ANALYSES OF THE AGE OF GENES AND THE FIRST
ARRIVAL TIMES IN A FINITE POPULATION

TAKEO MARUYAMA* AND PAUL A. FUERST†,

*National Institute of Genetics, Mishima 411, Japan, and †Department of Genetics, The Ohio State University, Columbus, Ohio 43210

Manuscript received August 23, 1982
Revised copy accepted August 25, 1983

ABSTRACT

The age of a mutant gene is studied using the infinite allele model in which every mutant is new and selectively neutral. Based on a time reversal theory of Markov processes, we develop a method of mathematical analysis that is considerably simpler for calculating the various statistics of the age than previous methods. Formulas for the mean and variance and for the distribution of age are presented together with some examples of relevance to cases in natural populations.—Theoretical studies of the first arrival time of an allele to a specified frequency, given an initially monomorphic condition of the locus, are presented. It is shown that, beginning with an allele that has frequency \( p = 1 \) or an allele with frequency \( p = 1/2N \), there is an initial lag phase in which there is virtually no chance of an allele with a specified intermediate frequency appearing in the population. The distribution of the first arrival time is also presented. The distribution shows several characteristics that are not immediately obvious from a consideration of only the mean and variance of first arrival time. Especially noteworthy is the existence of a very long tail to the distribution. We have also studied the distribution of the age of an allele in the population. Again, the distribution of this measure is shown to be more informative for several questions than are the mean and variance alone.

The probabilistic fate of mutant genes in finite populations, measured by statistics such as the fixation probability or the age of a gene, has been extensively studied by numerous authors. These stochastic analyses play an important role in population genetic and evolutionary theory (for a review, see Kimura and Ohta 1971; Nei 1975). The infinite allele model of Kimura and Crow (1964) is particularly important in molecular evolution theory and has, thus, been extensively studied. The model assumes that every mutant is new and is selectively neutral. Because of the neutrality assumption, the model is readily amenable to mathematical analysis, and many stochastic properties of the model have been obtained. However, almost all of these properties have been developed essentially for equilibrium populations only.

Recently, nonequilibrium approaches have become increasingly important. For instance, the role of founder effects during speciation and the effect of population bottlenecks on genetic variability and genetic change require math-
matical analyses involving nonequilibrium theory. These nonstationary properties are important, and some of them have been studied. For instance, the transient behavior of gene frequencies starting from a completely monomorphic population is biologically relevant for cases such as bottlenecks (NEI, MARUYAMA and CHAKRABORTY 1975).

We can define the age of an extant allele as the period of time that the allele has persisted in a population since it was introduced. Under the assumption of the infinite allele model, each allele will be introduced into the population exactly once, and, therefore, the age of an allele can be meaningfully studied. Aspects of the age of a gene have been studied previously by various mathematical methods (KIMURA and OHTA 1973; MARUYAMA 1974; LEVIKSON 1977; SAWYER 1977; WATTERSON 1977; NAGASAWA and MARUYAMA 1979).

Since we deal with selectively neutral genes in a finite population, an allele of a given frequency \( x \) can have had many different histories. Here, we are concerned with statistical properties of the age; the mean and the variance of ages of mutant genes are particularly important. The distribution itself of ages will be most informative, however, although this distribution is often difficult to obtain. The distribution may have a long tail contributing to a large variance; in such a case, the mean and variance may not provide sufficient information.

Recently, WATTERSON (1977) and NAGASAWA and MARUYAMA (1979) developed an approximation method that allows the distribution of age to be obtained easily. Using this theory, we will derive a formula for the variance of ages of alleles of a given frequency and present a numerical method for calculating the distribution of age.

There are other transient properties of interest. Various statistics of genetic variation in a population which starts from a completely monomorphic state are biologically interesting. If a population is founded from a few individuals or genomes through an extreme bottleneck, it often loses most genetic variability. It usually takes a long time for the population to recover the original level of the variability (see NEI, MARUYAMA and CHAKRABORTY 1975; LI and NEI 1975). In this paper we present a study on the time required for a mutant gene to reach a given frequency from a very low frequency, such as a mutant present singly in a population. This is referred to as the first arrival time. This statistic seems to be relevant for examining the hypothesis which claims that many natural populations are not in genetic equilibrium because they have undergone severe population restrictions. The hypothesis is intended to account for the fact that some species which presently have large population size show levels of genetic variation much lower than that expected from the equilibrium state of the infinite allele model (NEI 1980).

