GENETIC VARIABILITY AND RATE OF GENE SUBSTITUTION IN A FINITE POPULATION UNDER MUTATION AND FLUCTUATING SELECTION*

NAOYUKI TAKAHATA

National Institute of Genetics, Mishima, 411 Japan

ABSTRACT

By using a numerical method of solving stochastic difference equations, the level of genetic variability maintained in a finite population and the rate of gene substitution under several models of fluctuating selection intensities were studied. It is shown that mutation and random genetic drift both play an important role in determining genetic variability and the rate of gene substitution. Compared with the case of neutral mutations, the fluctuation of selection intensity caused by temporal and spatial heterogeneity of environments generally increases the rate of gene substitution, but the level of genetic variability may be increased or decreased, depending upon the model and the parameters used. Although such a type of selection *per se* can not be ruled out, when mutation is taken into account, it is difficult to explain both the observed amount of genetic variability and the rough constancy of evolutionary rate within a framework of fluctuating selection models.

IN a previous paper, TAKAHATA and KIMURA (1979) studied the genetic variability in a finite population under mutation and autocorrelated fluctuation of selection intensities. Recently, GILLESPIE (1977; 1979) investigated the same problem, as well as the genetic identity of NEI (1972) based on his mathematical model. His model is general in the sense that the effects of temporal and spatial heterogeneities in environments can both be treated by the same theory. However, the model is developed under the assumption that no mutation occurs, and the populations are so large that we can ignore the effect of random genetic drift. Although such a situation may be theoretically conceivable, it is not clear how realistic the theory is. Lack of quantitative study on this point has brought about a controversy (see NEI and GILLESPIE 1980). Another model relevant to this issue was proposed by NEI and YOKOYAMA (1976). However, these authors made a simplifying assumption about the variance of gene frequency change, so that it is not clear whether their formula gives a correct prediction.

Many fluctuating selection models, including those mentioned above, are formulated by using diffusion approximations, and it is possible to represent all of them in a single formulation. I shall not repeat the biological bases for those models, for which readers may refer to each paper cited above. In this paper, I shall investigate the effects of fluctuating selection on the level of genetic variability and the rate of gene substitution under the influence of mutation and ran-

* Contribution No. 1345 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411 Japan.

dom genetic drift. Since it is difficult to obtain an analytical solution for this case, I shall use a computer-simulation method. This method makes use of appropriate stochastic difference equations corresponding to the diffusion process. It has been applied to other problems in population genetics by MARUYAMA (1980) and MARUYAMA and NEI (1981). The results obtained will be discussed in relation to observed data on genetic variability and gene substitution.

MODEL AND ANALYSIS

Consider a panmictic population of diploid organisms with an effective size N_e. We assume that a gene consists of so many nucelotide sites that we can always regard a newly arisen mutant as a new, not pre-existing allele (KIMURA and CROW 1964). Let v be the mutation rate per gene per generation. Suppose that there are n segregating alleles at a locus in the population, and let $\phi(t; x_1, x_2, \ldots, x_{n-1}; y_1, y_2, \ldots, y_{n-1})$ be the transition probability density that the frequencies of alleles $A_1, A_2, \ldots, A_{n-1}$ change from $x_1, x_2, \ldots, x_{n-1}$ to $y_1, y_2, \ldots, y_{n-1}$ in time interval t. Then, the density ϕ satisfies the Kolmogorov backward equation

$$\frac{\partial \phi}{\partial t} = L_1 \phi,$$
$$L_1 = -\sum_{i=1}^{n-1} v x_i \frac{\partial}{\partial x_i} + \frac{1}{4N_e} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} x_i (\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} \tag{1}$$

for the case of neutral mutations (CROW and KIMURA 1956), where $\delta_{ii} = 1$ and $\delta_{ij} = 0$ if $i \neq j$. The time denoted by t is measured in generations.

In addition, we assume that natural selection acts on different genotypes in such a way that their selection intensities fluctuate randomly owing to environmental heterogeneities. A general diffusion equation for symmetric fluctuating selection models is then given by

$$\frac{\partial \phi}{\partial t} = L_2 \phi,$$
$$L_2 = \sigma^2 \beta \sum_{i=1}^{n-1} x_i (F - x_i) \frac{\partial}{\partial x_i} + \frac{\sigma^2}{2} \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} x_i x_j (\delta_{ij} + F - x_i - x_j) \frac{\partial^2}{\partial x_i \partial x_j} \tag{2}$$

where $F = \sum_{i=1}^{n} x_i^2$ and σ^2 is the strength of fluctuation of selection intensities (for example, see TAKAHATA, ISHII and MATSUDA 1975; FELSENSTEIN 1976; GILLESPIE 1977, 1979; TAKAHATA and KIMURA 1979 and references therein). In the above equation, β is a constant related to the so-called stabilizing effect (NEI and YOKOYAMA 1976) and it takes on nonnegative values depending on mathematical models. For instance, an appropriate extension of NEI and YOKOYAMA's model (1976) corresponds to the case of $\beta = 0$. Also, the cases of $\beta = 1$ and $\beta > 1$ have been studied independently by TAKAHATA and KIMURA (1979) and GILLESPIE (1977). Recently, GILLESPIE (1980b) extended his symmetric model to an asymmetric one. When the asymmetry is incorporated into the theory through the mean difference in fitnesses among different genotypes, the parts responsible for

