EFFECTS OF RANDOM FLUCTUATION OF SELECTION INTENSITY ON GENETIC VARIABILITY IN A FINITE POPULATION

MASATOSHI NEI
Center for Demographic and Population Genetics, University of Texas,
Houston, Texas 77030

and

SHOZO YOKOYAMA
Department of Genetics, University of Washington,
Seattle, Washington 98195

Received August 30, 1976

The average heterozygosity per locus is an important measure of genetic variability of a randomly mating population. Theoretical study on this quantity was initiated by Kimura and Crow (1964), who investigated the effective number of alleles defined as the reciprocal of one minus heterozygosity. In this study it was assumed that new mutations are always different from the alleles extant in the population. At the molecular level, the rate of nucleotide substitution per DNA base per generation is so low, that this assumption seems to hold approximately. Extending this work, Kimura (1969) studied the expected number of heterozygous nucleotide sites per individual and the expected number of segregating nucleotide sites in a population. Maruyama (1972) showed that there is a simple relationship between the expected number of heterozygotes and gene frequency.

All these studies are based on the assumption of neutral genes or selective genes with constant fitness. In nature, however, selection intensity may fluctuate from generation to generation, owing to environmental changes (Fisher and Ford 1947). This fluctuation of selection intensity is expected to increase the random fluctuation of gene frequencies and consequently decrease the amount of genetic variability maintained in a population (Wright 1948; Kimura 1955), though under certain circumstances it may produce an effect to stabilize gene frequency in large populations (Jensen and Pollak 1969; Gillespie 1973; and others). The importance of random fluctuation of selection intensity was emphasized by Crow (1972a) in explaining the apparent monomorphism of fibrinopeptide A in human populations despite a high rate of amino acid substitution in the evolutionary process. Mathematical studies on this subject have recently been reviewed by Karlin and Lieberman (1974).

In the present paper we shall study the effects of random fluctuation of selection intensity on the amount of genetic variability in populations. There are two ways to study this problem. One is to evaluate the expected heterozygosity per locus, using
the so-called infinite allele model (Kimura 1971). The other is to compute the expected number of heterozygous codons per locus (or nucleotide pairs), using the infinite site model. In the following we shall first use the second approach, since it is mathematically simpler than the first. Since our main purpose is to see the effects of random fluctuation of selection intensity, we shall assume that each mutant allele or codon is selectively neutral on the long-term basis but in a particular generation it may be subject to natural selection, the intensity of which varies at random over generation. This assumption simplifies the mathematical treatment considerably.

MODELS OF NATURAL SELECTION

Before going into the detail of our study, we need some discussion on the mathematical model of natural selection to be used. In the following we designate by $A'$ and $A$ the mutant and original alleles at a particular locus, respectively, and let $x$ be the frequency of $A'$.

In his study of stochastic change of gene frequency with random fluctuation of selection intensity, Kimura (1955) used a continuous time model, measuring genotype fitnesses by the Malthusian parameter. His mathematical treatment, however, includes several approximations which are not very clear to us (see, however, Karlin and Levikson 1974, p. 409). In recent years there have been some attempts to treat this problem more rigorously.

In Wright's discrete time model the relative fitnesses of three possible genotypes, $AA$, $AA'$, and $A'A'$, may be denoted by $1-s$, 1, and $1+s$, respectively, for the case of genic selection. We assume that $s$ fluctuates at random with mean $\bar{s}=0$ and variance $V_s$. (Here we are assuming a negative correlation between the fitnesses of $AA$ and $A'A'$. We shall discuss this problem later.) In diffusion approximations to discrete genetic processes it is customary to assume that $N$ adults produce many offspring, selection acts on the offspring deterministically, and then, independent of selection intensity, a random sample of size $N$ is drawn to form the adults for the next generation (Feller 1951). In the present case, the amount of change of gene frequency per generation is

$$dx=dx_s+dx_e,$$

where $dx_s$ and $dx_e$ are the changes due to selection and sampling error, respectively. Clearly, $dx_s$ and $dx_e$ are independent of each other. The mean and variance of $dx_s$ are given by 0 and $x(1-x)/(2N)+O(E(dx_s)/N)$, respectively. On the other hand, the expectation of $dx_e$ with respect to the distribution of $s$ is

$$E(dx_e)=E\left(\frac{sx(1-x)}{1+s(2x-1)}\right)$$

$$=V_s x(1-x)(1-2x)+O(E(s^4)).$$

Clearly, $E(dx)$ is equal to $E(dx_s)$. Therefore, the mean change of gene frequency per generation in diffusion approximations is given by

$$M_{dx}=V_s x(1-x)(1-2x).$$

(1)
Similar but slightly different formulations of $M_s$ have been obtained by Jensen and Pollak (1969), Gillespie (1973), Felsenstein (Crow 1972b), and Jensen (1973).

