MAINTENANCE OF GENETIC VARIABILITY UNDER THE JOINT EFFECT OF MUTATION, SELECTION AND RANDOM DRIFT*

WEN-HSIUNG LI

Center for Demographic and Population Genetics,
University of Texas Health Science Center at Houston,
Houston, Texas 77025

Manuscript received August 1, 1977
Revised copy received March 31, 1978

ABSTRACT

Formulae are developed for the distribution of allele frequencies (the frequency spectrum), the mean number of alleles in a sample, and the mean and variance of heterozygosity under mutation pressure and under either genic or recessive selection. Numerical computations are carried out by using these formulae and WATTERSON'S (1977) formula for the distribution of allele frequencies under overdominant selection. The following properties are observed: (1) The effect of selection on the distribution of allele frequencies is slight when $4Ns \leq 4$, but becomes strong when $4Ns$ becomes larger than 10, where N denotes the effective size and s the selective difference between alleles. Genic selection and recessive selection tend to force the distribution to be U-shaped, whereas overdominant selection has the opposite tendency. (2) The mean total number of alleles in a sample is much more strongly affected by selection than the mean number of rare alleles in a sample. (3) Even slight heterozygote advantage, as small as 10^{-5}, increases considerably the mean heterozygosity of a population, as compared to the case of neutral mutations. On the other hand, even slight genic or recessive selection causes a great reduction in heterozygosity when population size is large. (4) As a test statistic, the variance of heterozygosity can be used to detect the presence of selection, though it is not efficient when the selection intensity is very weak, say when $4Ns$ is around 4 or less. A model, which is somewhat similar to OHTA'S (1976) model of slightly deleterious mutations, has been proposed to explain the following general patterns of genic variation: (i) There seems to be an upper limit for the observed average heterozygosities. (ii) The distribution of allele frequencies is U-shaped for every species surveyed. (iii) Most of the species surveyed tend to have an excess of rare alleles as compared with that expected under the neutral mutation hypothesis.

THE selectionist *vs.* neutralist controversy over the maintenance of genic variation in natural populations has now continued unabated for ten years (see LEWONTIN 1974; NEI 1975 for reviews). What is needed to resolve this controversy is a theory by which the statistical properties of a population under the joint effect of mutation, selection and random drift can be examined thoroughly. Only by so doing can one evaluate the relative importance of these three

* Dedicated to SEWALL WRIGHT for his pioneering work on the stochastic distribution of multiple alleles.

factors in the maintenance of genetic variation or construct suitable statistics for testing hypotheses. Actually such a theory can be developed by using WRIGHT's (1949a) formula for the joint distribution of multiple alleles, but only very recently have workers begun to do so. Using this formula, WATTERSON (1977) has developed a statistical theory for the case of symmetrical overdominance among multiple alleles, and I (LI 1977) have developed a similar theory for the cases of genic selection and recessive selection; WATTERSON (1978) has later studied more general cases. In LI (1977), however, I have presented only a short summary of some of my findings. Here I present a detailed account. In Section I, I derive formulae for the distribution of the mean number of alleles at different frequencies or the frequency spectrum, and formulae for the mean number of alleles in a sample. Numerical computations are carried out by using these formulae and that of WATTERSON (1977) to see how selection changes the shape of the distribution of the mean number of alleles at different frequencies. Numerical computations are also carried out for the expected number of alleles whose sample frequency is equal to or less than q, $0 \leq q \leq 1$. In Section II, I develop formulae for the mean and variance of heterozygosity. These results, together with that of WATTERSON (1977), are applied to study how the average heterozygosity of a population changes with population size. The present result is also used to examine the effect of selection on the variance of heterozygosity. In Section III, I discuss the implications of the present findings for the maintenance of protein polymorphism.

DISTRIBUTION OF ALLELE FREQUENCIES

One of the most useful methods of describing a population is the distribution of allele frequencies or the frequency spectrum, which is conventionally denoted by $\Phi(x)$. Actually, $\Phi(x)$ is not a distribution in the probabilistic sense, but has the meaning that $\Phi(x)\mathrm{d}x$ represents the expected number of alleles whose frequency is between $x - \mathrm{d}x/2$ and $x + \mathrm{d}x/2$. Although a more precise term for $\Phi(x)$ is the distribution of the mean number of alleles at different frequencies, I follow the convention of calling $\Phi(x)$ the distribution of allele frequencies. The distribution $\Phi(x)$ for the case of neutral mutations has been studied by WRIGHT (1949b), KIMURA and CROW (1964), KARLIN and MCGREGOR (1967), and NEI and LI (1976), while that for the case of overdominant selection by WATTERSON (1977). Here, I study the cases of genic and recessive selection. Before this, I review WRIGHT's (1949a) formula for the joint distribution of multiple alleles, which is essential to this study.

Consider a random-mating population of effective size N. Let the number of possible allelic states at a locus be K, and let A_i denote the ith allele and x_i its frequency. Let the selective value of genotype A_iA_j be w_{ij}, which is assumed to be constant over time, and let $\overline{W}(x_1, \ldots, x_K)$ be the mean fitness of the population. Assume that in each generation A_j mutates to A_i with probability v_{ji}, $j \neq i$. In practice, it is likely that v_{ji} is a function of both A_j and A_i, but no solution seems to have been obtained under this general condition. However, for the special case where $v_{ji} = v_i$ for all $j \neq i$, WRIGHT (1949a) has obtained, using some heuristic arguments, the following formula for the equilibrium joint probability density for the first $L = K - 1$ gene frequencies:

$$\phi(x_1, \ldots, x_L) = C\bar{W}^{2N} \prod_{i=1}^{K} x_i^{\alpha_i - 1} \quad , \tag{1}$$

where $\alpha_i = 4Nv_i$ and $x_K = 1 - x_1 - \ldots - x_L$. A formal mathematical proof of this formula has later been provided by KIMURA (1956), WATTERSON (1977) and LI (1977), independently. Note that the mutation rate per gene per generation from A_i to all other alleles is $v_1 + \ldots + v_{i-1} + v_{i+1} + \ldots + v_K$ or $u - v_i$, where $u = v_1 + \ldots + v_i + \ldots + v_K$. If v_i is the same for all i, then $u - v_i = v$ and $v_i = v/L$ for all i. This symmetrical case has been known as the K-allele model (WRIGHT 1949b; KARLIN and MCGREGOR 1967; KIMURA 1968). The normalizing factor C in formula (1) is determined by the relation

$$\int \ldots \int_R \phi(x_1, \ldots, x_L) \mathrm{d}x_1 \ldots \mathrm{d}x_L = 1, \tag{2}$$

where the integration is over the region defined by

$$R\colon\ 0 \leq x_i \leq 1, \qquad x_1 + \ldots + x_L \leq 1.$$

This multiple integral may be evaluated by following WATTERSON'S (1977) method, if the w_{ij}'s are given explicitly.

The form of formula (1) is amazingly simple. Thus, to determine the joint probability density, we need only know the mean fitness of the population and the mutation rate (v_i) to A_i. This makes formula (1) also applicable to the case where A_i denotes a class of equally fit alleles instead of a single allele. To see how this works, we consider the following simple example. Suppose that the last $(K - j)$ allelic states are equally fit and let us regard them as a single class B_{j+1}. Let $B_i = A_i$, $i = 1, \ldots, j$. Let y_i be the frequency of B_i and u_i be the mutation rate to B_i, $i = 1, \ldots, j + 1$. This implies $y_i = x_i$ and $u_i = v_i$, $i = 1, \ldots, j$; $y_{j+1} = x_{j+1} + \ldots + x_K$ and $u_{j+1} = v_{j+1} + \ldots + v_K$. Obviously $\bar{W}(x_1, \ldots, x_K)$ can be written as $\bar{W}(y_1, \ldots, y_{j+1})$. Then, formula (1) says that the joint distribution of $y_1, \ldots, y_j$ is given by

$$\phi(y_1, \ldots, y_j) = C'\bar{W}^{2N} \prod_{i=1}^{j+1} y_i^{4Nu_i - 1} \quad .$$

This can be verified by following the simple proof procedure of formula (1) given by LI (1977) (see also KIMURA 1956; WATTERSON 1977). In particular, it is easy to see that when $j = 1$, this formula becomes identical with that for the diallelic case where the mutation rate from B_2 to B_1 is u_1 and that from B_1 to B_2 is u_2 (WRIGHT 1949a).

Genic Selection

We begin with the general case. We assume that the mutation rate, $v = u - v_i$, per gene per generation is the same for all alleles so that $\alpha_i = 4Nv/L = \alpha$ for all i.

We use the notations: $\theta = 4Nv = L\alpha$ and $\theta' = 4Nu = \theta + \alpha$. Let the selection coefficient for the ith allele be a_i so that the relative fitness of genotype A_iA_j is given by $w_{ij} = 1 + a_i + a_j$. Then $\overline{W} = 1 + 2a_1x_1 + \ldots + 2a_Kx_K$ and

$$\phi(x_1, \ldots, x_L) = C\overline{W}^{2N} \prod_{1}^{K} x_i^{\alpha-1} , \tag{3}$$

$$= C \sum_{n=0}^{2N} \binom{2N}{n} n!\, 2^n \sum_{(n)} \prod_{1}^{K} [a_i^{n_i} x_i^{n_i+\alpha-1} / n_i!], \tag{3'}$$

where $\binom{2N}{n}$ is the binomial coefficient and the summation $\sum\limits_{(n)}$ is over all vectors $(n) = (n_1, \ldots, n_K)$ of non-negative integers such that $n_1 + \ldots + n_K = n$. Using the Dirichlet integral formula (JOHNSON and KOTZ 1972) and the relation given by (2), we find that

$$C^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} n! 2^n \sum_{(n)} \prod_{1}^{K} [a_i^{n_i} \Gamma(n_i + \alpha)/n_i!]/\Gamma(n + \theta').$$

To derive $\Phi(x)$, we focus our attention on a particular allele, say A_j, and compute the probability, $\phi_j(x_j)\mathrm{d}x_j$, that the frequency of A_j in the population is in $(x_j - \mathrm{d}x_j/2,\ x_j + \mathrm{d}x_j/2)$. Note that $\phi_j(x_j)$ is the marginal probability density for x_j, and therefore

$$\phi_j(x_j) = \int \ldots \int_R \phi(x_1, \ldots, x_L)\mathrm{d}x_1 \ldots \mathrm{d}x_{j-1}\mathrm{d}x_{j+1} \ldots \mathrm{d}x_L, \tag{4}$$

where the integration is over the region

$$R'\text{: } 0 \leq x_i \leq 1 - x_j,\ \Sigma x_i \leq 1 - x_j, \qquad i \neq j.$$

Use of the transformation

$$z_i = x_i/(1 - x_j), \qquad i \neq j,$$

enables us to apply the Dirichlet integral formula to evaluate the multiple integral of (4) and we obtain

$$\phi_j(x_j) = C \sum_{n=0}^{2N} \binom{2N}{n} n! 2^n \sum_{(n)} [\Gamma(n - n_j + \theta)^{-1} \prod_{\substack{i=1 \\ i \neq j}}^{K} \frac{a_j^{n_i}}{n_i!} \Gamma(n_i + \alpha)$$
$$\times (a_j^{n_j}/n_j!) x_j^{n_j+\alpha-1}(1 - x_j)^{n-n_j+\theta-1}].$$

Since there are K allelic states,

$$\Phi(x) = \sum_{j=1}^{K} \phi_j(x) . \tag{5}$$

The subscript of x is now dropped because we are concerend with the mean number of alleles at a given gene frequency class, not any particular allele or alleles. Note that without losing generality, we can assume $a_K = 0$, and formula (5) is simplified to some extent.

Formula (5) is general but it is useful only when K or the magnitude of the $4Na_i$'s is small; otherwise it becomes computationally intractable because, for a large n or K, there are too many possible alternatives of the n_i's to be considered.

In order to make detailed computations feasible, we consider the following models.

Suppose that there are three classes of alleles, within each of which there are a number of equally fit alleles. Let the number of allelic states for the first, second, and third classes be I, M, and $Q = K - I - M$, and let the selection coefficient be s_1 for the first-class alleles, s_2 for the second-class alleles, and 0 for the third-class alleles; s_1 and s_2 can be positive or negative. $\Phi(x)$ can be obtained by putting these conditions into formula (5), but this creates the same computational difficulty as that of formula (5). The following approach overcomes this difficulty. Let

$$\begin{array}{lll} y_1 = x_1 + \ldots + x_I, & u_1 = Iv/L, & \theta_1 = 4Nu_1 = I\alpha, \\ y_2 = x_{I+1} + \ldots + x_{I+M}, & u_2 = Mv/L, & \theta_2 = 4Nu_2 = M\alpha, \\ y_3 = 1 - y_1 - y_2, & u_3 = Qv/L, & \theta_3 = 4Nu_3 = Q\alpha, \end{array}$$

where y_i is the sum of the frequencies of the ith class alleles and u_i is the sum of mutation rates to the ith class alleles. Note that formula (5) now becomes

$$\Phi(x) = I\phi_1(x) + M\phi_{I+1}(x) + Q\phi_K(x). \tag{6}$$

Note also that we may alternatively let the selection coefficients for the first-, second-, and third-class alleles be 0, t_2, and t_3, or r_1, 0, and r_3 so that $\overline{W}$ can be expressed in terms of the y_i's in three different ways:

$$W = 1 + 2s_1y_1 + 2s_2y_2 \tag{7a}$$

$$= (1 + 2s_1)(1 + 2t_2y_2 + 2t_3y_3) \tag{7b}$$

$$= (1 + 2s_2)(1 + 2r_1y_1 + 2r_3y_3) \tag{7c}$$

where $t_2 = (s_2 - s_1)/(1 + 2s_1)$, $t_3 = -s_1/(1 + 2s_1)$, $r_1 = (s_1 - s_2)/(1 + 2s_2)$, and $r_3 = -s_2/(1 + 2s_2)$. To evaluate $\phi_1(x_1)$, we consider the joint distribution of $x_1, \ldots, x_{I-1}, y_2, y_3$, and write $\overline{W}$ as (7b). As mentioned in the remark on formula (1), this distribution can be written as

$$\phi(x_1, \ldots, x_{I-1}, y_2, y_3) = C_1\Gamma(\alpha)^{-I}(1 + 2t_2y_2 + 2t_3y_3)^{2N} \times y_2^{\theta_2 - 1}y_3^{\theta_3 - 1} \prod_1^I x_i^{\alpha - 1}, \tag{8}$$

where C_1 is the normalizing constant. Using multinomial expansion and the Dirichlet integral formula, we find that

$$C_1^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n + \theta')} \sum_{(n)} \frac{t_2^{n_1}t_3^{n_2}}{n_1!n_2!} \Gamma(n_1 + \theta_2)\Gamma(n_2 + \theta_3).$$