fluctuation of selection intensities remain unchanged. It is not surprising that this asymmetry can efficiently reduce the level of genetic variability, but here I shall not consider the effect of mean difference in fitnesses, since the main purpose of this paper is to explore the relative roles of mutation, random genetic drift and fluctuating selection. The combined model of fluctuating selection with the effects of mutation and random genetic drift can be represented by the following Kolmogorov backward equation

$$\frac{\partial \phi}{\partial t} = (L_1 + L_2)\phi \tag{3}$$

where the operators L_1 and L_2 are given in (1) and (2), respectively. For convenience, we change the time scale from one generation to $T = \sigma^2 t$, dividing both sides of (3) by σ^2. Of course, it is possible to measure time in units of $2N_e$ generations, as usually done. Actually, I used this unit of time in the case of neutral mutation. However, in the present formulation, time was measured in units of $1/\sigma^2$ generations because I would like to treat the case of an infinite population, i.e., $N_e = \infty$. At any rate, equation (3) then becomes

$$\frac{\partial \phi}{\partial T} = \frac{1}{\sigma^2}(L_1 + L_2)\phi. \tag{4}$$

Note here that we can derive an interesting property of the quilibrium distribution that satisfies $(\tilde{L}_1 + \tilde{L}_2) = 0$, where $\tilde{L}_i$ ($i = 1$ and 2) denotes the adjoint operator of L_i. GILLESPIE (1977, 1979) has shown that the solution satisfying $\tilde{L}_2\phi = 0$ is given by the limit of a Dirichlet distribution with a parameter β (for $\beta > 1$), and it is irrelevant to σ^2. On the other hand, the solution of $\tilde{L}_1\phi = 0$ for the neutral model is expressed by a similar function with a single parameter of $4N_e v$ (see WRIGHT 1949; KIMURA and CROW 1964). Comparing both solutions obtained for the above cases separately, it is shown that the parameter β plays a role analogous to $2N_e v + 1$ in the neutral model, as pointed out by GILLESPIE (1977). Therefore, if the relationship of $\beta = 2N_e v + 1$ happens to hold, the solution of $\tilde{L}_2\phi$ must satisfy $\tilde{L}_1\phi = 0$. This indicates that the distribution with a single common parameter β (or $2N_e v + 1$) is the solution of $(\tilde{L}_1 + \tilde{L}_2)\phi = 0$. These special cases will be used for a check of the present method described below.

Unfortunately, however, it is not easy to solve (4) for other cases analytically, so that we simulate the process represented by stochastic difference equations relevant to (4). The simulation method in the infinite allele model has been recently applied to other problems in population genetics by MARUYAMA (1980), MARUYAMA and NEI (1981) and MARUYAMA, TAKAHATA and KIMURA (unpublished). If we use the theory of ITO (1944), we get the appropriate stochastic difference equations

$$\begin{aligned} \Delta x_i(T) &\equiv x_i(T + \Delta T) - x_i(T) \\ &= -v' x_i \Delta T + \beta x_i (F - x_i)\Delta T + \sum_{j=1}^{n-1} e_{ij} B_j(\Delta T) \end{aligned} \tag{5}$$

for $i = 1, 2, \ldots, n-1$, where $v' = v/\sigma^2$, $B_j(\Delta T)$ is an independent Brownian motion with mean 0 and variance ΔT, and e_{ij} is the (i, j)th element of a positive definite square root of the covariance matrix with the element of $V_{ij} = x_i x_j (\delta_{ij} + F - x_i - x_j) + \frac{1}{2N_e\sigma^2} x_i(\delta_{ij} - x_j)$. There are many other ways of representing diffusion processes (for example, see ITOH 1979; KIMURA 1980), but equations (5) seem to be not only general, but also the only way of representing the present case involving fluctuating selection.

In the actual calculation of (5), however, I modified the process by taking into account changes due to mutation. The modification seems unavoidable in order to make the process proceed in a reasonabl period of computation time. More importantly, several numerical studies suggest that we cannot simulate (5) accurately if we take the time interval ΔT larger than the initial frequnecy of mutants newly introduced into a population during ΔT, (although this finding is not rigorously justified).

Now, let A_i be one of the segregating alleles at time T, and $x_i(T)$ be the frequency in a population. To study the rate of gene substitution, we associate an integer $N(A_i)$ with each allele, which is referred to as the number of mutational events leading to A_i from the ancestral allele, say A_0, at $T = 0$. For convenience, we assume that $N(A_0) = 0$. The number of mutational events for each allele at any time T can be easily obtained by counting all mutations involved in the course of evolution of A_i from $T = 0$. For example, suppose that A_i mutates to A_{i1} in the time interval ΔT in which A_{i1} represents a new, not pre-existing allele under the assumption of the infinite allele model. Then, the number of mutational events associated with A_{i1} is increased by one from that of A_i, *i.e.*, $N(A_{i1}) = N(A_i) + 1$. If A_i still exists in a subsequent time and again mutates to another allele A_{i2} $(\neq A_{i1})$, $N(A_{i2})$ is equal to $N(A_i) + 1$, as before. But if A_{i1} mutates to A_{i3}, the number becomes $N(A_{i1}) + 1 = N(A_i) + 2$. Thus, the difference between $N(A_{i3})$ and $N(A_i)$ corresponds to the number of mutational events in a certain time interval during which A_{i3} has been introduced into a population through the lineage of $A_i \rightarrow A_{i1} \rightarrow A_{i3}$. If we continue this procedure from an initial population, we can unambiguously associate the number to all segregating alleles at any time. Note here that the same number of mutational events does not necessarily mean the same allele as shown in the above example.