The above formula for $M_s$ indicates that the gene frequency tends to return to 0.5 whenever it deviates from this value. This tendency is caused by the denominator in $A x = s x (1 - x) / (1 + s (2 x - 1))$. In diffusion approximations with constant fitness, the denominator or the mean fitness plays little role except as a normalizing factor for gene frequency. However, when selection coefficient fluctuates at random from generation to generation, this is no longer true. Thus, if we denote the fitnesses of $AA$, $AA'$, and $A'A'$ by $1$, $1 + s$, and $1 + 2 s$, respectively, then we obtain $M_s = -2 V_s x (1 - x)$. In this case the mutant gene behaves as if it were a deleterious recessive. On the other hand, if we denote the three fitnesses by $1 - 2 s$, $1 - s$, and $1$, respectively, the formula for $M_s$ is $2 V_s x (1 - x)^2$. The mutant gene now behaves as if it were a dominant favored. Note that all the above three models are the same type of genic selection and it is not easy to choose one of them for a theoretical study (Nei 1976). If selection coefficient is constant over generations, they all lead to the same formula for $M_s$.

In the above argument we have used a special case of correlated fitnesses for the three genotypes. However, our conclusion remains the same even if we consider a more general situation. Karlin and Levikson (1974) presented a general formula for $M_s$ (formula 3.13), taking into account all variances and covariances of genotype fitnesses. It is a complicated formula dependent on nine parameters but shows that a slight change in the variances and covariances of genotype fitnesses often results in a quite different value of $M_s$. In practice, we will almost never be able to determine the values of all parameters involved experimentally.

While there is nothing wrong in the above or Karlin and Levikson's mathematical treatment, we suspect that the model used is not very realistic for the present purpose. As was emphasized by Feller (1967), the mean fitness ($\bar{W}$) in the usual Wright model reflects the change of population size after selection compared with that in the previous generation at the same stage, assuming that the number of individuals at birth is the same for all generations. In the present case, $\bar{W}$ changes from generation to generation, owing to the change in gene frequency as well as to the fluctuation of selection coefficient. That is, the population size at a particular developmental stage at which natural selection operates changes in every generation. In nature, however, the size of a population is regulated and remains roughly constant at steady state. This regulation of population size seems to occur at all stages of life cycle, not just at the time of maturity as was assumed in the above formulation. In fact, in many organisms the number of individuals at an early stage of life cycle is often determined by the amount of limited resource or by the number of hiding places against predators or unfavorable weather factors. Namely, a large part of natural selection seems to occur through competition for limited resource or space. If this is the case, the competition model developed by Mather (1969) and extended by Nei (1971) considering population regulation may be more realistic, as will be discussed in the following.

In this competition model it is assumed that there are $N$ adults in the population and they produce a large number of offspring. Because of natural selection and ac-
cidental deaths, the number of individuals gradually declines after birth and $N$ individuals survive to maturity. The number of individuals at a particular developmental stage remains the same for all generations. This is because population size is determined by outside factors such as resource and space. Consider a pair of alleles, $A$ and $A'$, which act at an early stage of life, so that natural selection with respect to this locus can be regarded as deterministic. The number of survivors after selection may still be large but only $N$ individuals will eventually survive to maturity because of subsequent deaths independent of the locus. We assume that natural selection occurs through competition between a random pair of individuals for the limiting resource and space. Competition may occur between individuals of the same genotype or different genotypes, the probabilities of occurrence of intra- and inter-genotypic competitions depending on the genotype frequencies. The fitness of a genotype is the average survival of this genotype when it competes with the same and different genotypes. Mather (1969) and Nei (1971) have shown that when allele $A'$ is competitively stronger than $A$ and gene action is additive, the fitnesses of $AA$, $AA'$, and $A'A'$ are given by $1-2xs$, $1+(1-2x)s$ and $1+2(1-x)s$, respectively, where $s$ is the competitive advantage of gene $A'$ over $A$ such that when $AA$ and $AA'$ or $AA'$ and $A'A'$ are in competition their survival values are $1-s$ and $1+s$, respectively. Since the frequencies of $AA$, $AA'$, and $A'A'$ under random mating are $(1-x)^2$, $2x(1-x)$, and $x^2$, respectively, the mean fitness is always 1 irrespective of gene frequency and selection coefficient. After selection, $2N$ genes are chosen at random from the gene pool of survivors. Then, it can be shown that the expected amount of change of gene frequency per generation for a fixed value of $s$ is $E(dx|s)=sx(1-x)$. In the above we assumed pairwise competition, but the same result is obtained even if more than two individuals compete for a unit resource as long as each individual behaves independently (Nei 1971).