From formula (8), we obtain

$$\phi_1(x) = C_1\Gamma(\alpha)^{-1}x^{\alpha - 1}(1 - x)^{\theta - 1} \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n + \theta)} (1 - x)^n \times \sum_{(n)} \frac{t_2^{n_1}t_3^{n_2}}{n_1!n_2!} \Gamma(n_1 + \theta_2)\Gamma(n_2 + \theta_3). \tag{9}$$

(To keep the notation simple, we have written here and shall write $\phi_i(x_i)$ as $\phi_i(x)$ when it creates no ambiguity.) On the other hand, to evaluate $\phi_{I+1}(x_{I+1})$, we consider the joint distribution of $y_1, x_{I+1}, \ldots, x_{I+M-1}, y_3$, and write $\overline{W}$ as (7c).

$$\phi(y_1, x_{I+1}, \ldots, x_{I+M-1}, y_3) = C_2 \Gamma(\alpha)^{-M}(1 + 2r_1y_1 + 2r_3y_3)^{2N} \times y_1^{\theta_1-1} y_3^{\theta_3-1} \prod_{I+1}^{I+M} x_i^{\alpha-1}, \tag{10}$$

$$\phi_{I+1}(x) = C_2\Gamma(\alpha)^{-1}x^{\alpha-1}(1-x)^{\theta-1} \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+\theta)} \times (1-x)^n \sum_{(n)} \frac{r_1^{n_1} r_3^{n_2}}{n_1!n_2!} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3). \tag{11}$$

$$C_2^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+\theta')} \sum_{(n)} \frac{r_1^{n_1} r_3^{n_2}}{n_1!n_2!} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3).$$

In the same manner, we obtain

$$\phi(y_1, y_2, x_{I+M+2}, \ldots, x_K) = C_3\Gamma(\alpha)^{-Q}(1 + 2s_1y_1 + 2s_2y_2)^{2N} \times y_1^{\theta_1-1} y_2^{\theta_2-1} \prod_{I+M+1}^{K} x_i^{\alpha-1}, \tag{12}$$

$$\phi_K(x) = C_3\Gamma(\alpha)^{-1}x^{\alpha-1}(1-x)^{\theta-1} \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+\theta)} (1-x)^n \times \sum_{(n)} \frac{s_1^{n_1} s_2^{n_2}}{n_1!n_2!} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2), \tag{13}$$

$$C_3^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+\theta')} \sum_{(n)} \frac{s_1^{n_1} s_2^{n_2}}{n_1!n_2!} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2).$$

Putting together (6), (9), (11), and (13), we have

$$\Phi(x) = \frac{x^{\alpha-1}(1-x)^{\theta-1}}{\Gamma(1+\alpha)} \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+\theta)} (1-x)^n \sum_{(n)} \frac{1}{n_1!n_2!} \times [\theta_1 C_1 t_2^{n_1} t_3^{n_2} \Gamma(n_1+\theta_2)\Gamma(n_2+\theta_3) + \theta_2 C_2 r_1^{n_1} r_3^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3) + \theta_3 C_3 s_1^{n_1} s_2^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2)]. \tag{14}$$

If $|4Ns_i| << N$, $i = 1{,}2$, formula (14) may be approximated by

$$\Phi(x) = \frac{x^{\alpha-1}(1-x)^{\theta-1}}{\Gamma(1+\alpha)} \sum_{n=0}^{2\mathrm{N}} \frac{(1-x)^n}{\Gamma(n+\theta)} \sum_{(n)} \frac{1}{n_1!n_2!} \times [\theta_1 C_1 T_2^{n_1} T_3^{n_2} \Gamma(n_1+\theta_2)\Gamma(n_2+\theta_3) + \theta_2 C_2 R_1^{n_1} R_3^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3) + \theta_3 C_3 S_1^{n_1} S_2^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2)], \tag{14'}$$

where $S_i = 4Ns_i$, $T_i = 4Nt_i$, $R_i = 4Nr_i$, and

$$C_1^{-1} = \sum_{n=0}^{2N} \Gamma(n+\theta')^{-1} \sum_{(n)} T_2^{n_1} T_3^{n_2} \Gamma(n_1+\theta_2)\Gamma(n_2+\theta_3)/(n_1!n_2!),$$

$$C_2^{-1} = \sum_{n=0}^{2N} \Gamma(n+\theta')^{-1} \sum_{(n)} R_1^{n_1} R_3^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3)/(n_1!n_2!),$$

$$C_3^{-1} = \sum_{n=0}^{2N} \Gamma(n+\theta')^{-1} \sum_{(n)} S_1^{n_1} S_2^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2)/(n_1!n_2!).$$

The same approximation applies to all of the following formulae, but will not be repeated (see LI 1977). However, all numerical computations of this study are carried out under this approximation.

The key difference between formulae (5) and (14) is that in the latter for every n only the possible alternatives of n_1 and n_2 are to be considered, *i.e.*, the summations $\sum_{(n)}$ are now univariate summations over $n_1 = 0, 1, 2, \ldots, n$, with $n_2 \equiv n - n_1$. Since this is true for any K, the limiting case of infinitely many alleles is just a special case. Indeed, to apply formula (14) to the model of infinite alleles (WRIGHT 1949b; KIMURA and CROW 1964), we simply put $\alpha = 0$ and $\theta' = \theta$. This remark applied to all the following results.

The mean number of alleles in a population is given by

$$\bar{n} = \int_{1/4N}^{1} \Phi(x)\,\mathrm{d}x,$$

assuming that the effective population size, N, is equal to the actual size. (We use $1/4N$ instead of the conventional value $1/2N$ as the lower limit of integration because it allows a continuity correction.) When K is finite,

$$\bar{n} = \int_0^1 \Phi(x)\,\mathrm{d}x - \int_0^{1/4N} \Phi(x)\,\mathrm{d}x$$
$$= K - \int_0^{1/4N} \Phi(x)\,\mathrm{d}x,$$

in which the last integral represents the mean number of alleles that are not present in the population.

We now study the mean number of alleles in a sample. We consider only the case of infinitely many alleles, since the result for this case is somewhat simpler in form and numerical computations are usually carried out under this condition. OHTA (1976) shows that the mean number of alleles whose sample frequency is less than or equal to q is given by

$$n_q = \sum_{i=1}^{[i_s]} \int_{1/2N}^{1} \binom{2m}{i} x^i (1-x)^{2m-i} \Phi(x)\,dx,$$

where $2m$ is the number of genes sampled, $i_s = 2mq$, and $[i_s]$ denotes the integral part of i_s. (If we define n_q as the mean number of alleles whose sample frequency is less than q, then $[i_s] = 2mq - 1$ if $2mq$ is an integer, and $[i_s] =$ the integral part of $2mq$ if otherwise.) We assume that $N >> m$ so that the lower limit of integration can be replaced by 0 with a good approximation. Using the relation

$$\int_0^1 \binom{2m}{i} x^i (1-x)^{2m-i} (1-x)^n x^{-1} (1-x)^{\theta-1}\,\mathrm{d}x = \frac{(2m-i+1)_i}{i(2m+n+\theta-\mathrm{i})_i},$$

where $(a)_i = a(a+1) \ \ldots \ (a+i-1)$, we find that

$$n_q = \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+\theta)} \sum_{i=1}^{[i_s]} \frac{(2m-i+1)_i}{i(2m+n+\theta-i)_i} \sum_{(n)} \frac{1}{n_1!n_2!}$$
$$\times [\theta_1 C_1 t_2{}^{n_1} t_3{}^{n_2} \Gamma(n_1+\theta_2)\Gamma(n_2+\theta_3) + \theta_2 C_2 r_1{}^{n_1} r_3{}^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3)$$
$$+ \theta_3 C_3 s_1{}^{n_1} s_2{}^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2)]. \qquad (15)$$

If $q = 1$, the above formula can be simplified by noting that$[i_s] = 2m$ and

$$\sum_{i=1}^{2m} \frac{(2m-i+1)_i}{i(2m+n+\theta-i)_i} = \sum_{i=0}^{2m-1} \frac{1}{n+\theta+i} \, . \qquad (16)$$

The same substitution applies to the case of recessive selection below (formula (24)).When $q = 1$, formula (15) represents the expected total number of alleles in a sample of $2m$ genes, and it reduces to formula (11) of EWENS (1972) if all mutations are neutral.

The formulae for the case of $k(k > 3)$ classes of alleles can be written down immediately by analogy, but computer computation soon becomes impracticable as the number of classes increases. On the other hand, the case of two classes of alleles is just a special case of three classes of alleles. Since this case is of particular interest, because of its simplicity, we consider it in some detail. Let the selection coefficient be s for the first-class alleles and 0 for the second-class alleles. This is equivalent to assuming that $s_1 = s$ and $s_2 = 0$. Since there are no third-class alleles, $\theta_3 = 0$. Putting these conditions into formulae (14) and (15), we obtain

$$\Phi(x) = \frac{x^{\alpha-1}(1-x)^{\theta-1}}{\Gamma(1+\alpha)} \sum_{n=0}^{2N} \binom{2N}{n} \frac{2^n(1-x)^n}{\Gamma(n+\theta)}$$
$$\times [\theta_1 C_1 t^n \Gamma(n+\theta_2) + \theta_2 C_2 s^n \Gamma(n+\theta_1)], \qquad (17)$$

$$n_q = \sum_{n=0}^{2N} \binom{2N}{n} \frac{2^n}{\Gamma(n+\theta)} [\theta_1 C_1 t^n \Gamma(n+\theta_2) + \theta_1 C_2 s^n \Gamma(n+\theta_1)]$$
$$\times \sum_{i=1}^{[i_s]} \frac{(2m-i+1)_i}{i(2m+n+\theta-i)_i} \, , \qquad (18)$$

$$C_1^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} (2t)^n \Gamma(n+\theta_2)/\Gamma(n+\theta'),$$

$$C_2^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} (2s)^n \Gamma(n+\theta_1)/\Gamma(n+\theta'),$$

where $t = -s/(1+2s)$. Recently, WATTERSON (1978) has also studied the case of two classes of alleles under genic selection and has obtained an approximate formula for $\Phi(x)$, assuming that $S = 4Ns$ is very small. Incidentally, formula (6) of *Li* (1977) contains a typographic error: θ_2 should have been $\theta_2 - 1$.

The above model of two classes becomes identical with that of WRIGHT (1966) when there is only one allelic state in the first class, *i.e.*, $I = 1$. WRIGHT called the first allele the type allele and gave an estimate for the mean number of deleterious alleles present in the population. He assumed that each gene mutates at the rate of v per generation and that the rate of backward mutation from all other alleles to any allele is v_1. This is equivalent to the assumption that the number of allelic states is $K = v/v_1 + 1$. He obtained the following approximate formula for the mean number of deleterious alleles in the population, assuming S large.

$$n_d \approx (K-1)\left[1 - \frac{\Gamma(S+\alpha)}{\Gamma(S)\Gamma(\alpha)}\left(\frac{1}{2N}\right)^{\alpha}\right], \tag{19}$$

$$\approx (K-1)\left[1 - \left(\frac{S-1}{2N}\right)^{\alpha}\frac{1}{\Gamma(1+\alpha)}\right]. \tag{19'}$$

The general formula for n_d is obtained by integrating the second term of (17) from $1/4N$ to 1 and by noting that $\theta_1 = \alpha$ and $\theta_2 = \theta$.

$$n_d = K - 1 - \frac{\theta}{\Gamma(1+\alpha)} C_2 \sum_{n=0}^{2N} \binom{2N}{n} (2s)^n \frac{\Gamma(n+\alpha)}{\Gamma(n+\theta)}$$
$$\times \int_0^{1/4N} x^{\alpha-1}(1-x)^{n+\theta-1}dx. \tag{20}$$

Formula (20) applies to any values of N and s, but it is more complicated than WRIGHT'S (1966) formula. It is therefore interesting to find the condition under which WRIGHT'S formula holds approximately. WRIGHT considered $v_1 = 0.25 \times 10^{-9}$ and 0.25×10^{-10}, and found that n_d is almost the same for both values of v_1 if $v = 10^{-6}$, $s \geq 10^{-4}$, and $N \geq 10^5$. Since the approximation is more accurate when v_1 is larger, I have used the smaller value $v_1 = 0.25 \times 10^{-10}$ in computing Table 1. It is seen from this table that WRIGHT'S formula holds approximately when S is larger than 10 and that formula (19') gives a close approximation to formula (19).