Instead of (5), we stochastically introduce mutants every ΔT to avoid the necessity of taking ΔT as an indefinitely small value when v' becomes small. Let ε be an arbitrary but small positive number. Then, we choose ΔT as $v'\,\Delta T/\varepsilon < 1$ for each existing allele A_i at time T, whether or not a mutation occurs until $T + \Delta T$ is decided with probability $v'\,\Delta T/\varepsilon$. If a pseudo-uniform random number generated is less than this value, we mutate A_i to a new, not pre-existing allele A_j with the initial frequency $\varepsilon x_i(T)$. At the same time, the frequency of A_i is reduced by the same amount, and the number of mutational events for A_j is associated according to the above rule. This process is repeated for all n segregating alleles at time T before we compute changes due to fluctuating selection and random genetic drift. The formal representation may be written

$$x_i(T+\Delta T) = (1-\varepsilon)x_i(T),$$
$$x_j(T+\Delta T) = \varepsilon x_i(T) \qquad (6)$$

and

$$N(A_j) = N(A_i) + 1 \ ,$$

with probability $v' \Delta T/\varepsilon$. When more than one mutations occur in the above n repeats, each A_j should be chosen as new by incorporating the contribution coming from the mutant alleles that have already been generated in the present time interval.

After mutations, using the frequencies of (6) and ignoring the alleles with frequencies less than a certain small value (I tentatively set it at 10^{-6}), I computed the changes of frequencies caused by fluctuating selection and random genetic drift according to the method mentioned above. Namely,

$$\Delta x_i(T) = \beta x_i(F-x_i)\Delta T + \sum_{j=1}^{l-1} e_{ij}B_j(\Delta T) \qquad (7)$$

for $i = 1, 2, \ldots, l-1$ and $x_l = 1 - \sum_{i=1}^{l-1} x_i$, where l denotes the number of all existing alleles, including new mutants introduced in the time interval ΔT. As noted by MARUYAMA and NEI (1981), it is important to take ΔT as small as possible. I used $\Delta T = 0.001$ or less and $\varepsilon = 0.005$. These parameters, however, are somewhat arbitrary; therefore, the accuracy of the present method based on (6) and (7) should be checked from various aspects. This will be discussed later.

Using the number of mutational events and gene frequencies, we can define the mean number of gene substitutions $K(T)$ until time T as

$$K(T) = \Sigma N(A_i)x_i(T) \ , \qquad (8)$$

where the sum is taken for all existing alleles. The mean rate is given by dividing $K(T)$ by sufficiently large T. The value of $K(T)/T$ is used as a measure of the rate of gene substitution. Fixation of a particular allele must occur with a very low probability in a polymorphic population where new mutations are continuously introduced. Therefore, we cannot apply the theory of fixation probability to get the rate of gene substitution. This is particularly true when selection is taken into account. MARUYAMA and NEI (1981) considered the rate of gene substitution in a slightly different way, whereby they recorded the full lineages of any allele from the start of simulations. They counted the number of "fixation of mutational codons," which means complete loss of pre-existing alleles from a population. This number is equal to $K(T)$ if the population eventually becomes monomorphic at a specified time T. Since $K(T)$ in (8) has the contribution coming from polymorphic alleles, it is generally greater than that of MARUYAMA and NEI's, but the discrepancy between the two becomes negligible after a sufficiently long time has elapsed.

To obtain the mean rate of gene substitution per generation in units of mutation rate v, we divide $K(T)/T$ by v'. Then, we have

$$k/v = \frac{K(T)}{Tv'} = \frac{K(t)}{tv} \ . \qquad (9)$$

Based on (6) and (7), we can study various properties of both equilibrium and nonequilibrium populations. Starting from a monomorphic population, I repeatedly calculated gene frequencies up to about $T = 1100$, which corresponds to $10^3/\sigma^2$ generations. For example, in the case of $\sigma^2 = 10^{-4}$, the simulation was continued for 10^7 generations. To establish the equilibrium population, the first $10^2/\sigma^2$ generations were discarded; thereafter, every $0.1/\sigma^2$th generation gene frequencies were recorded, the total number of observations being 10^4. This scale of simulation seems to be sufficient for studying various properties of equilibrium populations. However, it does not give a very accurate estimate of the rate of gene substitution. As will be discussed later, the rate has a large sampling error. Because of the limitation of computer time, however, I will present here rough estimates of the rate of gene substitution to show the rapid turnover of alleles in populations resulting from fluctuation of selection intensity. Also, the equilibrium distribution of rare alleles with frequencies less than ε is probably unreliable, irrespective of the scale of simulation. The accuracy depends heavily on the values of ΔT and ε.