In the present paper we assume that $s$ fluctuates at random with mean $\bar{s}=0$ and variance $V_s$, as before. Therefore, $E(dx)\equiv 0$, and

$$M_s=0.$$  \hspace{1cm} (2)

On the other hand, the expected variance of $dx$ per generation is

$$E((dx)^2)=E((dx_s)^2) + E((dx_a)^2)
\nonumber
=V_s(x^2(1-x)^2+x(1-x)/(2N)) + O[E(s^3)/N].$$

The variance of gene frequency change per generation in diffusion approximations is therefore given by

$$V_{s} = V_s(x^2(1-x)^2+x(1-x)/(2N)).$$  \hspace{1cm} (3)

Incidentally, this formula applies also for the Wright model.

In the above formulation we assumed that selection operates at an early stage of development. Under certain assumptions, however, it can be shown that formulae (2) and (3) hold irrespective of the stage at which competitive selection occurs (Appendix). It is also noted that in his study of gene frequency distribution Kimura (1955) used the same formulae as (2) and (3), though his derivation is not very clear.

In the following we shall study the effect of random fluctuation in selection in-
tensity on genetic variability by using the Mather-Nei model (except in some special cases), since this model seems to be more realistic at least to us. (Data on molecular polymorphism and evolution are not consistent with the predictions from the Wright model; see DISCUSSION.) Furthermore, some studies based on the Wright model have already been reported by Karlin and Levikson (1974) and Takahata et al. (1975). Before leaving the models of natural selection, however, we should like to make a brief comment on the Wright model of natural selection. As mentioned earlier, this model does not appear to be adequate for the present purpose. However, if selection coefficient is small and remains constant, there is not much difference between the Wright model and the Mather-Nei model. Also, in a population which is growing geometrically the Wright model may be more realistic than the Mather-Nei model, though geometric growth cannot continue indefinitely. Nevertheless, experimental data indicate that the actual process of natural selection is very complicated and the Wright model often does not hold even approximately except for some deleterious genes (Nei 1971). Which model is better for describing natural selection should eventually be determined by conducting experiments.

**NUMBER OF HETEROZYGOUS CODONS PER LOCUS**

Let us now study the number of heterozygous codons per locus by using the competition model mentioned above. Consider a random mating population of size \( N \). Let \( v \) be the mutation rate per locus. Thus, if the mutation rate at the \( i \)-th codon is \( v_i \), then \( v = \sum_i v_i \), where \( n \) is the number of codons at a locus. We assume that whenever a mutant appears it represents a mutation at a previously homoallelic site. This assumption appears to hold approximately, since the mutation rate at a codon is apparently of the order of \( 10^{-9} \) per year. We designate by \( A' \) and \( A \) the mutant and original codons at a particular site, respectively, and by \( x \) the frequency of \( A' \). We assume that at this codon site natural selection occurs according to the competition model discussed above with respect to codons \( A \) and \( A' \), and any other codons varying in the same cistron are selected independently. Under this assumption, the mean and variance of the change in \( x \) per generation are given by (2) and (3), respectively. In the following we replace \( N \) in (3) by the effective population size \( N_e \) to take into account the effects of separate sexes, nonrandom mating, etc.