Recessive Selection

We consider two classes of alleles and assume that the number of allelic states is I for the first class and $M = K - I$ for the second class. Let the relative fitness of genotype A_iA_j be $1 - 2s$ if $i,j > I$, and be 1 otherwise. This means that the second-class alleles are completely recessive. The joint distribution of $x_1, \ldots, x_{I-1}$ and y_2 is

TABLE 1

Mean number of deleterious alleles in a population of size N

$4N$	4×10^4	8×10^4	10^5	2×10^5	3×10^5	4×10^5
$4Ns$	4	8	10	20	30	40
formula (20)	0.58	1.16	1.29	1.65	2.47	3.31
formula (19)	0.32	0.64	0.80	1.60	2.39	3.18
formula (19')	0.33	0.65	0.81	1.60	2.39	3.19

Note: $v = 10^{-6}$, $v_1 = 0.25 \times 10^{-10}$ and $s = 10^{-4}$.

$$\phi(x_1,\ldots,x_{I-1},y_2) = C_1\Gamma(\alpha)^{-I}(1-2sy_2^2)^{2N}y_2{}^{\theta_2-1}\prod_1^I x_i^{\alpha-1} , \tag{21}$$

$$C_1^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} (-2s)^n \frac{\Gamma(2n+\theta_2)}{\Gamma(2n+\theta')} ,$$

from which we obtain

$$\phi_1(x_1) = C_1\Gamma(\alpha)^{-1}\sum_{n=0}^{2N}\binom{2N}{n}(-2s)^n\frac{\Gamma(2n+\theta_2)}{\Gamma(2n+\theta)}x_1^{\alpha-1}(1-x_1)^{2n+\theta-1} .$$

Similarly we obtain

$$\phi(y_1,x_{I+2},\ldots,x_K) = C_2\Gamma(\alpha)^{-M}(1+2s_1y_1+2s_2y_1^2)^{2N}y_1^{\theta_1-1}\prod_{I+1}^K x_i^{\alpha-1} , \tag{22}$$

$$\phi_K(x_K) = C_2\Gamma(\alpha)^{-1}\sum_{n=0}^{2N}\binom{2N}{n}n!2^n\sum_{(n)}\frac{s_1{}^{n_1}s_2{}^{n_2}\Gamma(n_1+2n_2+\theta_1)}{n_1!n_2!\Gamma(n_1+2n_2+\theta)}$$
$$\times\, x_K^{\alpha-1}(1-x_K)^{n_1+2n_2+\theta-1} ,$$

$$C_2^{-1} = \sum_{n=0}^{2N}\binom{2N}{n}n!2^n\sum_{(n)}\frac{s_1{}^{n_1}s_2{}^{n_2}\Gamma(n_1+2n_2+\theta_1)}{n_1!n_2!\Gamma(n_1+2n_2+\theta')} ,$$

where $s_1 = 2s/(1-2s)$, $s_2 = -s/(1-2s)$. Thus,

$$\Phi(x) = I\Phi_1(x) + M\Phi_K(x)$$
$$= \frac{x^{\alpha-1}(1-x)^{\theta-1}}{\Gamma(1+\alpha)}\sum_{n=0}^{2N}\binom{2N}{n}[\theta_1C_1(-2s)^n\frac{\Gamma(2n+\theta_2)}{\Gamma(2n+\theta)}(1-x)^{2n}$$
$$+\,\theta_2C_2n!2^n\sum_{(n)}\frac{s_1{}^{n_1}s_2{}^{n_2}\Gamma(n_1+2n_2+\theta_1)}{n_1!n_2!\Gamma(n_1+2n_2+\theta)}(1-x)^{n_1+2n_2}] . \tag{23}$$

The formula corresponding to formula (15) is

$$n_q = \theta_1C_1\sum_{n=0}^{2N}\binom{2N}{n}(-2s)^n\frac{\Gamma(2n+\theta_2)}{\Gamma(2n+\theta)}\sum_{i=1}^{[i_s]}\frac{(2m-i+1)_i}{i(2m+2n+\theta-i)_i}$$
$$+\,\theta_2C_2\sum_{n=0}^{2N}\binom{2N}{n}n!2^n\sum_{(n)}\frac{s_1{}^{n_1}s_2{}^{n_2}\Gamma(n_1+2n_2+\theta_1)}{n_1!n_2!\Gamma(n_1+2n_2+\theta)}$$
$$\times\sum_{i=1}^{[i_s]}\frac{(2m-i+1)_i}{i(2m+n_1+2n_2+\theta-i)_i} . \tag{24}$$

As in the case of genic selection, WRIGHT (1966) gave an approximate formula for the mean number of recessive deleterious alleles in the population, assuming that there is a type allele. This formula is similar to formula (19), except that $4N\sqrt{2us}$ is to be substituted for S. The general formula for this number can be obtained by integrating the second term of formula (23), *i.e.*, $M\phi_K(x)$, from $1/4N$ to 1, and by noting that $\theta_1 = \alpha$ and $\theta_2 = \theta$. Numerical results indicate that Wright's formula again holds approximately if S is larger than 10. Note, however, that formula (19′) is applicable only if $4N\sqrt{2us}$ is larger than 1.

Numerical Examples

(A) *Distribution of allele frequencies*: Before considering numerical results, let us examine analytically the form of the distribution of allele frequencies under various types of selection. The distributions of allele frequencies for the cases of genic and recessive selection are given above, while that for the case of symmetrical overdominance is given by

$$\Phi(x) = \theta x^{-1}(1-x)^{\theta-1}e^{-\sigma x^2}W(\sigma(1-x)^2,\theta)/W(\sigma,\theta), \tag{25}$$

$$W(\sigma,\theta) = \sum_{n=0}^{\infty}\sum_{i=0}^{n}(-\sigma)^n C_{i,n}\theta^i/\Gamma(2n+\theta),$$

where $\sigma = 2Ns$ (WATTERSON 1977). This formula is derived under the assumptions that the relative fitness is 1 for all heterozygotes and $1-s$ for all homozygotes, that the number of allelic states is infinite, and that $|2Ns| << N$. The $C_{i,n}$ constants can be computed by use of the following recursive formula:

$$C_{i+1,n} = \sum_{j=1}^{n-i}\frac{C_{i,n-j}}{i+1}\,\frac{(2j-1)!}{j!}\,; \quad C_{0,0} = 1,$$

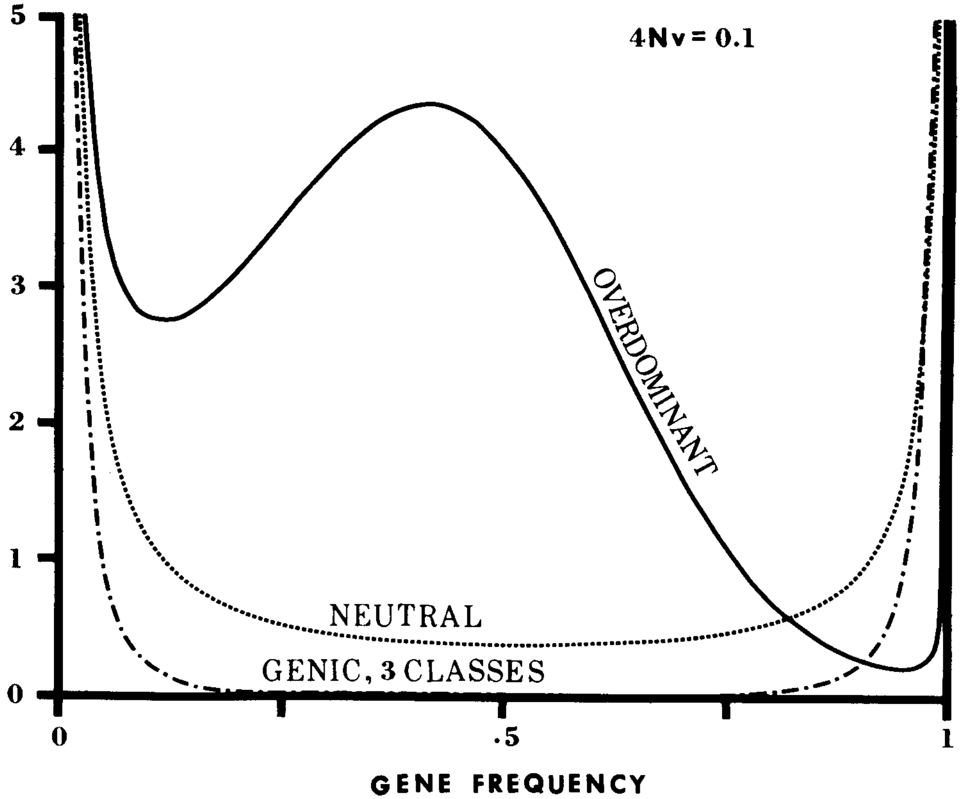
starting with $C_{1,n} = (2n-1)!/n!$ (STEWART 1976; WATTERSON 1977). We note from formulae (5), (14), (23), and (25) that regardless of the type of selection all the distributions under the model of infinite alleles have the common factor $x^{-1}(1-x)^{\theta-1}$, as does the distribution for the case of neutral mutations:

$$\Phi(x) = \theta x^{-1}(1-x)^{\theta-1}\ . \tag{26}$$

Thus, in all cases the value of $\Phi(x)$ at $x = 1$ is infinite if $\theta < 1$, finite if $\theta = 1$, and 0 if $\theta > 1$, while that at $x = 0$ is always infinite. It is also easy to see that $\Phi(x)$ given by (26) is U-shaped if $\theta < 1$ and L-shaped if $\theta \geq 1$. By continuity, the distribution under very weak selection pressure should be of the same shape as that for the case of neutral mutations; numerical results indicate that the shape of $\Phi(x)$ is not much affected if $s \leq 1/N$. These few properties are all that can be inferred analytically. To have a deeper understanding, numerical computations are necessary. A particular effort to be made is to see under what situation a U-shaped distribution can be obtained, because this shape of distribution is universally observed in nature for protein loci (unpublished studies of CHAKRABORTY, FUERST and NEI).

In all examples in this section, the model of infinite alleles is used. The examples given in Figures 1a and 1b are intended to show the effect of selection on the shape of the distribution of allele frequencies. The selection intensity is $4Ns_1 = 20$ and $4Ns_2 = 10$ for the case of three classes of alleles under genic selection, $4Ns = 20$ for the case of two classes of alleles under recessive selection and $2Ns = 10$ for the case of overdominant selection. In the case of genic selection, θ is divided into $\theta_1 = 0.01\theta$, $\theta_2 = 0.09\theta$, and $\theta_3 = 0.90\theta$; $\theta/2 = 2Nv$ represents the number of new alleles appearing in each generation, of which $\theta_i/2 = 2Nu_i$ belong to the ith class. In the case of recessive selection, $\theta_1 = 0.01\theta$ and

$\theta_2 = 0.99\theta$. In Figure 1a, $\theta = 0.1$ for all cases. The curve for neutral mutations is U-shaped, as expected. The curve for genic selection is also U-shaped but there are fewer intermediate- and low-frequency alleles and more high-frequency alleles as compared to the case of neutral mutations. To avoid crowding, the curve for recessive selection is not shown in Figure 1a, but it is U-shaped. The curve for overdominant selection has a peak at $x = 0.41$ and is far from being U-shaped—there are more intermediate- and low-frequency alleles and fewer high-frequency alleles as compared to the case of neutral mutations. Thus overdominant selection and genic selection have opposite effects on the shape of the distribution of allele frequencies. The mean heterozygosities for the cases of overdominant selection, neutral mutations, and genic selection are 0.485, 0.091, and 0.012, respectively. In Figure 1b, $\theta = 1$. Now only the curve for genic selection is U-shaped. However, the curve for recessive selection is nearly U-shaped and shows a similar tendency as that for genic selection. The curve for overdominant selection is again far from being U-shaped. As expected, the curve for neutral mutations is L-shaped. The mean heterozygosities for the cases of over-



FIGURES 1a and 1b.—Distribution of allele frequencies under various types of selection. The ordinate denotes $\Phi(x)$, which has the meaning that $\Phi(x)dx$ represents the expected number of alleles whose frequency is between $x - dx/2$ and $x + dx/2$. In Figure 1a, $\theta = 0.1$ while in dominant selection, neutral mutations, recessive selection, and genic selection are 0.697, 0.500, 0.238, and 0.114, respectively. The bearing of the above findings on protein polymorphism will be discussed later.

Figure 2 shows the effect of population size on the distribution of allele frequencies for the case of three classes of alleles under genic selection. The parameters used are $s_1 = 10^{-5}$, $s_2 = s_1/2$, $v = 1.12 \times 10^{-6}$, $u_1 = 0.02 \times 10^{-6}$, $u_2 = 0.1 \times 10^{-6}$, and $u_3 = 10^{-6}$. When $4N = 4 \times 10^5$, the curve is U-shaped. This is expected because $\theta = 0.448$ is smaller than one. The curve for $4N = 15 \times 10^5$ has a peak at $x = 0.95$, but there are now fewer intermediate-frequency alleles and more low-frequency alleles than those for the previous case. As expected, $\Phi(x)$ becomes 0 at $x = 1$ since θ is now 1.68. When $4N$ increases to 30×10^5, the peak becomes higher and moves to the left. This tendency will continue as population size increases—when N becomes infinite, all alleles will be concentrated near $x = 0$. Although the three curves for $4N = 4 \times 10^5$, 15×10^5, and 30×10^5 look different, they yield very similar mean heterozygosities: 0.268, 0.250, and 0.269.

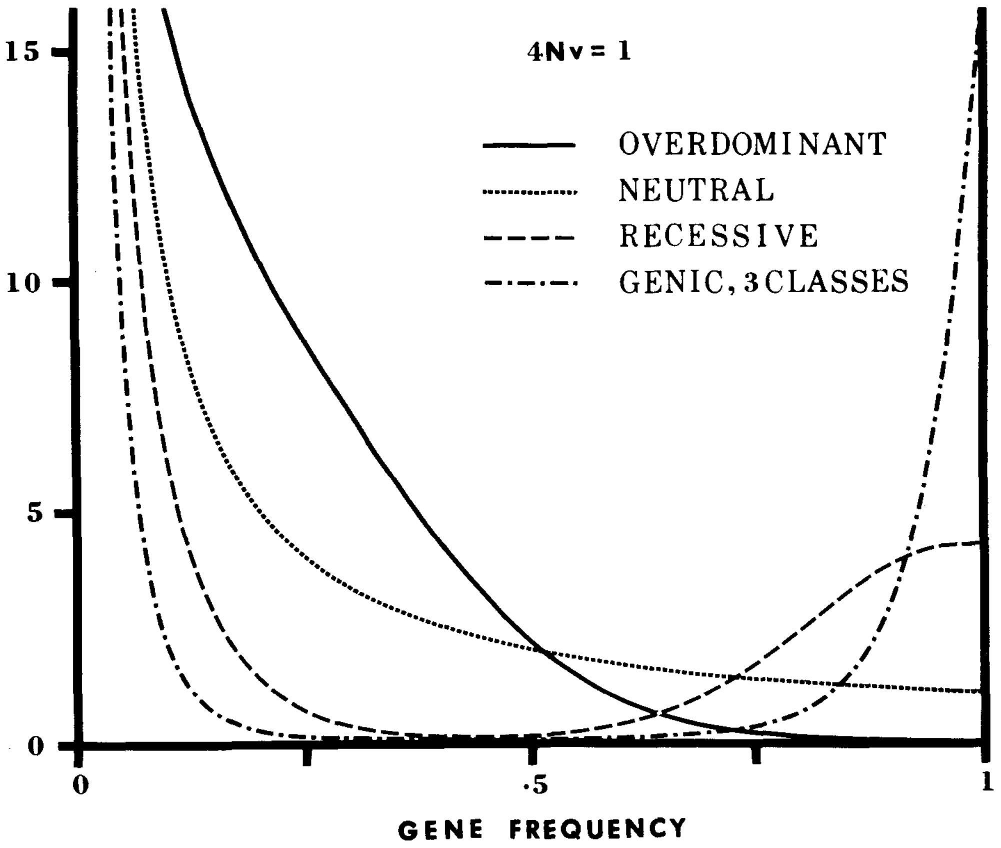


Figure 1b, $\theta = 1$. Overdominant selection: $\sigma = 2Ns = 10$. Recessive selection: $4Ns = 20$, $\theta_1 = 0.01\theta$, and $\theta_2 = 0.99\theta$. Genic selection: $4Ns_1 = 20$, $4Ns_2 = 10$, $\theta_1 = 0.01\theta$, $\theta_2 = 0.09\theta$, and $\theta_3 = 0.9\theta$.