RESULTS AND DISCUSSION

Before discussing the full details of my results, we first examine the validity of the present method of stochastic difference equations, because equations (6) and (7) with $\Delta T = 0.001$ or less may not necessarily be warranted in regard to our requirements. This can be done in three different ways. One is to study the neutral case where we have rigorous solutions for many quantities. The results are given in Table 1, in which the mean $\bar{F}$ and variance V_F of homozygosity (KIMURA and CROW 1964; STEWART 1976), the mean number n_a of different alleles observed in

TABLE 1

Comparison of the results of Monte Carlo experiments (obs.) with those (exp.) expected in the neutral model

	$2N_e v$	0.001	0.01	0.1	0.5	1.0
$\bar{F}$	exp.	0.9980	0.9804	0.8333	0.5000	0.3333
	obs.	0.9979	0.9816	0.8288	0.5005	0.3451
V_F	exp.	0.0007	0.0063	0.0395	0.0417	0.0222
	obs.	0.0008	0.0060	0.0406	0.0453	0.0253
n_a	exp.	1.01	1.12	2.12	5.88	9.77
	obs.	1.01	1.11	2.27	5.70	9.20
k/v	exp.	1.00	1.00	1.00	1.00	1.00
	obs.	1.04	0.89	1.18	0.86	1.17

$\bar{F}$ and V_F are the mean and variance of homozygosity, n_a stands for the number of different alleles in a sample of $2n$ genes ($n = 10^2$) and k/v for the rate of gene substitution relative to the mutation rate. The slight deviation of k from v is due to a sampling error at observed time $T = 10^3$.

a sample of $2n$ genes (EWENS 1972) and the rate of gene substitution k in units of mutation rate v are presented. It is seen from the table that the simulation results provide a quite satisfactory approximation in this case. In particular, note that the rate of gene substitution defined by (9) is roughly equal to unity, which is expected from the theory using the fixation probability (KIMURA 1968). The small discrepancy must come from sampling error. If we observe the rate at much later simulation stages or take the average for many independent repeats of simulation, it is expected that k is very close to v.

Another check of the method was made for some special cases of the parameters involved, *i.e.*, $\beta = 2N_e + 1$ holds. In these cases, we fortunately know the rigorous solution of the equilibrium distribution of gene frequency, in addition to the mean and variance of homozygosity. The solution of the equilibrium distribution of (4) is then given by the limit of a Dirichet distribution with a single common parameter of either β or $2N_ev$, as mentioned earlier. Therefore, the frequency spectrum is equal to $\theta x^{-1}(1 - \mathrm{x})^{\theta-1} = (2\beta - 2)x^{-1}(1 - x)^{2\beta-3}$, where $\theta = 4N_ev$, the mean homozygosity to $\bar{F} = \dfrac{1}{1+\theta} = \dfrac{1}{2\beta = 1}$, and so forth. For instance, if $2N_ev = 0.1$ and $\beta = 1.1$, or $2N_ev = 0.5$ and $\beta = 1.5$, the value of $\bar{F}$ should be equal to 0.8333 or 0.5, (see Figure 1 for the frequency spectrum). These special cases are marked by the symbol † in Table 2. We can see from Table 2 that the present stochastic difference equation method gives good agreement with the theoretical prediction. However, it is important that we can no longer expect that the rate of gene substitution k equals the mutation rate v, even in such cases.

The other check is possible by comparing the results for the case of $\beta = 1$ with those given by TAKAHATA and KIMURA (1979), where the underlying model population was simulated under the effect of autocorrelated selection. The agreement between the two is again satisfactory (see Figure 2 in their paper). However, as noted by MARUYAMA and NEI (1981), the results on rare alleles thus obtained may not be accurate enough in some cases. The values of the frequency spectrum in a low frequency class and the number of different alleles in a sample n_a tend to be smaller than the expected values. This is mainly due to the values of ΔT and ε used in the simulation.

Now let us examine our results with special reference to protein polymorphism and molecular evolution. First of all, it is clear that the amount of genetic variability depends heavily not only on the mutation rate relative to σ^2, but also on the value of β, particularly when the effect of random sampling of gametes is small (Table 2). For example, when $2N_e\sigma^2$ is greater than 10^3, $\beta = 1.1$ and $v/\sigma^2 = 0.001$, the mean homozygosity $\bar{F}$ considerably decreases to about one-half, compared with the expected value of 0.8333 in the case of no mutation in an infinite population, while it increases from 0.3333 in the case of $2N_ev = 1$ for neutral mutations. This increase or decrease of genetic variability compared with the above two extreme cases must result from the so-called stabilizing effect in the drift term, as well as the diffusion caused by fluctuation of selection intensity in equation (3). However, the stabilizing effect plays a more important role than the diffusion effect in determining the level of genetic variability. Actually,