The expected number of heterozygous codons per locus (\( H_e \)) can be obtained by the following formula:

\[
H_e = \int_0^1 2x(1-x)\Phi_1(x)dx,
\]

where \( \Phi_1(x)dx \) is the expected number of codons with gene frequencies between \( x \) and \( x + dx \). This function \( \Phi_1(x) \) is the same as the gene frequency distribution under irreversible mutation by Wright (1938). Thus, if \( \Phi_1(x) \) is found, \( H_e \) is readily obtained. The distribution \( \Phi_1(x) \) may be obtained by using Kimura's (1964; 1969) general formula:
\[
\Phi_1(x) = \frac{2v}{V_s G(x)} \int_x^1 G(y) dy / \int_0^1 G(y) dy,
\]
where \(1/(2N) \leq x \leq 1\) and
\[
G(x) = \exp \left\{ - \int_0^x \frac{2M_y}{V_s} dy \right\}.
\]
Therefore, using (2) and (3), we get
\[
\Phi_1(x) = \frac{4N_v}{2N_s V_s x^2 (1-x) + x},
\]
approximately, where \(K = 2N_s V_s\) and \(C = \sqrt{K^2 + 4K}\).

By using l'Hospital's rule it can be shown that when \(V_s \to 0\), formula (5) reduces to
\[
H_e = 4N_v v,
\]
which is equal to Kimura's (1969) formula for the number of heterozygous nucleotide sites per individuals if \(v\) is replaced by the mutation rate per gamete.

It is noted that (5) becomes the same as Crow's (1972a) formula for the average fixation time if we divide it by \(v\). Thus, \(H_e\) decreases as \(N_s V_s\) increases in the same way as the fixation time does (see Figure 3 in Crow's paper). Clearly, a large value of \(N_s V_s\) has the same effect as that of reducing the effective population size.

Maruyama (1972) showed that the expected number of heterozygotes for neutral mutations at a given gene frequency \(x\) is given by \(h(x) \approx 4(1-x)\). When \(s\) fluctuates at random, \(h(x)\) is given by
\[
h(x) = 2x(1-x) \Phi_1(x)
\]
\[
= \frac{8N_v v (1-x)}{2N_s V_s x (1-x) + 1}.
\]

Fig. 1. Relationship between the expected number of heterozygotes and gene frequency. The numbers beside the curves represent the value of \(N_s V_s\). The ordinate is in units of \(2N_v v\).
If $N_eV_s$ is 0, this becomes essentially the same as Maruyama’s (1972) formula. However, if $N_eV_s>0$, $h(x)$ is not linearly related with $x$. Figure 1 shows the relationships between $h(x)$ and $x$ for some values of $N_eV_s$. It is clear that $h(x)$ rapidly decreases as $x$ increases if $N_eV_s$ is large.

**AVERAGE HETEROZYGOITY PER LOCUS**

The heterozygosity at a locus is defined as $H=1-\Sigma x_i^2$, where $x_i$ is the frequency of the $i$-th allele in the population. Let us now study the expectation of this quantity, taking into account random fluctuation of selection intensity. In this section we consider an allele (codon sequence) as a unit.

Suppose that there are $k$ possible alleles at a locus and each allele mutates with a rate of $v/(k-1)$ to one of the $k-1$ alleles, so that $v$ is the mutation rate per locus per generation. Let $x$ be the frequency of a particular allele $A'$. We again assume that natural selection occurs according to the Mather-Nei model with respect to genotypes $AA$, $AA'$, and $A'A'$, where $A$ denotes a collection of all other alleles existing in the population. The mean of the change of $x$ per generation ($M_{sa}$) is now $-vx+(1-x)v/(k-1)$ or $-vx$ approximately, since $k$ is generally extremely large. On the other hand, the variance $V_{sa}$ is again given by (3).

Under this circumstance, the expected number of alleles of which the frequency is in the range $x$ to $x+dx$ is given by

$$\Phi(x)dx = \frac{C_0}{V_{sa}} e^{\int_0^x (M_{sa} - v_2y)dy} dx,$$

where $C_0$ is a constant such that $\int_0^1 x\Phi(x)dx = 1$ (Kimura and Crow 1964). Therefore we have

$$\Phi(x) = C_0x^{-1} (1-x)^{4N_e} (\lambda_1-x)^{4N_eA-1}(x-\lambda_2)^{4N_eB-1},$$

where

$$\lambda_1 = (1+\sqrt{1+2/(N_eV_s)})/2, \quad \lambda_2 = (1-\sqrt{1+2/(N_eV_s)})/2, \quad A = -v\lambda_1/(\lambda_1-\lambda_2), \quad B = v\lambda_2/(\lambda_1-\lambda_2).$$