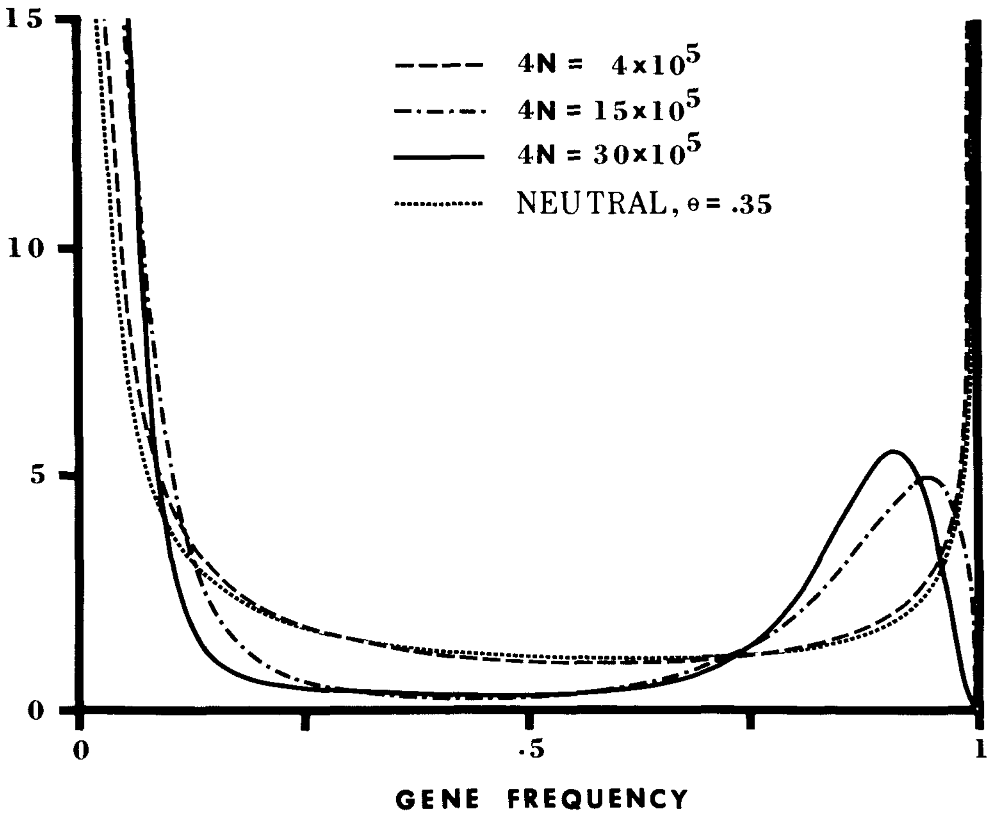


FIGURE 2.—The effect of population size on the distribution of allele frequencies under genic selection. The ordinate denotes $\Phi(x)$, which has the meaning that $\Phi(x)dx$ represents the expected number of alleles whose frequency is between $x - dx/2$ and $x + dx/2$. Mutations are divided into three classes with $u_1 = 0.02 \times 10^{-6}$, $u_2 = 0.1 \times 10^{-6}$, and $u_3 = 10^{-6}$, $s_1 = 10^{-5}$, and $s_2 = s_1/2$. The neutral case is for comparison. For details, see text.

Let us now consider a hypothetical population in which mutations are strictly neutral, but the mean heterozygosity $\bar{H}$ at equilibrium is about the same as those of the above three populations, say $\bar{H} = 0.26$. From this $\bar{H}$, we obtain $\theta = 0.35$ by using the relation $\bar{H} = \theta/(1 + \theta)$ (KIMURA and CROW 1964). Using this θ value and formula (26), we obtain the curve for the case of neutral mutations (Figure 2). This curve is very similar to the first curve, but different from the other two curves. Thus, the distribution of allele frequencies may be used to detect weak selection in large populations where the effect of random drift is relatively weak. One general property is that for a given level of mean heterozygosity there is an excess of rare alleles in the case of genic selection compared with the case of neutral mutations. On the other hand, in the case of overdominant selection the number of rare alleles tends to be small, while the number of intermediate-frequency alleles tends to be large compared with the case of neutral mutations. For example, the mean heterozygosity for the case of overdominant selection given in Figure 1a is 0.485, which is close to the value of 0.500 for the case of neutral mutations given in Figure 1b, but there are fewer rare alleles and more intermediate-frequency alleles in the former case than in the latter (note the difference in scale for the ordinates). It should be noted that in this method of comparison the distribution for the model of neutral mutations is computed by using $\theta = \bar{H}/(1 - \bar{H})$. Here, $\bar{H}$ is the expected heterozygosity for the population under study, while in practice it refers to the observed average heterozygosity. If the number of rare alleles for the population under study is larger (smaller) than that for the model of neutral mutations, then we say there is "an excess (deficiency) of rare alleles." This terminology is used in this sense throughout the present paper.

In Figure 3 we plot the distribution of the mean number of the third-class (most disadvantageous) alleles at different frequencies for the three cases of genic selection shown in Figure 2. It is seen that when $4N = 4 \times 10^5$ and $4Ns_1 = 4$, the third-class alleles are spread over the whole range of gene fre-

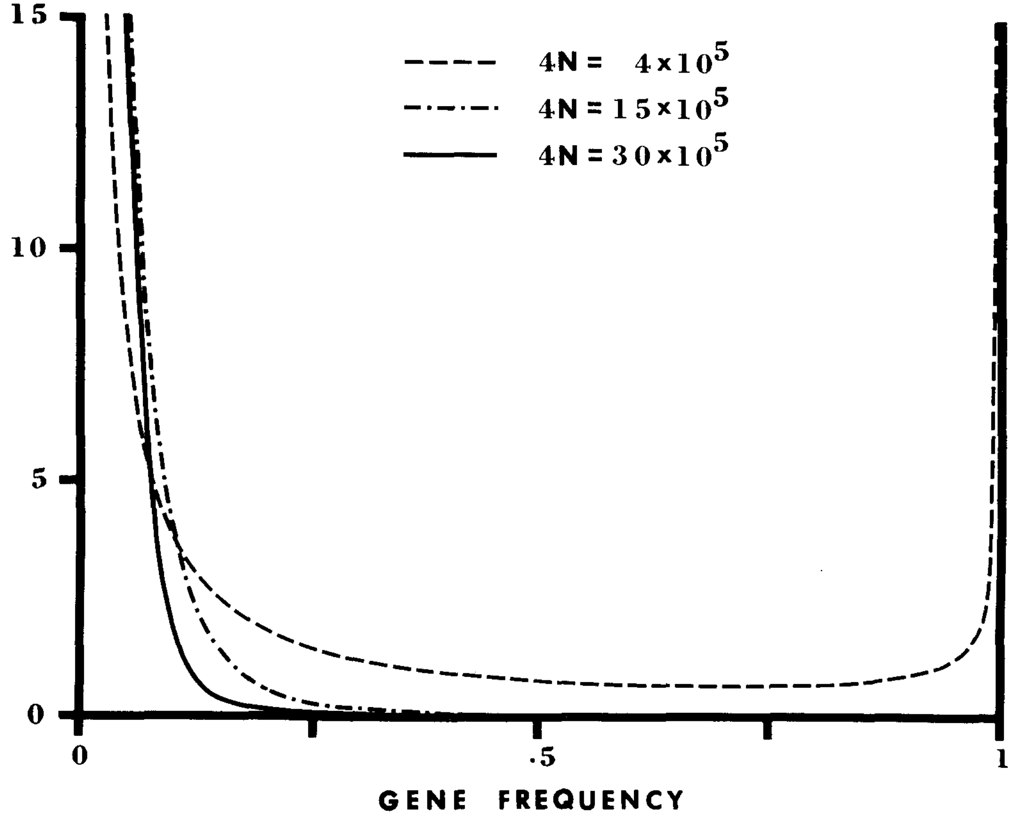


FIGURE 3.—Distributions of the number of the third-class (most disadvantageous) alleles under genic selection. The ordinate denotes $\Phi_3(x)$, which has the meaning that $\Phi_3(x)dx$ represents the expected number of the third-class alleles whose frequency is between $x - dx/2$ and $x + dx/2$. The parameters used are the same as those of Figure 2.

quency, including the fixation class. But when $4N$ increases to 15×10^5 so that $4Ns_1 = 15$, virtually none of the third-class alleles have a frequency higher than 0.3. As the population size increases more, the third-class alleles are pushed further down to lower frequencies. Thus, slightly deleterious mutations are able to spread over the whole population when N is of the order of $1/s$ or smaller, but are kept in low frequency when N is one order larger than $1/s$, where s is the selection coefficient against these mutations.

A property of prime importance emerges from the above results: the higher the potential for mutations to persist in the population, the lower the probability for $\Phi(x)$ to be U-shaped. (The potential is determined by N and the selective values of the mutations.) This is because the sum of allele frequencies must be one, and therefore the probability for any of the allele frequencies to be close to one becomes small when the number of alleles becomes large (*cf.*, Fig. 1b). Obviously the distribution $\Phi(x)$ cannot be U-shaped if there is no allele with frequency close to one. The distribution also cannot be U-shaped if there exists a peak at an intermediate frequency. Such a peak can arise if there is a force to keep some alleles at intermediate frequencies (*cf.*, Fig. 1a). It can also arise because of the accumulation of a large number of low frequency alleles (*cf.*, Fig. 2).

In the light of the above findings, the property of the distribution of allele frequencies under various situations can be recapitulated as follows: (1) For neutral mutations, there is no selective force to retain alleles in the population, but there is also no selective force to eliminate them so that alleles can become extinct only through random drift or mutation. Thus, the distribution is U-shaped if new alleles arise at a rate lower than one in every two generations, *i.e.*, $\theta = 4Nv < 1$, but it becomes L-shaped if new alleles arise at a higher rate, *i.e.*, $\theta \geq 1$. (2) Balancing selection not only has a high potential to retain alleles in the population, but also has a tendency to produce a peak at intermediate frequencies. Thus, a U-shaped distribution is unlikely to be observed under this mode of selection, unless the population size is very small so that random drift is strong and selection becomes ineffective. Numerical computations based on the model of symmetrical overdominance show that even if $2Ns$ is as small as 5 the distribution is non-U-shaped, because of the existence of a mild peak in the middle, if θ is 0.1; when θ becomes smaller this peak gradually becomes less conspicuous and the distribution tends to become U-shaped. One may argue that when the assumption of symmetry is removed the population will have a lower potential for holding alleles. However, it should be stressed that even severely deleterious alleles can accumulate in the population if they enjoy heterozygote advantage or minority advantage. The best example of this is the sickle-cell anemia gene in Africa; despite its lethality in homozygous condition, the frequencies of this gene in some African populations are as high as 0.15, sometimes even 0.20 (ALLISON 1961). Therefore, if balancing selection is prevalent, then even severely deleterious mutations may persist in the population for a long time and the distribution would soon become non-U-shaped as the population size increases. (3) "Purifying" (or "negative") selection, which includes genic selection, recessive selection, etc., tends to force the distribution to be U-shaped, for it is rather effective in eliminating disadvantageous mutations or in keepng them in low frequencies. Under this type of selection, the distribution is U-shaped if $\theta \leq 1$. When θ becomes larger than one, $\Phi(x)$ becomes 0 at $x = 1$, but has a peak at a high gene frequency (see Figure 2). The location of this peak depends on the intensity of selection as well as the magnitude of θ. If it is close to $x = 1$, the distribution, when plotted as a histogram, may become U-shaped. For example, the histogram for the curve with $\theta = 1.68$ in Figure 2 is U-shaped if the range of gene frequency is divided into ten equal intervals. This example shows that by incorporating even tiny selective differences such as $s_1 = 10^{-5}$ and $s_2 = 10^{-5}/2$ into the model of selective neutrality, a U-shaped histogram can be obtained even if θ is substantially larger than 1. In natural populations a U-shaped histogram may be obtained for an even larger θ value, because the majority of mutations are perhaps more deleterious than $s = 10^{-5}$. Note also that in practice only a finite number of genes are sampled from the population, so that low-frequency alleles are less likely to be observed than high-frequency ones. This sampling effect tends to move the aforementioned peak closer to $x = 1$. Numerical computations show that this effect increases the chance of observing a U-shaped histogram, though only to a small extent. Thus, under purifying selection the observed distribution, which is generally plotted as a histogram, can be U-shaped even if θ is considerably larger than one, say of the order of 10. However, it is unlikely that a U-shaped distribution can be observed if θ is much larger than one, say of the order of 100 or larger, unless an overwhelming majority of mutations are very deleterious so that they are quickly eliminated from the population or kept in exceedingly low frequencies.