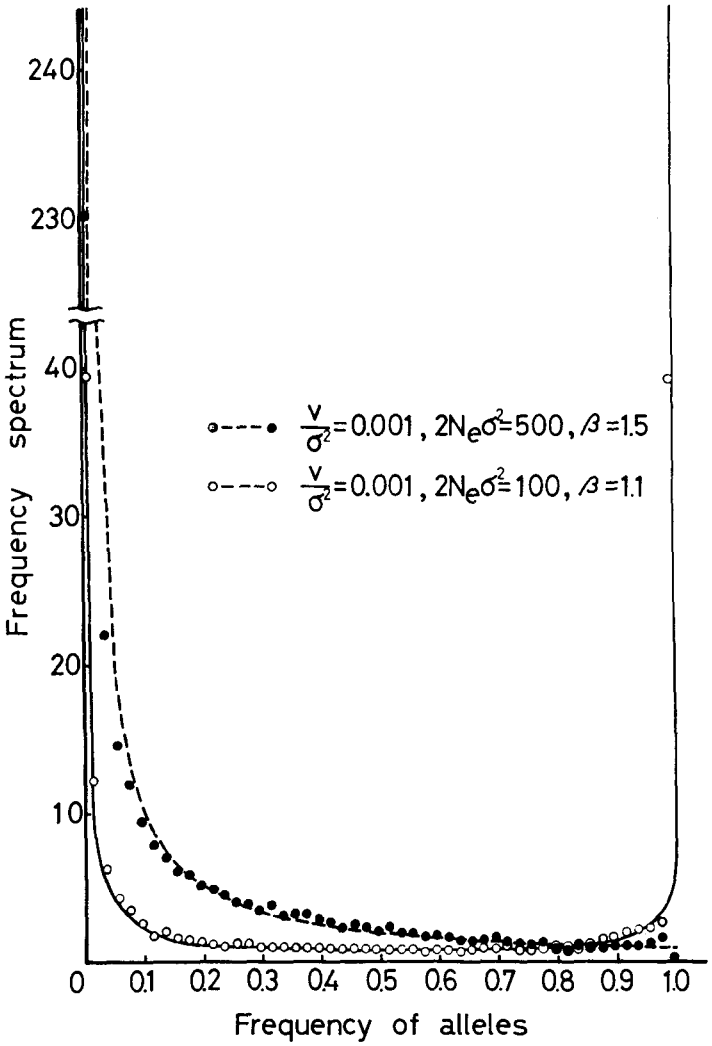

FIGURE 1.—Frequency spectrum. The solid and broken curves are drawn according to the equation $\theta x^{-1}(1-x)^{\theta-1}$, where $\theta = 4N_e v = 0.2$ and 1.0. The results of simulations for the cases of $v/\sigma^2 = 0.001$, $2N_e\sigma^2 = 100$ and $\beta = 1.1$, and of $v/\sigma^2 = 0.001$, $2N_e\sigma^2 = 500$ and $\beta = 1.5$ are, respectively, marked by open and solid circles.

when there is no other effect but diffusion, the amount of genetic variability is kept very low compared with the neutral model (see the results for $\beta = 0$ in Table 2). In the opposite case of $\sigma^2 = 0$ and $\beta \neq 0$, which is equivalent to a symmetrical overdominant model, the effect of β on genetic variability is quite pronounced (MARUYAMA and NEI 1981). Note here that in the range of parameters used in Table 2, the two-allele approximation of NEI and YOKOYAMA (1976) to the variance of gene frequency change seems to be fairly good. However, under more polymorphic situations, it somewhat underestimates the amount of genetic

variability. For example, in the case of $v/\sigma^2 = 0.2$ and $2N_e\sigma^2 = 10$, $\bar{F}$ was about 0.37 ($k = 0.8v$), contrasted with 0.47 in Table 1 in their paper.

The larger the value of β, the more sensitive is the mean homozygosity to changes in the ratio of v/σ^2 and the value of $2N_e\sigma^2$. It is obvious that if we keep the ratio of v/σ^2 constant, such as greater than 10^{-4}, we cannot ignore the mutational effect even in a large population. Unless we assume that v/σ^2 approaches zero as $2N_e\sigma^2$ becomes large, the combined model of (3) does not approach the simple fluctuating selection model of (2) (see Figure 2). Although such a limit is, of course, mathematically justified and conceivable, the biological meaning is obscure since neither v tends to be zero, nor can σ^2 be greater than unity. In this regard, GILLESPIE's discussion on the effect of mutation in his model (1979, p. 752) is not very clear.

Additionally, the effect of random genetic drift still plays a significant role in determining the level of genetic variability even when $2N_e\sigma^2$ exceeds 10^3. Therefore, if σ^2 is 10^3 times greater than v, the effect can be neglected only for the cases of $2N_e\sigma^2 >> 10^3$, although the critical value of $2N_e\sigma^2$ is dependent on β. Thus, it is important to realize that the level of genetic variability is strongly influenced by values of $2N_e\sigma^2$ and v/σ^2 (TAKAHATA and KIMURA 1979; NEI 1980) and that this tendency is more pronounced as the value of β increases. The heuristic arguments of GILLESPIE (1980a) in reply to NEI's comment that genetic variability is much dependent on the value of $2N_ev$ only when $\beta = 1$ (namely, the case of autocorrelated fluctuation of selection intensities) are not supported. On the contrary,