Throughout the present paper, we assume that the population consists of \( N \) diploid individuals and, thus, there are \( 2N \) genes. We assume, further, that all alleles are selectively neutral and every mutant is new. Therefore, two genes are identical only if they are identical by descent. We denote by \( v \) the mutation rate per gene per generation.

Let \( t \) be the time measured in units of generations and consider a particular allele \( A \) which is introduced into the populations at \( t = 0 \). We assume that \( N \)
is sufficiently large so that the diffusion approximation is valid. Let $\phi(t, p, x)$ be the transient probability density that the frequency of $A$ is $x$ at time $t$ given that it is $p$ at $t = 0$. Then the density function satisfies the Komologorov backward equation:

$$\frac{\partial \phi}{\partial t} = L[\phi]$$

(1)

where

$$L[\phi] = \frac{x(1 - x)}{4N} \frac{\partial^2}{\partial x^2} - vx \frac{\partial}{\partial x}$$

(2)

(Crow and Kimura 1970). Equation (1) and equations derived from (1) will be extensively used in what follows.

**AGE OF MUTANT GENES**

For a given $t$, let $g(t, x)$ be the probability that the age of an allele is $t$, provided that the present frequency is $x$.

Let

$$G(t, x) = \int_0^t g(\xi, x) d\xi.$$  

(3)

Namely, as a function of $t$, $G(t, x)$ is the distribution of ages of alleles of present frequency $x$, and $g(t, x)$ is its probability density. According to Watterson (1977) and Nagasawa and Maruyama (1979), $G(t, x)$ satisfies

$$\frac{\partial G(t, x)}{\partial t} = L[G(t, x)],$$

(4)

with $L[\phi]$ given by (2). The boundary conditions are

$$G(0, 0) = 1$$

$$G(0, x) = 0 \quad \text{for} \quad 0 < x \leq 1,$$

$$G(t, 0) = 1$$

$$\left. \frac{\partial G(t, x)}{\partial x} \right|_{x = 1} < \infty,$$

(Nagylaki 1974). Therefore, to obtain the density $g(t, x)$, we first calculate $G(t, x)$ and then take its derivative with respect to $t$. It is important to note that this method of calculating the age distribution is considerably simpler than that of Maruyama (1974) or Maruyama and Kimura (1975).

Although it is possible to express the solution of (4) as a series of eigenfunctions of the operator $L[\phi]$, numerical calculations are more practical. Numerical integration of (4) with the conditions given by (5) is quite simple. We approximate the solution of (4) by a solution of the difference equation corresponding to the differential equation (4). A method is given by Maruyama (1977, pp. 95–101).
A number of cases of $g(t, x)$ are given in Figure 1. Each curve in Figure 1 represents the graph of the density $g(t, x)$ of the age distribution for a specified value of $4N\nu$, given that the present gene frequency is $x = 0.5$. As shown by Maruyama (1974), the average age increases as $4N\nu$ decreases, and this is evident from the graphs in Figure 1. More detailed information on the mean is given in Table 1. However, note that the peaks of $g(t, x)$ lie near $t = N$ or less, and that the tails of $g(t, x)$ increase as the average increases. When the value of $4N\nu$ is large, the distribution of age is nearly symmetrical on both sides of the peak, whereas the distribution becomes strongly asymmetrical as $4N\nu$ becomes small. Therefore, if $4N\nu$ is small, such as 10 or less, the mean and the variance of the age may not give enough information for some problems, such as the youngest possible age for an allele of specified frequency. In fact, as summary statistics, there is a relatively constant relationship between the mean and standard deviation (see Table 1) which fails to convey the true relationship illustrated in the distribution. Since values for $4N\nu$ which are regarded to be important in natural populations are usually 10 or less, it will be necessary to know more than the mean and variance. The present numerical method of obtaining $g(t, x)$ based on Watterson (1977) and Nagasawa and Maruyama (1979) will be useful. Although most of the results presented in figures and tables are computed for values of $4N\nu = 5$ or greater, the methods can be used to obtain results for any value of $4N\nu$. If bottlenecks are important in natural populations, the apparent $4N\nu$ value will be much less than 10, even though the true value may be large. Since it is the true value which concerns us here, we generally present results for a larger set of $4N\nu$ values.