(B) *Number of alleles in a sample*: Table 2 presents the mean number, n_q, of alleles with sample frequencies less than or equal to q, when m individuals or $2m$ genes are randomly chosen from the population. The parameters are specified in the table. For each θ value, we consider three cases: (1) neutral mutations, (2) two classes of alleles under genic selection, and (3) three classes of alleles under genic selection. The third-class alleles of case 3 correspond to the second-class alleles of case 2, while the first- and second-class alleles of case 3 represent a further subdivision of the first-class alleles of case 2. The population size for the three cases with $\theta = 3.36$ is 7.5 times that for the three cases with $\theta = 0.448$. When there are two or three classes of alleles, the first value in parentheses denotes the number of alleles from the first class, the second value from the second class, and so on. A number of interesting properties emerge from this table. (1) For a given θ value, the mean number of rare alleles, defined as alleles whose sample frequency is less than or equal to 0.01, is almost the same for the two cases of genic selection and also for the case of no selection, particularly when the sample size is large. On the other hand, the expected total numbers of alleles in a sample for the two cases of genic selection differ considerably from that for the case of no selection. This finding supports NEI's (1977) contention that, for

TABLE 2

Mean number of different alleles in a sample of m *individuals* (2m *genes*)

		θ=0.448			θ=3.36			θ=0.37
		Neutral mutations	Two classes θ_1=0.048, θ_2=0.4; S=4	Three classes θ_1=0.008, θ_2=0.04, θ_3=0.4; S_1=4, S_2=2	Neutral mutations	Two classes θ_1=0.36, θ_2=3 S=30	Three classes θ_1=0.06, θ_2=0.3, θ_3=3 S_1=30, S_2=15	Neutral mutations
m=20	$n_{0.01}$	0	0	0	0	0	0	0
	$n_{0.10}$	0.96	0.87(0.11, 0.76)	0.91(0.02, 0.09, 0.80)	6.27	3.30(0.78, 2.52)	3.05(0.13, 0.37, 2.54)	0.80
	$n_{1.00}$	2.66	2.34(0.98, 1.36)	2.50(0.31, 0.41, 1.78)	9.08	4.92(2.32, 2.60)	4.27(1.24, 0.41, 2.62)	2.40
m=50	$n_{0.01}$	0.67	0.66(0.07, 0.58)	0.67(0.01, 0.06, 0.59)	4.89	3.77(0.55, 3.22)	3.71(0.09, 0.38, 3.24)	0.56
	$n_{0.10}$	1.34	1.24(0.15, 1.09)	1.29(0.03, 0.12, 1.14)	9.09	5.49(1.08, 4.41)	5.22(0.18, 0.60, 4.44)	1.11
	$n_{1.00}$	3.07	2.74(1.03, 1.71)	2.90(0.32, 0.44, 2.13)	12.02	7.12(2.66, 4.47)	6.43(1.30, 0.63, 4.50)	2.74
m=100	$n_{0.01}$	0.67	0.66(0.07, 0.59)	0.67(0.01, 0.06, 0.60)	4.96	4.30(0.55, 3.76)	4.27(0.09, 0.41, 3.77)	0.56
	$n_{0.10}$	1.64	1.54(0.18, 1.35)	1.59(0.03, 0.15, 1.41)	11.32	7.45(1.32, 6.13)	7.17(0.22, 0.78, 6.16)	1.35
	$n_{1.00}$	3.38	3.05(1.06, 1.99)	3.21(0.33, 0.47, 2.41)	14.30	9.08(2.90, 6.18)	8.38(1.34, 0.82, 6.22)	2.99
m=250	$n_{0.01}$	1.03	1.01(0.11, 0.90)	1.02(0.02, 0.09, 0.91)	7.59	6.89(0.82, 6.07)	6.86(0.14, 0.65, 6.07)	0.85
	$n_{0.10}$	2.04	1.94(0.23, 1.72)	1.99(0.04, 0.18, 1.77)	14.35	10.30(1.65, 8.64)	10.00(0.28, 1.05, 8.68)	1.69
	$n_{1.00}$	3.79	3.45(1.10, 2.35)	3.62(0.33, 0.51, 2.77)	17.35	11.92(2.23, 8.69)	11.20(1.39, 1.09, 8.72)	3.33
m=500	$n_{0.01}$	1.31	1.30(0.14, 1.16)	1.31(0.02, 0.12, 1.17)	9.76	9.04(1.06, 7.99)	9.01(0.17, 0.84, 8.00)	1.09
	$n_{0.10}$	2.34	2.25(0.26, 1.99)	2.30(0.04, 0.21, 2.04)	16.66	12.53(1.89,10.63)	12.22(0.32, 1.25,10.67)	1.94
	$n_{1.00}$	4.10	3.76(1.14, 2.62)	3.93(0.34, 0.54, 3.05)	19.67	14.17(3.48,10.68)	13.44(1.44, 1.29,10.72)	3.59

NOTE: $\theta = 4Nv$, where v is the mutation rate per gene per generation. $n_{0.01}$ denotes the number of alleles with sample frequencies ≤ 0.01, $n_{0.10}$ the number of alleles with sample frequencies ≤ 0.1, and $n_{1.00}$ the total number of alleles in the sample. When there are two or three classes of mutations, the first value in parentheses denotes the number of alleles from the first class, the second value from the second class and so on.

estimating mutation rate, it is better to use the number of rare alleles rather than the total number of alleles, for the former is much less affected by selection than the latter. (2) A great majority of the alleles in a sample are in low frequencies, *i.e.*, around or less than 0.01, and are largely due to mutations of the second or third class. (3) The n_q value is larger for case 3 than for case 2 when $S_1 = S = 4$, but the situation is reversed when $S_1 = S = 30$. A simple explanation for this phenomenon is as follows. In case 3, only a minority of mutations, less than 2 percent, belong to the first class. Therefore, when $S_1 = 4$ there is a high probability that the first-class alleles are in very low frequencies or even absent from the population (see the first values in parentheses for case 3 with $\theta = 0.448$). Compared with case 3, there should be more first-class alleles in case 2 because about 10 percent of the mutations belong to this class. Consequently, selection is weaker in case 3 than in case 2, so that n_q is larger for case 3 than for case 2. On the other hand, when $S_1 = S = 30$, selection becomes effective, so that the sum of the frequencies of the first-class alleles is high even in case 3. Now selection is stronger in case 3 than in case 2 because more alleles are selected against, and thus n_q is smaller for the former than for the latter.

The sampling property of allele frequencies can be used to detect the presence of selection. To see this, we consider the following example. Cases 3 with $\theta = 0.448$ and $\theta = 3.36$ are equivalent to the cases of $4N = 4 \times 10^5$ and $4N = 3 \times 10^6$ in Figure 2, respectively. As noted earlier, the mean heterozygosities for these two cases are virtually equal: 0.268 and 0.269. If we assume that the heterozygosity of a population is completely due to neutral mutations and use $\bar{H} = 0.270$ to estimate θ, then we get $\theta = 0.37$. The values of n_q for neutral mutations with $\theta = 0.37$ are given in the last column of Table 2. It is clear that the differences in these values between the cases of genic selection and neutral mutations are negligibly small if $S_1 = 4$ and $S_2 = 2$, but very large if $S_1 = 30$ and $S_2 = 15$. In particular, $n_{0.01}$ for the case of genic selection with $S_1 = 30$ and $S_2 = 15$ can be more than eight or nine times that for neutral mutations. We shall discuss the implication of this finding for protein polymorphism later.

MEAN AND VARIANCE OF HETEROZYGOSITY

Let H and J denote the heterozygosity and homozygosity of a locus. Under random mating,

$$J = x_1^2 + \ldots + x_K^2 \ ,$$

so that

$$\bar{J} \equiv \mathrm{E}(J) = \sum_{i=1}^{K} \mathrm{E}x_i^2 \ ,$$

$$\mathrm{E}(J^2) = \sum_{i=1}^{K} \mathrm{E}x_i^4 + \sum_{i=1}^{K} \sum_{j \neq i}^{K} \mathrm{E}[x_i^2 x_j^2] \ .$$

The variance of J is given by $V(J) = E(J^2) - \bar{J}^2$. Since $H = 1 - J$, $\bar{H} = 1 - \bar{J}$ and $V(H) = V(J)$. We shall determine the mean and variance of H by studying the moments of J. When there is no selection, the following results agree with those for the case of neutral mutations obtained by KIMURA and CROW (1964), STEWART (1976), WATTERSON (1974) and LI and NEI (1975).

Genic Selection

In the general case, the joint probability density of gene frequencies is given by formula (3′) and

$$\mathrm{E}(J) = C \int \cdots \int_R \sum_{i=1}^{K} x_i^2 \phi(x_1, \ldots, x_L) \mathrm{d}x_1 \ldots \mathrm{d}x_L$$

$$= C \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+2+\theta')} \sum_{(n)} \left[\left\{ \sum_{i=1}^{K} (n_i + \alpha)_2 \right\} \right.$$

$$\left. \times \prod_{i=1}^{K} \frac{s_i^{n_i}}{n_i!} \Gamma(n_i + \alpha) \right] . \qquad (27)$$

$$\mathrm{E}(J^2) = C \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+4+\theta')} \sum_{(n)} \left[\left\{ \sum_{i=1}^{K} (n_i + \alpha)_4 \right. \right.$$

$$\left. \left. + \sum_{i=1}^{K} \sum_{j \neq i}^{K} (n_i + \alpha)_2 (n_j + \alpha)_2 \right\} \prod_{i=1}^{K} \frac{s_i^{n_i}}{n_i!} \Gamma(n_i + \alpha) \right] . \qquad (28)$$

In the case of three classes of alleles, it is simpler to derive $\bar{J}$ from the distribution of allele frequencies given by formula (14).

$$\mathrm{E}(J) = \int_0^1 x^2 \Phi(x) \mathrm{d}x$$

$$= (1+\alpha) \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+2+\theta')} \sum_{(n)} \frac{1}{n_1!n_2!}$$

$$\times [\theta_1 C_1 t_2^{n_1} t_3^{n_2} \Gamma(n_1+\theta_2)\Gamma(n_2+\theta_3) + \theta_2 C_2 r_1^{n_1} r_3^{n_2}$$

$$\times \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3) + \theta_3 C_3 s_1^{n_1} s_2^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2)], \qquad (29)$$

in which the C_i's are the same as those in formula (14). To compute $\mathrm{E}(J^2)$, we notice that

$$\mathrm{E}(J^2) = I\mathrm{E}(x_1^4) + M\mathrm{E}(x_{I+1}^4) + Q\mathrm{E}(x_K^4)$$

$$+ I(I-1)\mathrm{E}(x_1^2 x_2^2) + M(M-1)\mathrm{E}(x_{I+1}^2 x_{I+2}^2) + Q(Q-1)\mathrm{E}(x_L^2 x_K^2)$$

$$+ 2IM\,\mathrm{E}(x_1^2 x_{I+1}^2) + 2IQ\mathrm{E}(x_1^2 x_K^2) + 2MQ\mathrm{E}(x_{I+1}^2 x_K^2).$$

From the joint distribution of $x_1, \ldots, x_{I-1}, y_2$, and y_3 given by formula (8) we obtain

$$\mathrm{E}(x_1^4) = C_1(\alpha)_4 \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+4+\theta')} \sum_{(n)} \frac{t_2^{n_1} t_3^{n_2}}{n_1!n_2!} \Gamma(n_1+\theta_2)\Gamma(n_2+\theta_3),$$

$$\mathrm{E}(x_1^2 x_2^2) = \alpha^2(\alpha+1)^2 \mathrm{E}(x_1^4)/(\alpha)_4.$$

From formulae (10) and (12), we obtain

$$\mathrm{E}(x_{I+1}^{4}) = C_2(\alpha)_4 \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+4+\theta')} \sum_{(n)} \frac{r_1^{n_1} r_3^{n_2}}{n_1!n_2!} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3),$$

$$\mathrm{E}(x_{I+1}^{2} x_{I+2}^{2}) = \alpha^2(\alpha+1)^2 \mathrm{E}(x_{I+1}^{4})/(\alpha)_4,$$

$$\mathrm{E}(x_K^4) = C_3(\alpha)_4 \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+4+\theta')} \sum_{(n)} \frac{s_1^{n_1} s_2^{n_2}}{n_1!n_2!} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2),$$

$$\mathrm{E}(x_L^2 x_K^2) = \alpha^2(\alpha+1)^2 \mathrm{E}(x_K^4)/(\alpha)_4.$$

To evaluate $\mathrm{E}(x_1^2 x_{I+1}^{2})$, we consider the joint distribution of x_1, z_1, $x_{I+1}, \ldots,$ x_{I+M-1}, and y_3, where $z_1 = y_1 - x_1$, and find that

$$\phi(x_1,z_1,x_{I+1},\ldots,x_{I+M-1},y_3) = C_4\Gamma(\alpha)^{-M}(1+2r_1x_1+2r_1z_1+2r_3y_3)^{2N}$$
$$\times x_1^{\alpha-1} z_1^{\theta_1-\alpha-1} y_3^{\theta_3-1} \prod_{I+1}^{I+M} x_i^{\alpha-1} ;$$

$$\mathrm{E}(x_1^2 x_{I+1}^{2}) = C_4\alpha(\alpha+1) \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+4+\theta')} \sum_{(n)} \frac{r_1^{n_1} r_3^{n_2} r_1^{n_3}}{n_1!n_2!n_3!}$$
$$\times \Gamma(n_1+\theta_1-\alpha)\Gamma(n_2+\theta_3)\Gamma(n_3+2+\alpha),$$

$$C_4^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+\theta')} \sum_{(n)} \frac{r_1^{n_1} r_3^{n_2} r_1^{n_3}}{n_1!n_2!n_3!} \Gamma(n_1+\theta_1-\alpha)$$
$$\times \Gamma(n_2+\theta_3)\Gamma(n_3+\alpha).$$