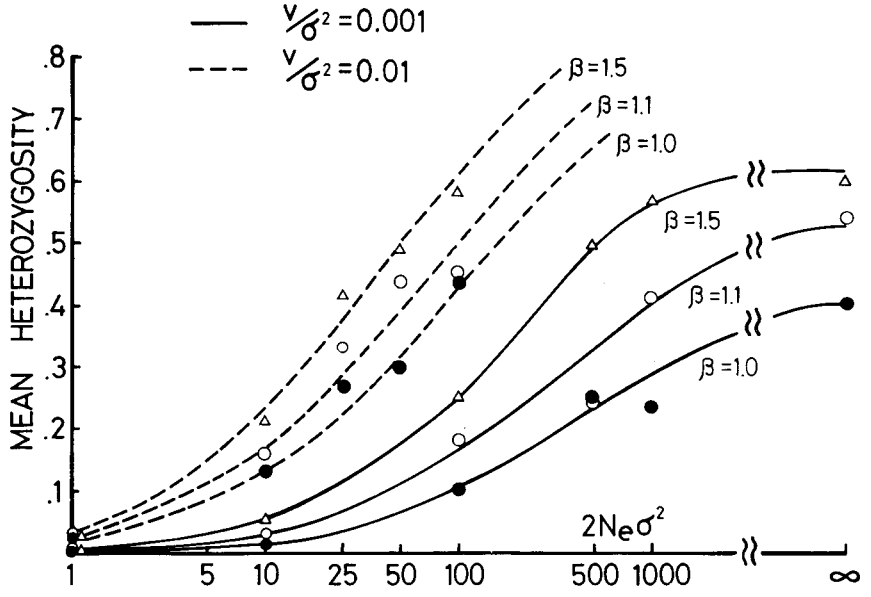

FIGURE 2.—Mean heterozygosity as a function of $2N_e\sigma^2$ when both values of v/σ^2 and β are specified. The solid circles (for $\beta = 1.0$), open circles (for $\beta = 1.1$) and open triangles (for $\beta = 1.5$) represent the results of simulations when $v/\sigma^2 = 0.01$ and 0.001. The solid ($v/\sigma^2 = 0.001$) and broken ($v/\sigma^2 = 0.01$) curves are solely inferred from Monte Carlo experiments.

TABLE 2

Mean and variance of homozygosity, mean number of different alleles in a sample of 200 genes and the rate of gene substitution observed in Monte Carlo experiments

$2N_e\sigma^2$ ($2N_e\nu$)		Neutral (expected)	$\beta=0$	$\beta=1.0$	$\beta=1.1$	$\beta=1.5$
		$\nu/\sigma^2 = 10^{-4}$				
	$\overline{F}$	0.9804	0.9970	0.9805	0.9635	0.9436
100	V_F	0.0063	0.0010	0.0061	0.0117	0.0178
(0.01)	n_a	1.12	1.03	1.13	1.18	1.24
	k/ν	1.00	——*	29.4	63.6	19.6
	$\overline{F}$	0.8333	0.9988	0.9506	0.8338†	0.6161
1000	V_F	0.0395	0.0001	0.0143	0.0402	0.0396
(0.1)	n_a	2.12	1.08	0.54	2.38	3.83
	k/ν	1.00	——*	41.8	78.6	88.0
	$\overline{F}$	0.5000	0.9990	0.7913	0.7750	0.4880†
5000	V_F	0.0417	0.0000	0.0512	0.0445	0.0251
(0.5)	n_a	5.88	1.06	3.75	4.02	6.48
	k/ν	1.00	——*	145.	145.	117.
	$\overline{F}$	0	0.9982	0.7886	0.6883	0.4934
∞	V_F	0	0.0001	0.0429	0.0474	0.0377
	n_a	∞	2.15	3.83	5.88	8.12
	k/ν	1.00	——*	111.	118.	109.
		$\nu/\sigma^2 = 10^{-3}$				
	$\overline{F}$	0.9980	0.9979	0.9986	0.9975	0.9980
	k/ν	1.00	0.97	0.97	2.91	——*
1	V_F	0.0007	0.0007	0.0004	0.0009	0.0007
(0.001)	n_a	1.01	1.01	1.01	1.01	1.01
	$\overline{F}$	0.9804	0.9943	0.9886	0.9707	0.9484
10	V_F	0.0063	0.0016	0.0035	0.0096	0.0179
(0.01)	n_a	1.12	1.07	1.09	1.17	1.22
	k/ν	1.00	1.94	3.65	3.88	8.05
	$\overline{F}$	0.8333	0.9916	0.8961	0.8180†	0.7520
100	V_F	0.0395	0.0012	0.0267	0.0406	0.0499
(0.1)	n_a	2.12	1.34	1.86	2.30	2.60
	k/ν	1.00	1.94	12.6	15.6	14.5
	$\overline{F}$	0.5000	0.9873	0.7521	0.7581	0.5936†
500	V_F	0.0417	0.0028	0.0513	0.0553	0.0411
(0.5)	n_a	5.88	1.63	4.13	3.99	6.50
	k/ν	1.00	2.40	17.9	14.6	30.8
	$\overline{F}$	0.3333	0.9952	0.7688	0.5879	0.4321
1000	V_F	0.0222	0.0002	0.0443	0.0504	0.0289
(1.0)	n_a	9.77	1.68	5.34	7.58	10.6
	k/ν	1.00	——*	23.3	21.4	34.1

TABLE 2—Continued

$2N_e\sigma^2$ ($2N_e v$)		Neutral (expected)	$\beta=0$	$\beta=1.0$	$\beta=1.1$	$\beta=1.5$
	$\bar{F}$	0	0.9870	0.5969	0.4593	0.4077
∞	V_F	0	0.0008	0.0557	0.0499	0.0341
	n_a	∞	5.31	17.0	18.7	18.0
	k/v	1.00	1.69	17.5	20.6	23.0
		$v/\sigma^2 = 10^{-2}$				
	$\bar{F}$	0.9804	0.9854	0.9831	0.9698	0.9820
1	V_F	0.0063	0.0042	0.0054	0.0102	0.0056
(0.01)	n_a	1.12	1.12	1.12	1.16	1.13
	k/v	1.00	0.97	1.07	1.94	1.55
	$\bar{F}$	0.8333	0.9454	0.8683	0.8437†	0.7877
10	V_F	0.0395	0.0131	0.0326	0.0369	0.0469
(0.1)	n_a	2.12	1.76	2.16	2.26	2.51
	k/v	1.00	0.78	2.23	2.62	4.27
	$\bar{F}$	0.5000	0.9569	0.6960	0.5646	0.5110†
50	V_F	0.0417	0.0071	0.0528	0.0500	0.0393
(0.5)	n_a	5.88	3.77	6.50	7.30	7.13
	k/v	1.00	0.76	3.79	4.81	4.69
	$\bar{F}$	0.3333	0.9042	0.5626	0.5488	0.4153
100	V_F	0.0222	0.0208	0.0507	0.0696	0.0309
(1.0)	n_a	9.77	4.87	9.41	10.1	11.7
	k/v	1.00	0.96	4.30	4.89	5.74