The average heterozygosity is then obtained by

$$H = 1 - \int_0^1 x^2\Phi(x)dx$$

$$= \int_0^1 P(x)dx / \int_0^1 Q(x)dx,$$

where

$$P(x) = (1-x)^{4N_e} (\lambda_1-x)^{4N_eA-1}(x-\lambda_2)^{4N_eB-1} \quad \text{and} \quad Q(x) = P(x)/(1-x).$$

To evaluate $H$, numerical integrations are required. For this computation we used the Gauss-Legendre method of numerical integration. The results obtained are given in Figure 2. As expected, for a given value of $N_e$ the average heterozygosity decreases
as \( V \), or \( N_i V \), increases. However, if \( N_i V \) is smaller than 1, \( H \) is almost the same as that for \( N_i V = 0 \). On the other hand, if \( N_i V \) is larger than 1, \( H \) is considerably reduced particularly in large populations.

**DISCUSSION**

In the foregoing sections we have studied the expected number of heterozygous codons per locus and the average heterozygosity per locus separately. Under certain conditions these two quantities can be related. Following Kimura (1969), the probability of a particular codon site being heterozygous is \( H_i/n \), where \( n \) is the number of codons at a locus and the mutation rate is assumed to be the same for all codons. (If there are any invariant codons, they are disregarded here.) Thus, the probability that the locus is heterozygous at one or more codon sites \( (H_i) \) is \( 1 - (1 - H_i/n)^n = 1 - e^{-H_i} \), since \( H_i/n \) is a small quantity. Thus, using (5) we have

\[
H_i = 1 - \left[ \frac{C - K}{C + K} \right]^{K/n} = \frac{s/e}{s/C}.
\] (10)

The above computation depends on the assumption that the frequency change of each codon is independent of all others. In practice the codons in a cistron are tightly linked, so that \( H_i \) does not necessarily agree with \( H \). Table 1, however, shows that the agreement between \( H \) and \( H_i \) are quite satisfactory unless \( H \) is very high. In the computation of the values in Table 1 the mutation rate per locus \( (v) \) was assumed to be \( 10^{-8} \).

The second method to relate the two quantities is intuitive. For a truly neutral locus, \( H \) is equal to

\[
H_2 = H_i/(1 + H_i),
\] (11)
Table 1. Relationships between the average heterozygosity ($H$) and the expected number of heterozygous codons per locus ($H_0$) (See the text for $H_1$ and $H_2$)

<table>
<thead>
<tr>
<th>$N_0$</th>
<th>$V_s$</th>
<th>$H_s$</th>
<th>$H$</th>
<th>$H_1$</th>
<th>$H_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^2$</td>
<td>$10^{-3}$</td>
<td>.00304</td>
<td>.00303</td>
<td>.00304</td>
<td>.00303</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>.00399</td>
<td>.00397</td>
<td>.00398</td>
<td>.00397</td>
</tr>
<tr>
<td></td>
<td>$10^{-6}$</td>
<td>.00400</td>
<td>.00398</td>
<td>.00399</td>
<td>.00398</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>.00400</td>
<td>.00398</td>
<td>.00399</td>
<td>.00398</td>
</tr>
<tr>
<td>$10^3$</td>
<td>$10^{-2}$</td>
<td>.01128</td>
<td>.01110</td>
<td>.01122</td>
<td>.01115</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>.03872</td>
<td>.03727</td>
<td>.03798</td>
<td>.03721</td>
</tr>
<tr>
<td></td>
<td>$10^{-6}$</td>
<td>.03999</td>
<td>.03845</td>
<td>.03920</td>
<td>.03845</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>.04000</td>
<td>.03846</td>
<td>.03921</td>
<td>.03846</td>
</tr>
<tr>
<td>$10^4$</td>
<td>$10^{-2}$</td>
<td>.02102</td>
<td>.01981</td>
<td>.02080</td>
<td>.02059</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>.30414</td>
<td>.23169</td>
<td>.26224</td>
<td>.23231</td>
</tr>
<tr>
<td></td>
<td>$10^{-6}$</td>
<td>.39867</td>
<td>.28502</td>
<td>.32879</td>
<td>.28504</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>.40000</td>
<td>.28571</td>
<td>.32879</td>
<td>.28504</td>
</tr>
<tr>
<td>$10^5$</td>
<td>$10^{-2}$</td>
<td>.03038</td>
<td>.02587</td>
<td>.02992</td>
<td>.02948</td>
</tr>
<tr>
<td></td>
<td>$10^{-4}$</td>
<td>1.12793</td>
<td>.52817</td>
<td>.67630</td>
<td>.53006</td>
</tr>
<tr>
<td></td>
<td>$10^{-6}$</td>
<td>3.87177</td>
<td>.79619</td>
<td>.97918</td>
<td>.79474</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>4.00000</td>
<td>.80000</td>
<td>.98168</td>
<td>.80000</td>
</tr>
</tbody>
</table>

