublished, see [6]). This approach allows one to compare observed with expected variation in gene frequencies for given migration rates and population sizes. In all these cases the observed variation, computed as an f value (variance of gene frequencies divided by $\bar{p}\ (1 - \bar{p})$, where $\bar{p}$ is the mean gene frequency) is in "semiquantitative" agreement with that expected under the balance of drift and migration and in the absence of selection.

These results thus suggest selection played a minor role in generating the observed variation between populations. In each case, however, only variation at a microgeographic scale was measured. The studies were also based on areas selected to have low population numbers or lower migration and thus relativey stronger drift effects so that they cannot be considered to represent the species as a whole. When variation is analyzed at a wider geographic level—for example, by comparing broad ethnic groups, then the effect of selection becomes apparent. The criterion used is a simple one. If drift alone were responsible for the observed variation, then every locus should show the same amount of variation in gene frequency between populations. Thus, we know that for genes that are polymorphic, or more precisely, that are not maintained by the balance of mutation and selection under drift alone, f should be the same for all genes, being a function only of N and of migration rates. The observed f values in interracial comparisons vary greatly from gene to gene (over a range of at least 10 fold, see Cavalli-Sforza [5]). This clearly suggests that selection is operating at this level of comparison. Selection may be disruptive for genes having relatively high values of f, in which case the genes are responding differently to selection in different environments. Selection may, on the other hand, be balancing for those genes giving low f values. In this case, similar balancing selection in different environments is presumably reducing the level of variation in comparison with that expected from drift alone. Unfortunately, however, the analysis of interracial variation cannot yet be carried to the level of comparing observed with expected f value, as in the case of the analysis of microgeographic variation. This is because we know too little about the demographic conditions that prevailed during the formation of races and this information is needed to compute the expected values of f.

On the whole, these analyses of the variation in polymorphic gene frequencies between different populations in mouse, *Drosophila*, and man, do suggest the existence of detectable differences due to selection.

Let us now consider the data derived from amino acid sequences on the rate of gene substitution which lead to a comparison of the observed and expected rate of evolution under different assumptions. We want values of N and μ, the latter possibly subdivided according to the selection effects of the mutational

change. In general, the relevant value of N depends on the population which is sampled. For molecular evolution this is the whole species and so at the upper limit of possible values of N. In all the cases of variation discussed so far, such as interracial comparisons, or the analysis of variation at a microgeographic level, the values of N involved were smaller as implied by the populations being sampled. We will limit our discussion to man as this is the species for which this quantity can be estimated most satisfactorily.

We should, of course, not consider the present world population as the basis for evaluating N for man. Very large increases in population size have occurred just during the last 10,000 or so years, that is, since the domestication of plants and animals has augmented the carrying capacity of the land for man. Most of our evolution, however, took place before this, while man was still a hunter and gatherer. The relevant estimates of population size which have been suggested, for example, 125,000 by Deevey [9], seem far too low. Today, there are still people who live with a hunting and gathering economy, such as, for example, the African Pygmies. These alone number over 100,000 and occupy a very small portion of the African continent, at a density of about 0.2 per Km^2 (see [6]). On this basis, a minimum estimate of the total human population size throughout the Paleolithic must be of the order of 10^6 to 10^7. Reduction of N to N_e the effective population size, involves two factors: (1) overlapping generations, which reduces N by a factor of about one third [6] and (2) isolation. With respect to the latter, a theorem by Moran [21] states that, if a population of N individuals is separated into k groups amongst which exchange of individuals takes place, and each group receives from the other groups k individuals per generation, then the effect of the subdivision on the drift experienced by the population as a whole is practically negligible. That is, the effective size of the whole group is still close to N. It would seem that the effective size of the human population as a whole should therefore, not be taken as less than 10^5 and is probably nearer to 10^6.

Mutation rates have been estimated in man using pedigree data and mutation-selection balance theory, but an important source of bias in these estimates has apparently so far been overlooked. Average published mutation rates are generally about 3×10^{-5} per gene per generation. These estimates, however, generally ignore the fact that mutations at the particular locus for which they were derived were known to occur before they were studied. This implies that the particular loci studied must have been selected at least to some extent on the basis of their mutation frequency. A simple statistical computation shows that this can lead to a considerable bias in the estimated mutation rate. If one assumes, as a first approximation, that the probability of a mutation being included in a survey is proportional to its mutation rate, it can be shown that the unselected average mutation rate is equal to the harmonic mean of the observed selected mutation rates [6]. The results of the calculations show that the average mutation rate, because of the extreme variation in mutation rates between

different loci, is 3×10^{-7}, two orders of magnitude lower than the values given before.