In the same manner, we obtain

$$\phi(x_1,z_1,y_2,x_{I+M+2},\ldots,x_K) = C_5\Gamma(\alpha)^{-Q}(1+2s_1x_1+2s_1z_1+2s_2y_2)^{2N}$$
$$\times x_1^{\alpha-1} z_1^{\theta_1-\alpha-1} y_2^{\theta_2-1} \prod_{I+M+1}^{K} x_i^{\alpha-1} ,$$

$$\mathrm{E}(x_1^2 x_K^2) = C_5\alpha(\alpha+1) \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+4+\theta')} \sum_{(n)} \frac{s_1^{n_1} s_2^{n_2} s_1^{n_3}}{n_1!n_2!n_3}$$
$$\times \Gamma(n_1+\theta_1-\alpha)\Gamma(n_2+\theta_2)\Gamma(n_3+2+\alpha),$$

$$C_5^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+\theta')} \sum_{(n)} \frac{s_1^{n_1} s_2^{n_2} s_1^{n_3}}{n_1!n_2!n_3!} \Gamma(n_1+\theta_1-\alpha)\Gamma(n_2+\theta_2)\Gamma(n_3+\alpha),$$

and

$$\phi(y_1,x_{I+1},z_2,x_{I+M+2},\ldots,x_K) = C_6\Gamma(\alpha)^{-Q}(1+2s_1y_1+2s_2x_{I+1}+2s_2z_2)^{2N}$$
$$\times y_1^{\theta_1-1} x_{I+1}^{\alpha-1} z_2^{\theta_2-\alpha-1} \prod_{I+M+1}^{K} x_i^{\alpha-1} ,$$

$$\mathrm{E}(x_{I+1}^{2} x_K^2) = C_6\alpha(\alpha+1) \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+4+\theta')} \sum_{(n)} \frac{s_1^{n_1} s_2^{n_2} s_2^{n_3}}{n_1!n_2!n_3!}$$

$$\times \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2-\alpha)\Gamma(n_3+2+\alpha),$$

$$C_6^{-1} = \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+\theta')} \sum_{(n)} \frac{s_1^{n_1}s_2^{n_2}s_2^{n_3}}{n_1!n_2!n_3!} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2-\alpha)$$

$$\times \Gamma(n_3+\alpha),$$

where $z_2 = y_2 - x_{I+1}$. By using these results, $\mathrm{E}(J^2)$ can be obtained. In the case of infinite alleles, it becomes

$$\mathrm{E}(J^2) = \sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+4+\theta)} \sum_{(n)} \frac{1}{n_1!n_2!} [\theta_1(6+\theta_1)C_1 t_2^{n_1} t_3^{n_2}$$

$$\times \Gamma(n_1+\theta_2)\Gamma(n_2+\theta_3) + \theta_2(6+\theta_2)C_2 r_1^{n_1} r_3^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3)$$

$$+ \theta_3(6+\theta_3)C_3 s_1^{n_1} s_2^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2)] + 2\sum_{n=0}^{2N} \binom{2N}{n} \frac{n!2^n}{\Gamma(n+4+\theta)}$$

$$\times \sum_{(n)} \frac{(n_3+1)}{n_1!n_2!} [\theta_1\theta_2 C_2 r_1^{n_1+n_3} r_3^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_3)$$

$$+ \theta_1\theta_3 C_3 s_1^{n_1+n_3} s_2^{n_2} \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2) + \theta_2\theta_3 C_3 s_1^{n_1} s_2^{n_2+n_3}$$

$$\times \Gamma(n_1+\theta_1)\Gamma(n_2+\theta_2)], \tag{30}$$

because $I\ C_4 \to \theta_1 C_2$, $I\ C_5 \to \theta_1 C_3$, and $M\ C_6 \to \theta_2 C_3$ as $K \to \infty$.

The corresponding formulae for the case of two classes of alleles are

$$\mathrm{E}(J) = (1+\alpha) \sum_{n=0}^{2N} \binom{2N}{n} 2^n [\theta_1 C_1 t^n \Gamma(n+\theta_2)$$

$$+ \theta_2 C_2 s^n \Gamma(n+\theta_1)]/\Gamma(n+2+\theta')\ , \tag{31}$$

$$\mathrm{E}(J^2) = \sum_{n=0}^{2N} \binom{2N}{n} \frac{2^n}{\Gamma(n+4+\theta)} [\theta_1(6+\theta_1)C_1 t^n \Gamma(n+\theta_2)$$

$$+ \theta_2(6+\theta_2)C_2 s^n \Gamma(n+\theta_1) + 2\theta_1\theta_2 C_2 s^n \sum_{(n)} \frac{n_1+1}{n_2!} \Gamma(n_2+\theta_1)]. \tag{32}$$

where C_1 and C_2 are given in formula (16). Note that formula (32) is written under the condition of $K = \infty$, but formula (31) holds for any K.

Recessive selection

We consider two classes of alleles and use the same notations as those of formula (23). The formulae corresponding to (31) and (32) are

$$\mathrm{E}(J) = (1+\alpha) \sum_{n=0}^{2N} \binom{2N}{n} [\theta_1 C_1 (-2s)^n \frac{\Gamma(2n+\theta_2)}{\Gamma(2n+2+\theta')}$$

$$+ \theta_2 C_2 n! 2^n \sum_{(n)} \frac{s_1^{n_1} s_2^{n_2} \Gamma(n_1+2n_2+\theta_1)}{n_1!n_2!\Gamma(n_1+2n_2+2+\theta')} \Bigg] , \tag{33}$$

$$\mathrm{E}(J^2) = \sum_{n=0}^{2N} \binom{2N}{n} \Big[\theta_1(6+\theta_1)C_1(-2s)^n \frac{\Gamma(2n+\theta_2)}{\Gamma(2n+4+\theta')}$$

$$+ \theta_2(6+\theta_2)C_2 n!2^n \sum_{(n)} \frac{{s_1}^{n_1}{s_2}^{n_2}\Gamma(n_1+2n_2+\theta_1)}{n_1!n_2!\Gamma(n_1+2n_2+4+\theta')}\Big] \tag{34}$$

$$+ 2\theta_1\theta_2 C_1 \sum_{n=0}^{2N} \binom{2N}{n} \frac{(-2s)^n(2n)!}{\Gamma(2n+4+\theta')} \sum_{(2n)} (n_1+1)\frac{\Gamma(n_2+\theta_2)}{n_2!}, \tag{35}$$

in which the last summation $\sum_{(2n)}$ is over all vectors $(2n) = (n_1,n_2)$ of nonnegative integers such that $n_1 + n_2 = 2n$.

Numerical examples

(A) *Mean heterozygosity*: Table 3 shows the mean heterozygosity of a population under various types of selection. The parameters are specified in the table and the footnotes of the table. In all cases, the model of infinite alleles is used. The mean heterozygosity for the case of overdominant selection is computed by numerical integration of formula (25), while those for the other cases are computed by using the above formulae. The case of neutral mutations is given for comparison. A number of interesting properties are observed. (1) Overdominant selection increases considerably the amount of mean heterozygosity, even if the heterozygote advantage is as tiny as 10^{-5}. (2) In large populations, the mean heterozygosities for the cases of genic and recessive selection are much less than those for the case of neutral mutations. Thus, in large populations even slight purifying selection causes a great reduction in heterozygosity. This is particularly so in the cases of genic selection and confirms OHTA and KIMURA's (1975) result by simulation. (3) The $\bar{H}$ value for the case of neutral mutations is somewhat

TABLE 3

Mean heterozygosity under various types of selection

$4N$ $\theta = 4Nv$	4×10^4 0.04	10^5 0.1	2×10^5 0.2	4×10^5 0.4	10^6 1	2×10^6 2	3×10^6 3	4×10^6 4
Neutral mutations	0.0385	0.0909	0.167	0.286	0.500	0.667	0.750	0.800
Overdominant selection*	0.0405	0.1038	0.208	0.384	0.636	0.765	0.820	0.850
Recessive selection†:								
Case 1	0.0389	0.0928	0.168	0.257	0.354	0.435	0.489	0.541
Case 2	0.0385	0.0911	0.167	0.273	0.322	0.360	0.380	0.398
Genic selection‡:								
Case 1 (2 classes)	0.0383	0.0893	0.153	0.208	0.245	0.308	0.361	0.407
Case 2 (2 classes)	0.0385	0.0907	0.165	0.259	0.207	0.204	0.212	0.219
Case 3 (3 classes)	0.0384	0.0904	0.163	0.253	0.237	0.220	0.227	0.234

* The heterozygote and homozygote fitnesses are 1 and $1 - s$, where $s = 10^{-5}$.

† Case 1: $\theta_1 = 0.1\theta$, $\theta_2 = 0.9\theta$, $s = 10^{-5}$; Case 2: $\theta_1 = 0.01\theta$, $\theta_2 = 0.99\theta$, $s = 10^{-5}$.

‡ Case 1: $\theta_1 = 0.1\theta$, $\theta_2 = 0.9\theta$, $s = 10^{-5}$; Case 2: $\theta_1 = 0.01\theta$, $\theta_2 = 0.99\theta$, $s = 10^{-5}$; Case 3: $\theta_1 = 0.01\theta$, $\theta_2 = 0.09\theta$, $\theta_3 = 0.9\theta$, $s_1 = 10^{-5}$, $s_2 = s_1/2$.

smaller than those for the cases of recessive selection if $4N$ is about 2×10^5 or less. This seems peculiar but may be explained as follows. If mutations are neutral and N is small or intermediate, there is a high probability that the population is monomorphic at the time of observation. In the case of recessive selection, on the other hand, there exists some sort of mutation-selection balance because unfavorable mutations are only slightly selected against, and they occur more often than favorable mutations. This balance reduces slightly the probability of being monomorphic (see LI 1977) and, consequently, increases the mean heterozygosity to a small extent. As an example, when $4N = 10^5$, $\Phi(0.99)$ is 6.373 for the case of neutral mutations, but 6.339 for the case of recessive selection with $\theta_1 = 0.1\theta$ (case 1). Note that this balance is strongly affected by random drift, so that there is a high probability that the first-class alleles will become very rare or even absent from the population, if the proportion of favorable mutations is very small, say 1% or less. This explains why the $\bar{H}$ value for the second case of recessive selection is very close to that for the case of neutral mutations, if $4N$ is about 2×10^5 or less. This also explains why $\bar{H}$ is larger for the second case of recessive selection than for the first case of recessive selection when $4N$ is around 4×10^5. However, as the population size increases, the $\bar{H}$ value for the second case becomes smaller than that for the first case, because selection becomes effective and the proportion of unfavorable mutations is larger in the second case than in the first case. (4) In the cases of recessive selection and the first case of genic selection, $\bar{H}$ increases with increasing N, but in the second and third cases of genic selection $\bar{H}$ first increases, then decreases and then increases again as N increases.

The following is a simple explanation for this phenomenon. When N is small, selection is not effective and all alleles behave almost as neutral alleles, so that even unfavorable alleles contribute significantly to heterozygosity. But as N increases, unfavorable alleles are selected against and their contribution to heterozygosity is diminished while favorable alleles increase their contribution. Whether or not $\bar{H}$ increases with increasing N depends on whether or not the increase due to favorable alleles can compensate for the decrease due to selection against unfavorable alleles. In the case of recessive selection, the compensation seems always more than enough; a number of other parameter values were tried, but no contrary result was obtained. In the case of genic selection, the compensation is not enough if the proportion of favorable mutations is much smaller than that of unfavorable mutations and N is small or intermediate, but it will eventually become more than enough as N becomes large. (5) $\bar{H}$ is smaller for the first case than for the second case of genic selection when N is small, but the situation is reversed when N is large. That is, when the proportion of favorable mutations is reduced from 10% to 1%, $\bar{H}$ increases if N is small but decreases if N is large. In the third case of genic selection, the first and second classes represent a further subdivision of the first class of case 1. It is interesting to note that the $\bar{H}$ value for case 3 always lies between those for cases 1 and 2. These observations can again be explained in terms of the interaction between selection and random drift.

TABLE 4

Mean heterozygosity under genic selection when there are I *optimal states*

$4N$		4×10^4	2×10^5	4×10^5	10^6	2×10^6	3×10^6	∞
Two classes*	$I=1$	0.0388	0.166	0.260	0.202	0.190	0.190	0.190
	$I=2$	0.0392	0.165	0.246	0.204	0.205	0.212	0.595
	$I=10$	0.0420	0.167	0.225	0.254	0.309	0.356	0.919
Three classes†	$I=1$	0.0388	0.164	0.255	0.236	0.208	0.208	0.208
	$I=2$	0.0391	0.164	0.246	0.230	0.223	0.230	0.604
	$I=10$	0.0420	0.167	0.232	0.270	0.323	0.369	0.921

* Two classes: $s=10^{-5}$, the mutation rate to slightly deleterious alleles is 10^{-6}, and the mutation rate to an optimal state is 10^{-8}.

† Three classes: $s_1=10^{-5}$, $s_2=0.5\times10^{-5}$, the mutation rate to the third class of alleles is 0.9×10^{-6}, the mutation rate to the second class is 0.1×10^{-6}, and the mutation rate to an optimal state is 10^{-8}.

Table 4 shows the mean heterozygosity for the case where the number, I, of the first-class allelic states is small rather than infinite. The first-class alleles are called the optimal alleles. In all cases the mutation rate from all other alleles to an optimal allele is 10^{-8}. In the case of two classes of alleles, the number of allelic states of the second class is infinite with $u_2=10^{-6}$. In the case of three classes of alleles, the numbers of the second- and third-class allelic states are infinite with $u_2=10^{-7}$ and $u_3=9\times10^{-7}$. The mean heterozygosity for $N=\infty$ is computed by using a deterministic model. Namely, in the case of two classes of alleles, the sum of the equilibrium frequencies of the second-class alleles is $q=u_2/s=10^{-6}/10^{-5}=0.1$, and the alleles of this class do not contribute to homozygosity. The frequency of an optimal allele is $(1-q)/I$, so that the homozygosity of the population is $\bar{J}=I[(1-q)/I]^2=(1-q)^2/I$ and the heterozygosity is $\bar{H}=1-\bar{J}$. The heterozygosity for the case of three classes of alleles for $N=\infty$ is computed in a similar manner. It is seen that, if $I=1$, the mean heterozygosity reaches the deterministic value when $4Ns$ is 20 or larger. This means that the effect of random genetic drift is negligible when $4Ns$ is of this order of magnitude. On the other hand, if $I\geq2$, the mean heterozygosity is still far from the deterministic value, even if $4Ns=30$. This is because the optimal alleles are neutral with respect to each other and their frequencies are much affected by random genetic drift. Note that in the two cases of $I=10$, $\bar{H}$ always increases with increasing N, while in the other four cases $\bar{H}$ first increases, then decreases and then increases again. The explanation is similar to that given above.