where $H_0 = 4N_v \nu$. Random fluctuation of selection intensity produces approximately the same effect as that of random sampling of genes. Thus, the above relation may also hold for the present case. In fact, as seen from Table 1, $H_0$ is very close to $H$ except for the case where $N_s V_s$ is large.

In the present paper we have been concerned only with the amount of heterozygosity per locus. Another important parameter in population genetics is the number of segregating codons (Kimura 1969) or the number of segregating loci (Fisher 1930) in a population. The expected number of segregating codons per locus can be obtained as follows:

$$S = \int_{1/2N}^{1} \phi(x) dx$$

$$= 4N_v \nu \left\{ \log_e (2N) - \frac{K}{C} \log_e \left( \frac{C+K}{C-K} \right) \right\}$$

(12)

approximately, when $V_s \ll 1$ and $N \gg 1$. When $V_s \to 0$, the above formula reduces to

$$S = 4N_v \nu \log_e (2N),$$

(13)

which is essentially the same as Kimura's (1969) formula. In the case where one mutation occurs in each generation ($2N_v = 1$), it becomes the same as Fisher's (1930) formula for the number of segregating loci in a population, except for the correction factors for discrete gene frequencies.

Formula (12) indicates that the expected number of segregating codons decreases as $N_s V_s$ increases, as expected.

At the present time the amount of the fluctuation of selection intensity is not very
well known. The most well-studied case is that of the *medionigra* gene in the moth, *Panaxia dominula*. Analyzing nine-generation data, Fisher and Ford (1947) showed that the variance of gene frequency change is much larger than the value expected from random sampling of genes alone. Wright's (1948) estimate of $V_r$ for these data is 0.0483. This value would be exceptionally high, as noted by Wright himself, and generally $V_r$ must be much smaller. However, it should be noted that a small amount of random fluctuation such as $V_r=10^{-8}$ to $10^{-6}$ has a profound effect on genetic variability in large populations, though in large populations $V_r$ probably cannot be very large. For a given pair of alleles, $V_r$ is expected to decrease as $N_e$ increases. This is because in a large population the effects of many environmental fluctuations would be averaged out.

In the present study we have studied the effect of fluctuation of selection intensity as a random noise. As mentioned earlier, fluctuation of selection intensity may produce a stabilizing effect for gene frequency under certain circumstances. Karlin and Levikson (1974) and Takahata et al. (1975) studied the distribution of gene frequency for these cases. However, their formula is for a pair of alleles with forward and backward mutations, so that it cannot be directly compared with our formulae. Takahata et al. also derived a formula for the expected heterozygosity for a special case. Their formula is independent of population size and mutation rate. Here again it cannot be compared with our formula, since it is valid only for a special case.

Another quantity Takahata et al. studied by using (1) and (3) is the rate of gene substitution per generation ($\alpha$). They obtained
\[ \alpha = 2N_eV_r/\log_e(4N_eV_r) , \] (14)
which depends on population size, mutation rate, and the variance of selection intensity. If we use (2) and (3), the equivalent formula becomes
\[ \alpha = v , \] (15)
since the number of mutations occurring in a generation is $2N_e$ and the probability of fixation of a new mutant gene is $1/(2N)$ (Kimura 1962). It is known that the rate of amino acid substitution in proteins in evolution is roughly constant per year (Zuckerkandl and Pauling 1965; Margoliash and Smith 1965). This approximate constancy of the rate of gene substitution is more easily explained by our result than by Takahata et al.'s, since it would be very difficult to keep $\alpha$ in (14) constant for a long period of time. This suggests that the stabilizing effect caused by random fluctuation of selection intensity, if any, is unimportant in molecular evolution.