These mutation rate estimates in man refer only to deleterious alleles. The proportion of all mutations that are deleterious in man is not known though attempts have been made to estimate it in other organisms. It at least seems unlikely that the order of magnitude of the mutation rate to neutral and to advantageous alleles is higher than that to deleterious alleles.

Data on amino acid differences between proteins of different species suggest a median rate of evolution corresponding to 10^{-9} amino acid substitutions per year per amino acid position [30], [18]. In other words, the average time between amino acid substitutions at a given position in a protein is 10^9 years. When multiplied by three, to allow for the fact that three nucleotide pairs are needed to code for one amino acid, this gives 3×10^9 years as the mean time taken for the number of mutational transitions, as given by our computer model, to increase by one. As already discussed, the mean expected rate of gene substitution per generation, assuming only neutral mutations, is the mutation rate μ. Since the amino acid substitution data comes mainly from mammals, the relevant generation time should be an average for mammals, which can reasonably be taken to be four years. The molecular data thus suggests a mutation rate of $4/3 \times 10^9$ or 1.3×10^{-9} per nucleotide pair per generation, on the assumption that all or most mutations are neutral. If we assume that the mutation rate to neutral alleles is equal to that to deleterious alleles, and that there are on average about 1,000 nucleotide pairs per gene, then using the mutation rate estimate to deleterious alleles of 3×10^{-7} per gene, we obtain a neutral mutation rate per nucleotide pair of $3 \times 10^{-7}/1{,}000 = 3 \times 10^{-10}$. This is three times less than that suggested by observations on amino acid substitutions assuming neutrality of all mutations. At face value, this would argue against the idea suggested by Kimura, King and Jukes, and others, that most observed amino acid substitutions are due to neutral or quasi neutral mutations. However, the fact that Kimura can come to an opposite conclusion, using similar arguments and published data should stand as a warning against taking these numerical data too seriously as evidence either for or against neutrality. The figures involved are known with insufficient accuracy to make precise statements.

Consider now the situation when there can be both neutral and advantageous mutations. Assume that a proportion p_n of all mutations are effectively neutral (that is, lead to fitness differences in the range $\pm 1/2N$) and a proportion p_a are advantageous, that is, lead to fitness differences greater than $1/2N$. Since there will also be a fraction of mutants that are deleterious,

$$p_n + p_a < 1. \tag{1}$$

For the neutral mutants, the rate of gene substitution is simply obtained from the mutation rate to neutral changes, μp_n. For the advantageous mutants, the rate will be $k\mu p_a$, where k, a factor greater than one, represents the average

effects of selection on the rate of gene substitution. The overall rate of substitution, taking into account both neutral and advantageous mutants, is therefore given by

$$M = \mu(p_n + kp_a) \tag{2}$$

(see [6]). Clearly, M can be much greater than μ (even by a factor of ten or more) depending on the magnitudes of k and p_a, that is, depending on the distribution of fitness values among mutants. Thus, even reducing the number of variables from seven in Table IV to a minimum of three, as we have now done, the expected mean evolutionary times, based on population genetic models, are compatible with practically any reasonable observed rate of evolution.

The order of magnitude of N in man determines the order of magnitude of a selection differential that can be considered neutral, namely, $< 10^{-5}$ or even $< 10^{-6}$. The estimation of selection coefficients is in practice, however, very difficult. Selective differentials for advantageous mutations have only been estimated in a few cases mainly limited to malarial environments, such as for sickle cell anaemia heterozygotes, and for the G6PD gene. These two are both of the order 0.1 and even selection coefficients of this order of magnitude already require for their estimation the detailed examination of a considerable number of individuals. In most experimental situations, it is difficult or impossible to estimate selective coefficients smaller than 0.01. Only in very special situations has it proved possible to estimate small selection coefficients. Thus, the relative advantage of *ABO* alleles that protect against duodenal ulcer is of the order of 10^{-4} in males and 10^{-5} in females [6]. These estimates, however, depend on the assumption that differential mortality from ulcer is the sole cause of selection. Many such small selective differences could exist, usually unmeasurable, that could account for an observed rate of gene substitution which is higher than that expected for only neutral mutations.

Kimura ([14] and later) has suggested, following Haldane's earlier work on the cost of natural selection [12], that most mutations that eventually become substituted in a population must be neutral, because the genetic load implied by substitution at the rate indicated by observations on amino acid differences between species would be excessive. His computations are based, however, on the somewhat arbitrary assumption of independent action of different loci at the level of fitness. If a threshold model for selection is assumed, as has been suggested by Sved, Reed, and Bodmer [29], King [17], and Milkman [20] for heterotic polymorphisms, then the apparently excessive substitutional load disappears. It has actually been shown by Sved [27] that, assuming a threshold model, the observed rates of gene substitution can be readily accommodated with relatively minimal selective loads.