Figure 1.—Distribution density of the age of mutant genes whose present frequency is 0.5. a, $4N\nu = 0.5$; b, $4N\nu = 1$; c, $4N\nu = 2$; d, $4N\nu = 5$; e, $4N\nu = 10$; f, $4N\nu = 20$; g, $4N\nu = 30$; h, $4N\nu = 50$. 
AGE AND ARRIVAL TIME

TABLE 1

The mean and variance of the age of gene whose frequency is \( x \)

<table>
<thead>
<tr>
<th>4N_v</th>
<th>10</th>
<th>5</th>
<th>2.0</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.483</td>
<td>0.679</td>
<td>1.021</td>
<td>1.888</td>
</tr>
<tr>
<td></td>
<td>2.654</td>
<td>4.600</td>
<td>9.803</td>
<td>37.613</td>
</tr>
<tr>
<td>0.3</td>
<td>0.808</td>
<td>1.230</td>
<td>2.062</td>
<td>4.462</td>
</tr>
<tr>
<td></td>
<td>3.725</td>
<td>7.349</td>
<td>19.612</td>
<td>117.54</td>
</tr>
<tr>
<td>0.5</td>
<td>0.988</td>
<td>1.562</td>
<td>2.770</td>
<td>6.553</td>
</tr>
<tr>
<td></td>
<td>5.620</td>
<td>12.364</td>
<td>37.778</td>
<td>263.82</td>
</tr>
<tr>
<td>0.7</td>
<td>1.115</td>
<td>1.805</td>
<td>3.327</td>
<td>8.383</td>
</tr>
<tr>
<td></td>
<td>8.146</td>
<td>19.265</td>
<td>63.634</td>
<td>476.15</td>
</tr>
<tr>
<td>0.9</td>
<td>1.213</td>
<td>1.998</td>
<td>3.791</td>
<td>10.021</td>
</tr>
<tr>
<td></td>
<td>11.190</td>
<td>27.791</td>
<td>96.599</td>
<td>753.23</td>
</tr>
</tbody>
</table>

The upper numbers are the means and the lower numbers the variances. The unit of time = \( N \) generations.

THE MEAN VARIANCE OF AGE

Let \( m_1(x) \) and \( m_2(x) \) be the first and second moment of \( i.e., \ g(t, x) \)

\[
m_1(x) = \int_0^\infty t g(T, x) dt = \int_0^\infty t dG(t, x). \tag{6}
\]

and

\[
m_2(x) = \int_0^\infty t^2 g(t, x) dt = \int_0^\infty t^2 dG(t, x). \tag{7}
\]

Then it can be shown that these quantities satisfy

\[
L[m_1(x)] + 1 = 0, \tag{8}
\]

and

\[
L[m_2(x)] + 2m_1(x) = 0, \tag{9}
\]

where \( L[\phi] \) is the operator given by (2) [MARUYAMA and KIMURA (1971); NAGYLAKI (1974); see also KARLIN and TAYLOR (1975) for a general approach using Green's function]. The boundary conditions are the same as for (5), accounting for the absence of \( t \) as a parameter in (8). Solution of (8) yields

\[
m_1(x) = m_1 = \frac{4N_v}{F} \left[ \int_0^x \frac{1 - (1 - y)^e}{y(1 - y)^e} + [1 - (1 - x)^e] \int_x^1 \frac{1}{y(1 - y)^e} dy \right] \tag{10}
\]

where \( F = 1 - 4N_v \) and
where $F = 1 - 4Nu$ and $m_1(y)$ is given by (10). Note that formula (11) for the second moment of the age distribution is not much more complicated than formula (10) for the mean. Compared with formula (11), the method given by MARUYAMA (1974) for the second moment is considerably more difficult.