Another indication of unimportance of the stabilizing effect is obtained by examining the relationship between gene frequency and heterozygosity for protein loci. If we use formulae (1) and (3) instead of (2) and (3), the theoretical relationship corresponding to (7) is given by
\[ h(x) = 2x(1-x)\Phi_1(x) , \] (16)
where
\[ \Phi_i(x) = \frac{4N_i\nu(-Kx^2+Kx+1)}{x(1-x)} \left\{ \frac{1}{2B} \left[ 1 + \frac{1-2x}{-Kx^2+Kx+1} \right] \right\} \]

in which \( B = 1 + \frac{1}{2(C)} \log \frac{(C+K)}{(C-K)} \). The numerical relationship between \( x \) and \( h(x) \) is given in Figure 3. It is clear that if \( N_iV_i \) is large, the relationship between \( x \) and \( h(x) \) is not linear, and the expected number of heterozygotes (or heterozygosity) for gene frequency classes near 0.5 are much higher than those for the case of \( N_iV_i = 0 \). In practice, however, heterozygosity data for enzyme loci do not show such a relationship. The observed relationship between \( x \) and \( h(x) \) is roughly linear (Yamazaki and Maruyama 1972), or if it is not, there is an excess of heterozygosity in low gene frequency classes rather than in intermediate gene frequency classes (Latter 1975). This observation is more consistent with the prediction based on (2) and (3) (see Figure 1), though the excess of heterozygosity in low gene frequency classes can also be explained by the hypothesis of slightly deleterious genes.

![Diagram](image)

Fig. 3. Relationship between the expected number of heterozygotes and gene frequency when \( M_{ss} = V_i\nu(1-x)(1-2x) \) and \( V_{ss} = V_i\nu(x(1-x)^2 + x(1-x))/(2N_i) \) are used. The numbers beside the curves represent the value of \( N_iV_i \). The ordinate is in units of \( 2N_i\nu \).

In the last few years a number of authors (e.g. Gillespie and Langley 1974) suggested that the temporal fluctuation of selection intensity is an important mechanism for maintaining protein polymorphism. This suggestion is based on the stabilizing effect that occurs in a special case of the Wright model of natural selection, as discussed earlier. As we have seen above, however, data on molecular evolution and polymorphism do not support this view. Rather, the opposing view that fluctuating selection is a random noise seems to be more consistent with these data. The frequent observation that the average heterozygosity for protein loci is smaller than the level predicted by the neutral mutation hypothesis may also be an indication of the importance of the latter view. As we have seen in Figure 2, the fluctuating selection as a random noise is very effective in reducing heterozygosity particularly in large popula-
tions. On the other hand, if a stabilizing effect is generated by random fluctuation, it would increase the level of heterozygosity more than that expected under the neutral mutation hypothesis (Figure 3).

Random fluctuation of selection intensity creates some problems in distinguishing between neutral and selective genes in natural populations. If selection intensity fluctuates from generation to generation, single generation data on natural selection are not reliable, and a long-term study is required to determine whether or not a particular allele is selectively advantageous over the other allele or alleles. This would be technically difficult for organisms with a long generation time, such as man. Of course, when $s$ fluctuates at random, $\hat{s}$ would rarely become strictly 0 in nature. If $\hat{s}$ is not 0 for many mutant genes, the genetic variability in a population will further be reduced if the rate of gene substitution remains the same.

**SUMMARY**

The effects of random fluctuation of selection intensity on the expected number of heterozygous codons, average heterozygosity, and the relationship between gene frequency and heterozygosity are studied, assuming that the fluctuation of selection intensity is a random noise. In small populations these effects are relatively small, and the genetic variability in a population is determined mainly by the mutation rate and population size. In large populations, however, even a small amount of random fluctuation reduces the genetic variability to a considerable extent. A formula for the expected number of segregating codons per locus is also worked out. Roughly speaking, the effect of random fluctuation of selection intensity on this parameter is similar to that on the expected number of heterozygous codons per locus. In addition to the above studies, mathematical models of random fluctuation of selection intensity in diffusion approximations are discussed. It is shown that data on molecular evolution and polymorphism are more consistent with the predictions from the hypothesis that temporal fluctuation of selection intensity is a “random noise” than those from the alternative hypothesis that it is a “stabilizing factor for gene frequency.”