A number of other arguments, not based on the theoretical considerations we have discussed so far, have been put forward by King and Jukes [18] and Kimura ([14] and other papers) in favor of neutrality of most new mutations or "non-Darwinian evolution," as it has been called by King and Jukes. These

arguments concern, for example, the distribution of the number of amino acid substitutions per amino acid position in a protein, the question of "synonymous" substitutions, the apparent equivalence, according to some protein chemists, of different amino acid substitutions at many positions in many proteins and the apparent uniformity of the rate of evolution of some proteins over a wide evolutionary time span. Though we do not propose to elaborate further on these questions in this paper, we do not find any of these arguments particularly convincing as has been discussing by Richmond [25] and Clarke [7], [8].

The work of Lewontin and Hubby [19] in *Drosophila*, and Harris [13] in man has indicated average heterozygosity level per locus of 10 to 30 per cent corresponding to F values of from 0.7 to 0.9. If we take the minimal suggested value of N, namely 10^5, and use a minimal value of $\mu = 10^{-7}$ for the mutation rate to neutral alleles, then the formula used by Kimura and Crow [15] to evaluate F on the assumption of only neutral mutations, namely,

$$F = \frac{1}{1 + 4N\mu} \tag{3}$$

gives $F = 0.96$ which is almost certainly too high. If, on the other hand, we take $N = 10^6$ and $\mu = 10^{-6}$, this gives $F = 0.2$ which is clearly too low. Moreover, the high variance of F which was already mentioned makes the test insensitive. As already mentioned, it has been shown by Ewens (unpublished) that F is a poor statistic. Thus, observed levels of polymorphism could, in principle, be accounted for by neutral mutations, but the test is a weak one. This of course says nothing about the extent to which selection for fixable alleles is actually involved in maintaining observed levels of polymorphism. In Table III, we can notice that the introduction of selection does not alter the mean F values where F and μ are the same.

It seems worth recalling that in microorganisms, situations are available in which the rate of formation of advantageous mutants can be measured with some precision as illustrated by early work by Atwood, Ryan, and Schneider [1]. These authors noticed that asexual bacterial populations in which the equilibrium between a specific mutant and the rest of the population due to mutation selection balance was being investigated, occasionally underwent significant shifts in the relative frequency of the mutant in the population. These shifts could be interpreted on the hypothesis that new mutations with an increased fitness had occurred somewhere in the bacterial genome in one individual of the population, usually not of the original mutant type. These new advantageous mutations then wiped out the original mutant type whose equilibrium was being investigated. Once these new fitter types have replaced the old types, the specific mutant being investigated can reappear among the fitter types and return slowly to its former equilibrium. The estimate of the rate of mutation to such advantageous types under these conditions, was extremely low, namely, of the order of 10^{-12}, leaving plenty of scope for neutral or quasi neutral mutations. This system, however, only uncovers mutations with an increase in fitness that

is above a certain threshold and this may account for their extremely low rate of appearance. Though it is, of course, clear that results with bacteria and other microorganisms cannot readily be extrapolated to higher organisms, it does seem likely that further studies of the kinetics of such selective processes in microorganisms, both at a theoretical and at an experimental level, might well be rewarding.

6. Conclusions

The main point of constructing and describing our model has been to try and clarify the issues involved in matching population genetic theory to observed data on evolutionary rates and polymorphism. The results, perhaps unfortunately, are so far inconclusive, though we hope that further elaboration of the models and data will lead to a clearer understanding of the problems and, in particular, of the relative importance of neutral *versus* advantageous gene substitution. Although it appears that there is no major discrepancy between theory and data, the data do not yet clearly indicate what should be the prevailing values of N, μ, and the fitness differences to account for the observed properties of evolving populations. The major question of the extent to which new mutants are or are not associated with selective differences is, apparently, no nearer resolution today than it was well before the recent revival of discussion about "non-Darwinian" evolution.