THE FIRST ARRIVAL TIME

In the previous section, we dealt with the age of an allele having a specified frequency. If a mutant gene visits a specified frequency more than once, each visit becomes the subject of the calculation of age discussed before. Consider, however, the first arrival time to a particular frequency, disregarding all subsequent visits. This is relevant if we want to know the approximate time until a specified allele frequency will reappear following a complete loss of population variation. Figure 2 illustrates the concept of the first arrival time of a path.

When the method of MARUYAMA and KIMURA (1971) is used, the first arrival time for a mutant gene having some deleterious effects was studied by LI (1975). He concluded that the required time until a selectively deleterious mutant arrives at a certain frequency decreases as the gene effect becomes large. MARUYAMA (1972, 1974) reached similar conclusions in a study of age of genes with either a positive or negative effect. Whether the effect was advantageous or deleterious, the first arrival time decreased to the same extent as the absolute selective effect increased, although the proportion of deleterious genes arriving at a particular frequency declined rapidly with increasing selection (MARUYAMA 1972).

We select for discussion here only those mutants that eventually arrive at the specified frequency at least once before disappearing from the population.

![First Arrival Time](image-url)
To carry out the analysis we need a mathematical tool. Let $u_s(p)$ be the probability that a mutant starting with frequency $p$ at $t = 0$ reaches frequency $x$ at least once. Then under the present model,

$$
\begin{align*}
  u_s(p) &= \frac{1 - (1 - p)^x}{1 - (1 - x)^p} \quad \text{if} \quad p < x \\
  u_s(p) &= 1 \quad \text{if} \quad p > x
\end{align*}
$$

(12)

where $F = 1 - 4Nv$ and $x$ is a fixed frequency [see Karlin and Taylor (1975) for alternative approaches using Green's function]. Let $a_s(t, x)$ be the probability density of the first arrival time for a mutant starting at frequency $p$ and arriving at $x$. Let

$$
A_s(t, p) = \int_0^t a_s(\xi, p) d\xi.
$$

(13)

Now let

$$
\hat{L} = \frac{1}{u_s(p)} \left\{ \frac{p(1 - p)}{4N} \frac{\partial^2}{\partial p^2} - np \frac{\partial}{\partial p} \right\} u_s(p)
$$

(14)

where $u_s(p)$ is given by (12). Then according to Maruyama (1977), $A_s(t, p)$ satisfies

$$
\frac{\partial A_s(t, p)}{\partial t} = \hat{L}[A_s(t, p)]
$$

(15)

with the boundary conditions

$$
\begin{align*}
  A_s(0, p) &= 0 \quad \text{for} \quad p < x, \\
  A_s(t, x) &= 1 \quad \text{for all} \quad t \geq 0 \\
  A_s(t, 0) &= \text{finite} \\
  \left. \frac{dA_s(t, p)}{dp} \right|_{p = 1} &= \text{finite}
\end{align*}
$$

Analytic solution of (15) appears to be difficult. Numerical integration can, however, be easily carried out by the same method discussed in the preceding section. Note that $a_s(t, p)$ is a probability density for random variable $t$ with two parameters $x$ and $p$, where $x$ is the frequency arrived at and $p$ is the starting frequency. With respect to the parameter $p$, the cases of particular interest would be $p = \frac{1}{2N}$ and $p = 1$, representing the first arrival time beginning from a very rare frequency and that beginning as a fixed mutant. Since we cannot substitute 0 for $p$ in $a_s(t, p)$ we define

$$
a_s(t, 0) = \lim_{p \to 0} a_s(t, p).
$$

(16)

Figure 3 gives graphs of $a_s(t, 0)$ for $x = 0.5$. It is interesting to note that
Figure 3.—Distribution density of the first arrival time from a very low frequency \((1/2N)\) to \(x = 0.5\). a, \(4Nv = 0\); b, \(4Nv = 10\); c, \(4Nv = 20\); d, \(4Nv = 30\); e, \(4Nv = 40\).