**ACKNOWLEDGEMENT**

This work was supported by grants from the US National Science Foundation and the US National Institutes of Health. Part of this work was done while the authors were affiliated with Brown University.

**APPENDIX**

We assume that in each generation a large number of individuals are born and a small proportion of them ($N$ individuals) survive to maturity, the majority of prematurity deaths being accidental or due to genes other than the locus under considera-
tation. We further assume that natural selection due to a particular pair of alleles, $A$ and $A'$, occurs only once in a life cycle through competition between a random pair of individuals. For simplicity, we consider the following haploid model due to Nei (1971).

<table>
<thead>
<tr>
<th>Competition between</th>
<th>Frequency</th>
<th>Probability of survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A': A'$</td>
<td>$x^2$</td>
<td>1</td>
</tr>
<tr>
<td>$A': A$</td>
<td>$2xy$</td>
<td>$(1+s)/2$</td>
</tr>
<tr>
<td>$A: A$</td>
<td>$y^2$</td>
<td>$(1-s)/2$</td>
</tr>
</tbody>
</table>

In the above model $x$ and $y$ denote the frequencies of $A'$ and $A$ at birth, respectively, and $s$ is a constant such that $-1 \leq s \leq 1$. If we assume that the $N$ individuals surviving to maturity are winners in $N$ randomly chosen pairwise competitions at a certain developmental stage, then the probability that they come from $n_1 A': A'$, $n_2 A': A$, and $n_3 A: A$ competitions is

$$P(n_1, n_2, n_3) = \frac{N!}{n_1! n_2! n_3!} (x^2)^{n_1} (2xy)^{n_2} (y^2)^{n_3}.$$  \hfill (A1)

From each of the $n_1 A': A'$ competitions, one $A'$ survives, while from each $A: A$ competition one $A$ survives. On the other hand, from each $A': A$ competition $A'$ survives with probability $(1+s)/2$. Thus, the probability that $r A'$ individuals survive from $n_2 A': A$ competitions is

$$P(n_2; r) = \frac{n_2!}{r!(n_2-r)!} \left[ \frac{1+s}{2} \right]^{n_2-r} \left[ \frac{1-s}{2} \right]^{n_2-r}.$$  \hfill (A2)

Therefore, the mean of the frequency $\langle x' \rangle$ of $A'$ for a fixed value of $s$ in the adult population is

$$E(\langle x' \rangle | s) = \Sigma_N \Sigma_r \frac{n_1+r}{N} P(n_1, n_2, n_3) P(n_2; r)$$

$$= x + sxy,$$

where $\Sigma_N$ refers to the summation over all combinations of $n_1$, $n_2$, and $n_3$ with the restriction $n_1+n_2+n_3=N$ and $\Sigma_r$ the summation over all possible values of $r$. Thus, the mean of gene frequency change $\langle dx = x' - x \rangle$ for a fixed value of $s$ is $E(\langle dx \rangle | s) = sxy$. If $s$ varies at random with mean $\bar{s} = 0$ and variance $V_s$, then the expected change of $x$ per generation is

$$E(\langle dx \rangle) = \bar{s}xy = 0.$$  

On the other hand, the expected variance of $\Delta x = \Delta x_r + \Delta x_s$ (see the text) can be shown to be

$$E(\langle dx \rangle^2) = E(\langle dx_r \rangle^2) + E(\langle dx_s \rangle^2)$$

$$= V_s x^2 y^2 + xy/N + O(E(s^2)/N).$$
The usual procedure to get a diffusion approximation is to measure time in units of $N$ generations, and let $N \to \infty$ and $V_s \equiv E(s^2) \to 0$, such that $NV_s$ stays constant. Then, the mean $(M_{s^2})$ and variance $(V_{s^2})$ of gene frequency change per generation are given by

$$M_{s^2} = 0,$$

$$V_{s^2} = V_s x^2 (1-x)^2 + x(1-x)/N.$$  

(A3)  

(A4)

In diploid organisms, $N$ in (A4) must be replaced by $2N$, since $2N$ rather than $N$ genes are eventually sampled to form the adult population. Therefore, under the above assumptions $M_{s^2}$ and $V_{s^2}$ in diploids are given by (2) and (3) in the text, irrespective of the developmental stage at which natural selection operates.

LITERATURE CITED


GENETIC VARIABILITY