REFERENCES

[1] K. C. Atwood, L. K. Schneider, and F. J. Ryan, "Selective mechanisms in bacteria," *Cold Spring Harbor Symp. Quant. Biol.*, Vol. 16 (1951), pp. 345–355.
[2] W. F. Bodmer and L. L. Cavalli-Sforza, "A migration matrix model for the study of random genetic drift," *Genetics*, Vol. 59 (1968), pp. 565–592.
[3] W. F. Bodmer and J. Felsenstein, "Linkage and selection: theoretical analysis of the deterministic two locus random mating model," *Genetics*, Vol. 57 (1967), pp. 237–265.
[4] W. F. Bodmer and P. A. Parsons, "Linkage and recombination in evolution," *Advan. Genet.*, Vol. 11 (1962), pp. 1–100.
[5] L. L. Cavalli-Sforza, "Population structure and human evolution," *Proc. Roy. Soc. London B Biol. Sci.*, Vol. 164 (1966), pp. 362–379.
[6] L. L. Cavalli-Sforza and W. F. Bodmer, *The Genetics of Human Populations*, San Francisco, Freeman, 1971.
[7] B. Clarke, "Darwinian evolution of proteins," *Science*, Vol. 168 (1970), p. 1009.
[8] ———, "Selective constraints on amino-acid substitutions during the evolution of proteins," *Nature*, Vol. 228 (1970), pp. 159–160.
[9] E. S. Deevey, "The human population," *Sci. Amer.*, Vol. 203 (1960), pp. 194–204.
[10] I. Franklin and R. C. Lewontin, "Is the gene the unit of selection?" *Genetics*, Vol. 65 (1970), pp. 707–734.
[11] E. Hadorn, *Letalfaktoren*, Stuttgart, Georg Thieme Verlag, 1955.
[12] J. B. S. Haldane, "The cost of natural selection," *J. Genet.*, Vol. 55 (1957), pp. 511–524.
[13] H. Harris, "Enzyme and protein polymorphism in human populations," *Brit. Med. Bull.*, Vol. 25 (1969), pp. 5–13.

[14] M. KIMURA, "Evolutionary rate at the molecular level," *Nature*, Vol. 217 (1968), pp. 624–626.

[15] M. KIMURA and J. F. CROW, "The number of alleles that can be maintained in a finite population," *Genetics*, Vol. 49 (1964), pp. 725–738.

[16] M. KIMURA and T. OHTA, "Protein polymorphism as a phase of molecular evolution," *Nature*, Vol. 229 (1971), pp. 467–489.

[17] J. L. KING, "Continuously distributed factors affecting fitness," *Genetics*, Vol. 55 (1967), pp. 483-492.

[18] J. L. KING and T. H. JUKES, "Non-Darwinian evolution," *Science*, Vol. 164 (1969), pp. 788–798.

[19] R. C. LEWONTIN and J. L. HUBBY, "A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*," *Genetics*, Vol. 54 (1966), pp. 595–609.

[20] R. D. MILKMAN, "Heterosis as a major cause of heterogeneity in nature," *Genetics*, Vol. 55 (1967), pp. 493–495.

[21] P. A. P. MORAN, *The Statistical Processes of Evolutionary Theory*, Oxford, The Clarendon Press, 1962.

[22] M. L. PETRAS, J. D. REIMER, F. G. BIDDLE, J. E. MARTIN, and R. S. LINTON, "Studies of natural populations of *Mus*. V. A survey of nine loci for polymorphisms," *Can. J. Genet. Cytol.*, Vol. 11 (1969), pp. 497–513.

[23] E. M. PRAGER and A. C. WILSON, "Multiple lysozymes of duck egg white," *J. Biol. Chem.*, Vol. 246 (1971), pp. 523–530.

[24] S. PRAKASH, R. C. LEWONTIN, and J. L. HUBBY, "A molecular approach to the study of genic variation in central, marginal and isolated populations of *Drosophila pseudoobscura*," *Genetics*, Vol. 61 (1969), pp. 841–848.

[25] R. C. RICHMOND, "Non-Darwinian evolution: A critique," *Nature*, Vol. 225 (1970), pp. 1025–1028.

[26] J. A. SVED, "Linkage disequilibrium and homozygosity of chromosome segments in finite populations," *Theor. Pop. Biol.*, Vol. 3 (1971), pp. 125–141.

[27] ———, "Possible rates of gene substitution in evolution," *American Naturalist*, Vol. 102 (1968) pp. 283–293.

[28] ———, "The stability of linked systems of loci with a small population size," *Genetics*, Vol. 59 (1968), pp. 543–563.

[29] J. A. SVED, T. E. REED, and W. F. BODMER, "The number of balanced polymorphisms which can be maintained in a natural population," *Genetics*, Vol. 55 (1967), pp. 469–481.

[30] E. ZUCKERKANDL and L. PAULING, "Evolutionary disease, evolution and genic heterogeneity," *Horizons in Biochemistry* (edited by M. KASHA and B. Pullman), Chicago, Academic Press, 1965.