The columns indicated by "neutral" show the expected values when σ^2 tends to zero, keeping $2N_e v$ constant.

† Denotes the cases where $\beta = 2N_e v + 1$ holds, and the values of F, V_F and n_a should be equal to those in the neutral case in the same row.

* Indicates the cases in which no gene substitution occurred in a simulation (for details see text).

it is much more sensitive to the increased value of $2N_e v$ in the cases of $\beta > 1$ than when $\beta = 1$. General features of fluctuating selection models treated here indicate that if β is greater than $2N_e v + 1$, the level of polymorphism is larger than that expected in the neutral model having the same value of $2N_e v$; in the opposite case of $\beta < 2N_e v + 1$, it is reduced from the neutral level depending upon the value of $2N_e\sigma^2$. Furthermore, if $\beta = 2N_e v + 1$, the level does not differ from that predicted in the neutral model, irrespective of the value of either v/σ^2 or $2N_e\sigma^2$.

The variance of homozygosity (or heterozygosity) can take a large value of fluctuating selection models. In some cases where fluctuation of selection intensity predominates (*e.g.*, $v/\sigma^2 = 0.01$ and $2N_e\sigma^2 = 100$ in Table 2 and also Table 1 of TAKAHATA and KIMURA 1979), the value exceeds 0.06, which is never attained in the neutral case. This trend is caused by the rapid changes of alleles with intermediate frequencies on which selection can act efficiently. As compared with the neutral alleles, the rapid turnover of alleles followed by fluctuating selection can

create a large variance of homozygosity. This circumstance is characteristic of the models of fluctuating selection.

The rate of gene substitution relative to the mutation rate is also represented in Table 2, measured in units of the mutation rate. Except for the cases of neutral mutations and $\beta = 0$, the rate is accelerated as $2N_e\sigma^2$ increases. However, the value of k is not so large compared with value expected from the theory of fixation probability under the assumption of two segregating alleles (see for example, TAKAHATA, ISHII and MATSUDA 1975 for $\beta = 1$). This reduction has already been discussed by TAKAHATA and KIMURA (1979), who noted the intereference of alleles having higher fitnesses with the fixation of other alleles. More importantly, however, when $2N_e\sigma^2$ is changed, we can no longer expect the rate to be constant. This acceleration of evolutionary rate is also explained by the stabilizing effect in the drift term. New mutants with low frequencies tend to be incorporated as common alleles into a population with the help of this factor. Indeed, when there is no such factor, *i.e.*, $\beta = 0$, the observed rates are very close to the mutation rate. In Table 2, I did not present the values of variance of gene substitutional rate, since I performed only a few independent simulations for each case where β, v/σ^2 and $2N_e\sigma^2$ were specified. Therefore, I could not estimate the variance of the rate at time $T = 10^3$, but our studies suggest that this variance is great and that the standard deviation may be as large as 50% of the mean rate. If this is the the case, the mean rates obtained here give only rough features of gene substitution. Nevertheless, it is unlikely that the rough constancy of the evolutionary rate holds in a wide range of $2N_e\sigma^2$ values under fluctuating selection models.

The rate of gene substitution used here is similar to that of ISHII, MATSUDA and OGITA (1978), and may be useful when the ultimate fixation of alleles is rarely observed. However, their mathematical model so far studied is developed under the assumption that selection acts not on different genotypes, but on different states of $N(A_i)$ in the present notation. The biological meaning of the model is therefore quite obscure. We can also make use of NEI's (1972) genetic distance as an appropriate measure of evolutionary distance. However, I did not use it in the present study because I had to perform simulations for a long time to obtain good information on equilibrium. Actually, except for a few cases marked by the symbol * in Table 2, I observed no common allele between ancestral and descendent populations even after several million generations elapsed.

From the above studies, we can make the following conclusions. The stabilizing factor β caused by fluctuation of selection intensity increases the amount of genetic variability within a population and accelerates the rate of gene substitution. However, a large value of β may not necessarily give a higher rate of gene substitution than a small value of β when $2N_e v$ remains the same. This situation has been noted in the overdominant model (NEI and ROYCHOUDHURY 1973; MARUYAMA and NEI 1981). On the other hand, diffusion due to fluctuation of selection intensity plays a role in decreasing genetic variability without change of evolutionary rate. These opposite forces can balance each other and establish an equilibrium in the case of $\beta > 1$, as studied by GILLESPIE (1977). The effect of mutation, however, is very important in determining the amount of genetic

variability and the rate of gene substitution. Now that we have several estimates of mutation rate, and knowing that gene is mutable, neglect of mutational effects leads to a serious misunderstanding of the evolutionary process in natural populations. In addition, random genetic drift has the effect of decreasing the amount of genetic variability up to at least $2N_e\sigma^2 = 10^3$.