(B) *Mean heterozygosity for many classes of mutations*: So far we have considered discrete classes of mutations, but in reality the fitness spectrum of mutations seems to be continuous (Crow 1972; King 1972; Bodmer and Cavalli-Sforza 1972). Ohta (1978) has recently studied the mean number of heterozygous nucleotide sites per individual and the fixation probability of new mutations for a case of continuous spectrum. Here I am concerned with the mean heterozygosity. Theoretically, regardless of the shape of the spectrum, the mean heterozygosity for the discrete case should approach that for the continuous case as the number of classes increases. In practice, however, we can compute only a limited number of classes, as pointed out earlier. Thus, in order to get some idea about the mean heterozygosity for the continuous case, we must make some simplifications. We consider only genic selection. Let the relative fitness of the best genotype be 1 and let s be the selection coefficient against any particular mutation. It is clear from the above computations that genic selection causes a great reduction in mean heterozygosity when s is considerably larger than $1/N$ (see Tables 3 and 4). We therefore consider only mutations with selection coefficient of $0 \leq s \leq 1/N$. Assume that $\theta = 4Nv = 0.1$. We first make an effort to see how fast the mean heterozygosity for the discrete case approaches the value for the continuous case as the number of classes increases. To this end, we use a simple model in which the selection coefficient of mutations is uniformly distributed over the interval $[0, 1/N]$. If mutations are not divided into classes and are assumed to be equally fit (neutral), then $\bar{H} = \theta/(1+\theta) = 0.091$. Next, we approximate the continuous model by a model of two equal classes of mutations with $\theta_1 = \theta_2 = \theta/2$. It is easily computed that the mean selection coefficient against the first-class alleles is $1/(4N)$ and that against the second-class alleles is $3/(4N)$. If we assume that the relative mean selection coefficient against the first-class alleles is 0, then that against the second-class alleles is $1/(2N)$. Putting these parameters into formula (31), we obtain $\bar{H} = 0.0815$. Third, we approximate the continuous model by a model of three classes of mutations with $\theta_1 = \theta_2 = \theta_3 = \theta/3$. It is again easily computed that the relative mean selection coefficients against the first-, second- and third-class alleles are 0, $1/(3N)$ and $2/(3N)$. Using formula (29), we obtain $\bar{H} = 0.0796$. For the models of four and five classes of mutations, we obtain $\bar{H} = 0.0789$ and $\bar{H} = 0.0786$, respectively. These results suggest that the mean heterozygosities for the discrete models quickly approach a limit as the number of classes increases—the limit must be larger than 0.0779, the value for the first of the two models given below. Since the difference between 0.0786 and 0.0779 is small, we consider the model of five classes of alleles a good approximation to the continuous model. We now consider another two models of five classes of mutations. In the first model, $\theta_1 = \theta/15$, $\theta_2 = 2\theta/15$, . . . , $\theta_5 = 5\theta/15$, and the selection coefficients against the five classes are 0, $1/(5N)$, . . . ,$4/(5N)$. The mean heterozygosity for this model is $\bar{H} = 0.0779$. In the second model, $\theta_1 = 5\theta/15$, $\theta_2 = 4\theta/15$, . . . , $\theta_5 = \theta/15$, and the selection coefficients are again 0, $1/(5N)$, . . . ,$4/(5N)$. The mean heterozygosity for this model is $\bar{H} = 0.0827$. Thus the $\bar{H}$ value is somewhat larger for the case where the distribution of s is skewed toward 0 than for the case where the distribution of s is skewed toward $1/N$. In all cases, however, the $\bar{H}$ values are at most 15 percent less than 0.091, the value for the model of strictly neutral mutations with $\theta = 0.1$. We note that the prevalent mode of selection in nature is probably less effective than genic selection. We may therefore conclude that, regardless of what the fitness spectrum of mutations may be, all mutations with $s \leq 1/N$ are capable of contributing significantly to the mean heterozygosity of a population. Of course, the dichotomy between $s \leq 1/N$ and $s > 1/N$ is somewhat arbitrary, but it will make discussion easier when we consider protein polymorphism later.

(C) *Variance of heterozygosity*: This variance has been used by NEI and his associates as a test statistic to test the neutral theory (NEI 1975; FUERST, CHAKRABORTY and NEI 1977). The procedure is as follows. Under the null hypothesis of neutral mutations, θ is estimated by equating the observed average heterozygosity to its theoretical expectation, $\bar{H} = \theta/(1+\theta)$. The θ value is then used to compute the expected variance of heterozygosity by using

$$V(H) = 2\theta/[(1+\theta)^2(2+\theta)(3+\theta)] \; ,$$

which is derived under the assumption of neutral mutations (*cf.*, STEWART 1976). The expected variance is compared with the observed variance of heterozygosity over the loci studied. If there is no discrepancy between the two variances, the neutral mutation hypothesis is thought to be tenable; otherwise it is rejected. In Table 5 we examine what selection intensity can be detected by this test pro-

TABLE 5

Variance of heterozygosity

$S = S_1$		1	4	6	10	20
$\theta = 0.05$	$\bar{H}_1$*	0.0473	0.0386	0.02564	0.01204	0.00589
	$V(H_1)$	0.0144	0.0114	0.00694	0.00229	0.00071
	$V(H_1)'$	0.0144	0.0120	0.00815	0.00392	0.00194
	$\bar{H}_2$†	0.0475	0.0395	0.02336	0.01056	0.00544
	$V(H_2)$	0.0145	0.0118	0.00618	0.00180	0.00061
	$V(H_2)'$	0.0145	0.0122	0.00746	0.00345	0.00180
	$\bar{H}_3$‡	0.0478	0.0453	0.03714	0.02713	0.01933
	$V(H_3)$	0.0146	0.0136	0.01047	0.00659	0.00376
	$V(H_3)'$	0.0145	0.0139	0.01154	0.00860	0.00621
$\theta = 0.20$	$\bar{H}_1$*	0.1659	0.1415	0.10014	0.04804	0.02345
	$V(H_1)$	0.0393	0.0345	0.02442	0.00888	0.00277
	$V(H_1)'$	0.0394	0.0355	0.02744	0.01462	0.00748
	$\bar{H}_2$†	0.1663	0.1448	0.09289	0.04205	0.02166
	$V(H_2)$	0.0394	0.0354	0.02264	0.00700	0.00237
	$V(H_2)'$	0.0394	0.0360	0.02584	0.01295	0.00693
	$\bar{H}_3$‡	0.1670	0.1586	0.13147	0.09667	0.06949
	$V(H_3)$	0.0395	0.0374	0.03028	0.01977	0.01166
	$V(H_3)'$	0.0395	0.0382	0.03368	0.02668	0.02027
$\theta = 1.00$	$\bar{H}_1$*	0.4992	0.4743	0.41272	0.23710	0.11400
	$V(H_1)$	0.0417	0.0453	0.05083	0.03641	0.01216
	$V(H_1)'$	0.0417	0.0440	0.04844	0.04729	0.03033
	$\bar{H}_2$†	0.4997	0.4805	0.40661	0.20748	0.10544
	$V(H_2)$	0.0417	0.0448	0.05243	0.03026	0.01048
	$V(H_2)'$	0.0417	0.0435	0.04876	0.04458	0.02857
	$\bar{H}_3$‡	0.5002	0.4839	0.42606	0.32220	0.23798
	$V(H_3)$	0.0416	0.0427	0.04418	0.03526	0.02353
	$V(H_3)'$	0.0416	0.0432	0.04765	0.05077	0.04735

* Three classes of alleles: $\theta_1 = 0.01\theta$, $\theta_2 = 0.09\theta$, $\theta_3 = 0.9\theta$, $S_2 = S_1/2$.

† Two classes of alleles under genic selection: $\theta_1 = 0.01\theta$, $\theta_2 = 0.99\theta$.

‡ Two classes of alleles under recessive selection: $\theta_1 = 0.01\theta$, $\theta_2 = 0.99\theta$.

cedure. The values of $\bar{H}_i$ and $V(H_i)$, $i=1, 2, 3$, are computed by using those formulae given above, whereas $V(H_i)'$ is computed from $\bar{H}$ by using NEI's (1975) procedure under the null hypothesis of neutral mutations; $\bar{H}_1$ and $V(H_1)$ refer to the case of three classes of alleles under genic selection, $\bar{H}_2$ and $V(H_2)$ to the case of two classes of alleles under genic selection, and $\bar{H}_3$ and $V(H_3)$ to the case of two classes of alleles under recessive selection. It is seen that in all of the cases where $4Ns$ is 10 or larger, $V(H_i)'$ deviates considerably from $V(H_i)$, particularly in the case of three classes of alleles under genic selection. Thus, it seems that this test procedure is able to detect a selection intensity of this order. On the other hand, if $4Ns$ is around 4 or less, there is virtually no difference between $V(H_i)'$ and $V(H_i)$. When $4Ns=6$, two discrepancies are appreciably large, while others are not. When $4Ns$ is of this order of magnitude, the discrepancy between $V(H_i)'$ and $V(H_i)$ depends on the mutation rate and the type of selection. Notice that except for some cases with $\theta=1$, $V(H_i)$ is usually smaller than $V(H_i)'$. Since the majority of mean heterozygosities observed so far are less than those for the cases with $\theta=1$, we may conclude that purifying selection tends to reduce the variance of heterozygosity.

DISCUSSION

To emphasize the point that even slight selection has a drastic effect on genetic variability when the effective population size N is large, all the above numerical results were computed for small s values such as $s=10^{-5}$. The results, however, are also applicable to other combinations of N and s, because the effect of selection can be considered in terms of the product Ns when $|4Ns| << N$. For example, the mean heterozygosity should be almost the same for both the case of $s=10^{-2}$ and $N=10^2$ and that of $s=10^{-5}$ and $N=10^5$, provided that Nv is the same for both cases. Note, however, that when dealing with slight selective differences, we may need to consider one additional factor—the random fluctuation of selection intensities. This factor reduces the effectiveness of selection when the mean, $\bar{s}$, of s is larger than its variance $V(s)$. But, when $V(s)$ becomes larger than $\bar{s}$, it increases the random fluctuation of gene frequencies and consequently reduces the amount of genetic variability maintained in a population (WRIGHT 1948; KIMURA 1955; KARLIN and LEVIKSON 1974; NEI and YOKOYAMA 1976), though under certain circumstances it may produce a stabilizing effect on gene frequencies (JENSEN and POLLACK 1969; GILLESPIE 1973; KARLIN and LEVIKSON 1974; and others).

In the present study all the statistical properties of the maintenance of genetic variation are derived under the assumption that every mutation creates a new allele—the model of infinite alleles. This model seems to be appropriate if allelic variants are identified at the nucleotide or codon level. At present, however, genetic variation is mostly studied by electrophoresis. At the electrophoretic level, alleles (electromorphs) presumably mutate only to their nearby states so that mutations are to some extent recurrent and back mutations may occur. Thus, there are some differences between these two levels of detectability and caution should be taken when applying the present results to interpret polymorphism data collected by electrophoresis. The distribution of allele frequencies at the electrophoretic level has been studied only for the case of neutral mutations (KIMURA and OHTA 1975). KIMURA and OHTA's results show that in this case the distribution obtained under the model of stepwise change of electrophoretic mobility is similar to that obtained under the model of infinite alleles when $\theta < 1$. Presumably the previous statement that under balancing selection the distribution tends to be non-U-shaped holds as well at the electrophoretic level. In the case of purifying selection, it is not clear whether or not the distribution at the electrophoretic level is similar to that under the model of infinite alleles. However, as in the latter model, it is unlikely that the distribution can be U-shaped if θ is much larger than one, unless a special arrangement is made of the mutation rates and selection coefficients for the electromorphs. The rate of increase of mean heterozygosity with increasing population size is known, however, to be considerably slower at the electrophoretic level than that predicted by the model of infinite alleles, regardless of whether there is selection or not (OHTA and KIMURA 1973, 1975; LI 1976).

In applying the present results to data, we should also take into consideration the effect of variation in mutation rate over loci. For neutral mutations, this effect has recently been studied by NEI, CHAKRABORTY and FUERST (1976). Their conclusions are that this effect reduces the mean of heterozygosity, but inflates the variance, and that a U-shaped distribution can be obtained for a larger range of (average) $4Nv$ values than that for the case of constant mutation rate; however, it is a tilted U-shape if the mean of $4Nv$ over loci is larger than unity. Presumably these conclusions hold qualitatively for the cases of balancing selection and purifying selection. It should, however, be stressed that this effect causes no major change in the conclusions drawn under the assumption of constant mutation rate.

I now discuss the implications of the present findings for protein polymorphism. For ease of discussion, I first summarize the general patterns of genic variation that have emerged from the huge amount of gene-frequency data collected by electrophoresis (a compilation of available data has recently been made by FUERST, CHAKRABORTY and NEI 1977). (1) A very striking general pattern of genic variation is that the observed distribution (histogram) of allele frequencies is U-shaped for every species studied (unpublished result of CHAKRABORTY, FUERST and NEI). (2) When the observed distributions are compared with those expected under the model of selective neutrality, about one-third of the 140 species examined show a significant excess of rare alleles, but only one species shows a significant deficiency of rare alleles (unpublished result of CHAKRABORTY, FUERST and NEI). (3) There seems to be an upper limit for the observed average heterozygosities (LEWONTIN 1974; NEI 1975; unpublished result of NEI, FUERST and CHAKRABORTY). An hypothesis of the maintenance of genic variation is tenable only if it can explain these observations at least reasonably well.

Interestingly, none of these observations appear to be explicable by the hypothesis of balancing selection. First, the general pattern of U-shaped distribution is not expected under balancing selection (*cf.*, Figures 1a and b). Second, balancing selection should lead to deficiencies rather than excesses of rare alleles. This has been shown for overdominant selection, but should also be true for other types of balancing selection. Finally, the third observation has been taken by Lewontin (1974) and by the selectionists (*e.g.*, Ayala 1972; Milkman 1975) as strong evidence against the neutralist view, but is actually more incompatible with the selectionist view, for under balancing selection the average heterozygosity should be larger than under selective neutrality. For instance, the numerical examples given in the previous two sections show that even very slight overdominant selection, as small as $s = 10^{-5}$, increases the average heterozygosity considerably as compared to the case of no selection. Because of these difficulties, we are reluctant to accept balancing selection as an important cause for the maintenance of genic variation.