GILLESPIE (1979) estimated $\beta = 1.1$ and $\sigma^2 = 10^{-3} \sim 10^{-4}$ as typical values in comparing his theoretical results with observed data on protein polymorphism and molecular evolution. In the case of $\beta = 1.0$, MATSUDA, TAKAHATA and GOJOBORI tentatively concluded that the value of σ^2 is about 100 times greater than the mutation rate, (see also NEI and YOKOYAMA 1976). On the other hand, spontaneous band-morph mutation rate has been estimated as 10^{-6} to 10^{-5} by MUKAI and COCKERHAM (1977), NEI (1977), VOELKER, SCHAFFER and MUKAI (1980) and others. If we accept these estimates, the range of v/σ^2 may be from 10^{-1} to 10^{-4}, and the present simulations covered most of this range. If we assume that $\beta = 1.0 \sim 1.1$, $v = 10^{-6}$ and $\sigma^2 = 10^{-3}$, the effective size seems to be 10^5 at most in view of the observed level of average heterozygosity ($0 \sim 0.3$; FUERST, CHAKRABORTY and NEI 1977). I therefore conclude that mutation and random genetic drift cannot be neglected in the study of protein polymorphism and molecular evolution, even within a framework of fluctuating selection models.

I would like to express my thanks to T. OHTA, T. MARUYAMA and M. NEI for their interest and many useful suggestions.

LITERATURE CITED

CROW, J. F. and M. KIMURA, 1956 Some genetic problems in natural populations. Proc. 3rd Berkeley Symp. Math. Stat. Prob. **4**: 1–22.

EWENS, W. J., 1972 The sampling theory of selectively neutral alleles. Theoret. Pop. Biol. **3**: 87–112.

FELSENSTEIN, J., 1976 The theoretical population genetics of variable selection and migration. Ann. Rev. Genetics **10**: 253–280.

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., 1977 Sampling theory for alleles in a random environment. Nature **226**: 443–445. ——, 1979 Molecular evolution and polymorphism in a random environment. Genetics **93**: 737–754. ——, 1980a Protein polymorphism and the SAS-CFF model. Genetics **94**: 1089–1090. ——, 1980b The stationary distribution of an asymmetrical model of selection in a random environment. Theoret. Pop. Biol. **17**: 129–140.

ISHII, K., H. MATSUDA and N. OGITA, 1978 Molecular evolution by stochastic selection. pp. 24–33. In: *Proc. Intern. Symp. on Mathematical Topics in Biology*. Edited by E. TERAMOTO. Kyoto Univ. Press, Kyoto, Japan.

ITO, K., 1944 Stochastic integral. Proc. Imperial Acad. **20**: 519–524.

ITOH, Y., 1979 Random collision process of oriented graph. Institute of Statistical Mathematics (Japan), Research Memorandum. No. **154**: 1–20.

KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature **217**: 624–626. ——, 1980 Average time until fixation of a mutant allele in a finite population under continued mutation pressure: Studies by analytical, numerical and pseudo-sampling methods. Proc. Natl. Acad. Sci. U.S. **77**: 522–526.

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

MARUYAMA, T., 1980 On an overdominant model of population genetics. Adv. Appl. Prob. **12**: 274–275.

MARUYAMA, T. and M. NEI, 1981 Genetic variability maintained by mutation and overdominant selection in finite populations. Genetics **98**: 441–459.

MATSUDA, H., N. TAKAHATA and T. GOJOBORI, 1977 A diagnosis of the mechanism of protein polymorphism. pp. 168–188. In: *Proc. 2nd Taniguchi Inter. Symp. on Biophysics.* Edited by M. KIMURA. National Institute of Genetics Press, Mishima, Japan.

MUKAI, T. and C. C. COCKERHAM, 1977 Spontaneous mutation rates at enzyme loci in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S. **74**: 2514–2517.

NEI, M., 1972 Genetic distance between populations. Am. Naturalist **106**: 283–292. ——, 1977 Estimation of mutation rate from rare protein variants. Am. J. Hum. Genet. **29**: 225–232. ——, 1980 Protein polymorphism and the SAS-DFF model. Genetics **94**: 1085–1087.

NEI, M. and A. K. ROYCHOUDHURY, 1973 Probability of fixation and mean fixation time of an overdominant mutation. Genetics **74**: 371–380.

NEI, M. and S. YOKOYAMA, 1976 Effects of random fluctuation of selection intensity of genetic variability in a finite population. Japan J. Genetics **51**: 355–369.

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

TAKAHATA, N., K. ISHII and H. MATSUDA, 1975 Effect of temporal fluctuation of selection coefficient on gene frequency in a population. Proc. Natl. Acad. Sci. U.S. **72**: 4541–4545.

TAKAHATA, N. and M. KIMURA, 1979 Genetic variability maintained in a finite population under mutation and autocorrelated random fluctuation of selection intensity. Proc. Natl. Acad. Sci. U.S. **76**: 5813–5817.

VOELKER, R. A., H. E. SCHAFFER and T. MUKAI, 1980 Spontaneous allozyme mutations in *Drosophila melanogaster*: Rate of occurrence and nature of the mutants. Genetics **94**: 961–968.

WRIGHT, S., 1949 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, New Jersey.

Corresponding editor: M. NEI