We now consider the neutralist explanation. The selectionists strongly maintain that the neutral theory is wrong because the average heterozygosities in the *Drosophila willistoni* group (Ayala *et al.*, 1974) and in *Escherichia coli* (Milkman 1975) are much less than would be expected from the balance between mutation and random genetic drift. While the neutralists believe that this difficulty is resolvable, they disagree with each other to some extent on how to resolve it. On the one hand, Ohta (1974, 1976) thinks that some modification of the original theory is necessary and has proposed a modified hypothesis—the hypothesis of slightly deleterious mutations. On the other hand, Nei thinks that no modification is necessary because the long-term effective size of these species may be small or these species may have gone through a bottleneck in the recent past (Nei 1975, 1976; Nei, Maruyama and Chakraborty 1975).

Let us first examine Nei's resolution. The general pattern of U-shaped distribution strongly supports Nei's view that the effective sizes of natural populations are rather limited, because if the effective size is very large, a U-shaped distribution is unlikely to be obtained under any hypothesis. However, to explain the apparent upper limit of observed average heterozygosities by the neutral theory, the θ values for the species studied must be at most of the order of 0.6. This is because the highest average heterozygosity observed so far is 0.309 (in *Otiorrhynchus scaber*, Suomalainen and Saura 1973), from which we obtain $\theta = 0.60$ by using the formula $\overline{H} = 1 - 1/\sqrt{1+2\theta}$ derived under the model of stepwise mutation (Ohta and Kimura 1973). (A somewhat larger θ value is obtained if variation of mutation rate over loci is taken into account (Nei, Chakraborty and Fuerst 1976).) Whether this condition is reasonable or not is difficult to tell because we are unable to determine the long-term effective size of populations. This hypothesis, however, can be tested by considering some other aspects of genic variation, such as the variance of heterozygosity, the incidence of rare alleles, etc. A detailed analysis of available gene frequency data shows that in most species the observed variance of heterozygosity agrees reasonably well with that expected under the neutral theory (FUERST, CHAKRABORTY and NEI 1977). The excess of rare alleles in the *D. willistoni* group has been taken by OHTA (1976) as an indication of the prevalence of slightly deleterious mutations. However, this observation is not necessarily incompatible with the neutral theory because it is also explainable by recent population expansion (NEI and LI 1976).

Next let us examine OHTA's hypothesis that the genic variation of a population is mostly due to slightly deleterious mutations. The essence of OHTA's (1976) theory is that in small populations slightly deleterious alleles behave just like neutral alleles, while in large populations "the stable mutation-selection balance is reached and this provides an upper limit to heterozygosity." OHTA (1976) believes that her hypothesis can account for the relative uniformity of observed average heterozygosities over various species (LEWONTIN 1974) and the excess of rare alleles in some species. The present findings seem to support her contentions. However, her assumption (OHTA 1976) that large natural populations such as the *D. willistoni* group species are at the stable mutation-selection balance leads to the following two difficulties: (1) This assumption implies that the θ values for these populations are very large. [OHTA (1976) seems to agree with AYALA *et al.* (1974) that the population sizes of the *D. willistoni* group are extremely large, though she may not accept their estimate of $\theta = 400$ for these species.] But a U-shaped distribution is unlikely to be observed if $\theta >> 1$. Thus, under this assumption it is difficult to explain the general pattern of U-shaped distribution. (2) Because of the stable mutation-selection balance, evolution is supposed to stop in large populations, as OHTA (1976) herself noted. In practice, however, studies on genetic distance suggest that even in large populations such as the *D. willistoni* species (AYALA *et al.*, 1974) gene substitution has proceeded continuously with time.

OHTA's purpose in making the assumption of mutation-selection balance is to explain why the observed average heterozygosity is relatively uniform over various species surveyed (LEWONTIN 1974) and why there is an apparent upper limit for average heterozygosity (OHTA 1974, 1976). I believe that these two problems can be resolved without making this assumption, but by simply assuming that the genic variation of a population is mainly due to slightly deleterious mutations. Consider the first problem. To simplify the argument, let us assume that only genic selection is operating. In a population of effective size N, mutations can be divided somewhat arbitrarily into the following two classes: mutations with selective disadvantage $s \leq 1/N$ and mutations with $s > 1/N$. As seen earlier, the mean heterozygosity due to mutations with $s \leq 1/N$ is comparable to the case of neutral mutations, but that due to mutations with $s > 1/N$ is much less than what is expected under selective neutrality. Obviously, as N increases the proportion of the first-class mutations decreases, though the proportion of the second-class mutations increases. For example, if $N = 10^3$ the first class includes all mutations with $s \leq 10^{-3}$ but if $N = 10^4$ it includes only all mutations with $s \leq 10^{-4}$. When N increases from 10^3 to 10^4, the mean heterozygosity due to mutations with $s \leq 10^{-4}$ increases, but that due to mutations with $s > 10^{-4}$

decreases. Therefore, the $\bar{H}$ values for $N = 10^4$ may not be much larger than that for $N = 10^3$. The rate of increase of $\bar{H}$ with increasing N depends on how s is distributed, but should be much slower than that for strictly neutral mutations, unless the distribution of s is very skewed toward 0. This conclusion should also be qualitatively true for other types of purifying selection. Thus the first problem is resolved.

In the above model it is possible for $\bar{H}$ to become very large if N becomes exceedingly large. Then why is there an apparent upper limit for the observed average heterozygosities? My answer is that it is due to restricted effective sizes. As mentioned above, this view is strongly supported by the general pattern of U-shaped distribution. AYALA *et al.* (1974) estimate that the effective size for the species of the *D. willistoni* group is about 10^{10} and $v = 10^{-8}$ so that $4Nv$ is about 400. The fact that the distributions of allele frequencies for these species are U-shaped strongly suggests that this is a gross overestimate. Although presently the actual total size of each of these species may be large, it should be noted that the effective population size in the long evolutionary history is generally much smaller than the total size. (The human population is such an example.) As NEI, MARUYAMA and CHAKRABORTY (1975) have emphasized, if a population occasionally goes through a bottleneck the effective size is greatly reduced. It is interesting to note that even the size of a laboratory population that is maintained at a constant temperature and humidity fluctuates greatly (NOGUÉS 1977). For a similar reason, MILKMAN's (1975) estimate of $N = 10^{10}$ for *E. coli* for the last 40 million years has been challenged by NEI (1976) and WILSON (1976). At any rate, these estimates of effective sizes are highly speculative, since we do not even know their present actual sizes.

The advantage of this modified version over OHTA's original hypothesis is that it removes the difficulties created by the assumption of mutation-selection balance. In particular, it now can explain the general pattern of U-shaped distribution at least as well as the original neutral mutation hypothesis and the selectionist hypothesis, because purifying selection has the highest potential to maintain a U-shaped distribution, as shown above. Furthermore, like OHTA's hypothesis, it can better account for the excess of rare alleles in some species than the original neutral mutation hypothesis and the selectionist hypothesis. Hence the present hypothesis seems to explain better the general patterns of genic variation than other current ones.

However, the difference between KIMURA's and my hypothesis is very subtle. KIMURA (1968) called an allele "almost neutral" when $2Ns << 1$. This definition seems too strict. The present results suggest that a more reasonable definition of almost neutrality is $s \leq 1/N$, for the effect of random drift remains strong for mutations with selection coefficients of this order or smaller. I consider my hypothesis equivalent to this extended form of KIMURA's hypothesis. My modified version of the neutral mutation hypothesis is similar to OHTA's, but I do not assume that large natural populations are at the stable mutation-selection balance. I consider this balance unlikely to occur in nature because as the selective difference becomes very small it is unlikely to stay constant and to play a decisive role so as to maintain a stable balance.

I am greatly indebted to M. Nei for his constant help and valuable suggestions. I also thank G. A. Watterson, W. J. Schull, J. L. King and R. C. Lewontin for helpful comments and P. Fuerst for showing me the electrophoretic data that he has compiled. This study is supported by Public Health Service grant GM 20293 and National Science Foundation grant DEB 77-09120.

LITERATURE CITED

Allison, A. C., 1961 Genetic factors in resistance to malaria. Ann. N. Y. Acad. Sci. **91**: 710-729.

Ayala, F. J., 1972 Darwinian versus non-Darwinian evolution in natural populations of Drosophila. Proc. 6th Berkeley Symp. Math. Statist. Probab. **5**: 211-236.

Ayala, F. J., M. L. Tracey, L. G. Barr, J. F. McDonald and S. Perez-Salas, 1974 Genetic variation in natural populations of five Drosophila species and the hypothesis of selective neutrality of protein polymorphisms. Genetics **77**: 343-384.

Bodmer, W. F. and L. L. Cavalli-Sforza, 1972 Variation in fitness and molecular evolution. Proc. 6th Berkeley Symp. Math. Statist. Probab. **5**: 255-275.

Crow, J. F., 1972 Darwinian and non-Darwinian evolution. Proc. 6th Berkeley Symp. Math. Statist. Probab. **5**: 1-22.

Ewens, W. J., 1972 The sampling theory of selectively neutral alleles. Theoret. Popul. Biol. **3**: 87-112.

Fuerst, P. A., R. Chakraborty and M. Nei, 1977 Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. Genetics **86**: 455-483.

Gillespie, J. H., 1973 Natural selection with varying selection coefficients—a haploid model. Genet. Res. **21**: 115-120.

Jensen, L. and E. Pollack, 1969 Random selective advantages of a gene in a finite population. J. Appl. Probab. **6**: 19-37.

Johnson, N. L. and S. Kotz, 1972 *Distribution in Statistics: Continuous Multivariate Distributions.* John Wiley & Sons, New York.

Karlin, S. and B. Levikson, 1974 Temporal fluctuations in selection intensities: case of small population size. Theoret. Popul. Biol. **6**: 383-412.

Karlin, S. and J. McGregor, 1967 The number of mutant forms maintained in a population. Proc. 5th Berkeley Symp. Math. Statist. Probab. **4**: 415-438.

Kimura, M., 1955 Stochastic processes and distribution of gene frequencies under natural selection. Cold Spring Harbor Symp. Quant. Biol. **20**: 33-53. ——, 1956 Stochastic processes in population genetics. Ph.D. Thesis, University of Wisconsin, Madison. ——, 1968 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11**: 247-269.

Kimura, M. and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. Genetics **49**: 725-738.

Kimura, M. and T. Ohta, 1975 Distribution of allelic frequencies in a finite population under stepwise production of neutral alleles. Proc. Nat. Acad. Sci. U.S. **72**: 2761-2764.

King, J. L., 1972 The role of mutation in evolution. Proc. 6th Berkeley Symp. Math. Statist. Probab. **5**: 69-100.

Lewontin, R. C., 1974 *The Genetic Basis of Evolutionary Change.* Columbia Univ. Press, New York.

Li, W.-H., 1976 A mixed model of mutation for electrophoretic identity of proteins within and between populations. Genetics **83**: 423–432. ——, 1977 Maintenance of genetic variability under mutation and selection pressures in a finite population. Proc. Nat. Acad. Sci. U.S. **74**: 2509–2513.

Li, W.-H. and M. Nei, 1975 Drift variances of heterozygosity and genetic distance in transient states, Genet. Res. **25**: 229–248.

Milkman, R., 1975 Allozyme variation in *E. coli* of diverse natural origins. pp. 273–285. In: *Isozymes* VI. Edited by C. L. Markert. Academic Press, New York.

Nei, M., 1975 *Molecular Population Genetics and Evolution.* North-Holland, Amsterdam. ——, 1976 The cost of natural selection and the extent of enzyme polymorphism. Trends Biochem. Sci., Nov: N247–N248. ——, 1977 Estimation of mutation rate from rare protein variants. Am. J. Hum. Genet. **29**: 225–232.

Nei, M., R. Chakraborty and P. A. Fuerst, 1976 Infinite allele model with varying mutation rate. Proc. Nat. Acad. Sci. U.S. **73**: 4164–4168.

Nei, M. and W.-H. Li, 1976 The transient distribution of allele frequencies under mutation pressure. Genet. Res. **28**: 205–214.

Nei, M., T. Maruyama and R. Chakraborty, 1975 The bottleneck effect and genetic variability in populations. Evolution **29**: 1–10.

Nei, M. and S. Yokoyama, 1976 Effects of random fluctuation of selection intensity on genetic variability in a finite population. Japan. J. Genet. **51**: 355–369.

Nogués, R. M., 1977 Population size fluctuation in the evolution of experimental cultures of *Drosophila subobscura.* Evolution **31**: 200–213.

Ohta, T., 1974 Mutational pressure as the main cause of molecular evolution and polymorphisms. Nature **252**: 351–354. ——, 1976 Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theoret. Popul. Biol. **10**: 254–275. ——, 1978 Extension to the neutral mutation random drift hypothesis. pp. 148–167. In: Proc. 2nd Taniguchi Internat'l. Symp. on Biophysics, Molecular Evolution and Polymorphism. Edited by M. Kimura.

Ohta, T. and M. Kimura, 1973 A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genet. Res. **22**: 201–204. ——, 1975 Theoretical analysis of electrophoretically detectable polymorphism: Models of very slightly deleterious mutations. Am. Naturalist **109**: 137–145.

Stewart, F. M., 1976 Variability in the amount of heterozygosity maintained by neutral mutations. Theoret. Popul. Biol. **9**: 188–201.

Suomalainen, E. and A. Saura, 1973 Genetic polymorphism and evolution in parthenogenetic animals. I. Polyploid Curculionidae. Genetics **74**: 489–508.

Watterson, G. A., 1974 Models for the logarithmic species abundance distributions. Theoret. Popul. Biol. **6**: 217–250. ——, 1977 Heterosis or neutrality? Genetics **85**: 789–814. ——, 1978 The homozygosity test of neutrality. Genetics **88**: 405–417.

Wilson, A. C., 1976 Irrelevant to evolution studies. Trends Biochem. Sci., Aug: N180–N181.

Wright, S., 1948 One the roles of directed and random changes in gene frequencies in the genetics of populations. Evolution **2**: 279–294. ——, 1949a Adaptation and selection. pp. 365–389. In: *Genetics, Paleontology and Evolution.* Edited by G. L. Jepson, G. G. Simpson and E. Mayr. Princeton Univ. Press, Princeton. ——, 1949b Genetics of populations. Encyclopaedia Britanica, 14 ed., **10**: 111–112. ——, 1966 Polyallelic random drift in relation to evolution. Proc. Nat. Acad. Sci. U.S. **55**: 1074–1081.

Corresponding editor: D. L. Hartl