The graphs for \(4Nv = 0\) and \(4Nv = 5\) or 10 are not much different. It is also worth noting that as \(4Nv\) increases the graphs of \(a(t, 0)\) become symmetrical around the means. Note that until \(0.2N\) or \(0.3N\) generations there is virtually no possibility of mutants arriving at frequency higher than 0.3. In addition, unless \(4Nv\) is fairly large (30 or greater), the distribution tends to have a long tail. This indicates that in many populations there would be a delay of \(N\) generations or more before a new mutant would reach a frequency of 0.5. This numerical study asserts that there is at least some delay in time before a new mutant at appreciable frequency can appear in the populations under the neutral model.

Figure 4 presents the graphs of \(a(t, 1)\) for \(x = 0.5\). In these cases, too, the first arrival time distribution has a very long tail when \(4Nv\) is small (less than 2). Compared with the first arrival at \(x = 0.5\) for new mutants (starting frequency \(p = 1/2N\)), that of fixed mutants (starting \(p = 1\)) is faster. In particular, the time span in which there will be virtually no arrival is considerably less when beginning at \(p = 1\). If \(4Nv\) is of the order of 10, the waiting time is only about \(0.05N\) generations for a fixed allele to reach \(x = 0.5\) by mutation and random drift.

The time required for a population to have a new mutant of frequency \(x\) is...
Figure 4.—Distribution density of the first arrival time for a mutant originally fixed in a population to arrive at $x = 0.5$ by mutation and random drift. $4Nv$ values are given along curves. More complicated than the first arrival time $a_\delta(t, 0)$ discussed before. That first arrival time has been calculated on the assumption that a mutant in question reaches the frequency before becoming extinct from the population. In fact the probability that a new mutant reaches the frequency $x$ can be calculated by substituting $p = 1/2N$ in the top formula of (12). Namely,

$$u_x \left( \frac{1}{2N} \right) = \frac{1 - \left( 1 - \left( \frac{1}{2N} \right) \right)^F}{1 - \left( 1 - x \right)^F} \approx \frac{F}{2N \left( 1 - \left( 1 - x \right)^F \right)}$$

Note that every new mutant has equal chance to reach the frequency $x$, and that the mean number of mutants to be introduced into a population in time $\Delta t$ is equal to $2Nv\Delta t$. Therefore, the waiting time for the first mutant that will reach $x$ is given by a negative exponential law,

$$e^{-2Nq(t)}$$

where $q = u_x(1/2N) \approx F/2N[1 - (1 - x)^F]$ which is the probability for a new mutant to reach $x$. Since the waiting time for the occurrence of the first mutant that will reach $x$ is given by (18), the probability density, $b_x(t)$, that the first
A successful mutant will appear at time $t$ is equal to

$$b_x(t) = 2N\sqrt{q}e^{-2Nqv}. \quad (19)$$

If a population becomes completely monomorphic, the distribution density, $\alpha_x(t, 0)$, of the time required for the first mutant to reach the frequency $x$ is equal to a convolution of $a_x(t, 0)$ given by (16) and $b_x(t, 0)$ of (19). Namely,

$$\alpha_x(t, 0) = \int_0^t b_x(t - \xi)a_x(\xi, 0)d\xi. \quad (20)$$

A case of the first arrival time distribution is given in Figure 5, where the three densities, $a_x(t, 0)$, $b_x(t)$ and $\alpha_x(t, 0)$ are illustrated.

The two first arrival times given by $a_x(t, 0)$ and $\alpha_x(t, 0)$ are quite different. The distribution $a_x(t, 0)$ is the time required for a mutant to reach the frequency $x$, provided that the mutant reaches $x$.

On the other hand, the time specified by $\alpha_x(t, 0)$ is the sum of the waiting time for the first successful mutant to arise and the time required for that mutant to travel from a very low frequency to $x$.

![Figure 5](image_url)

**Figure 5.**—The time required for a population to have a mutant of frequency $x = 0.5$ which have arised by mutation. Curve $a$ is the first arrival time, given that a particular mutant reaches $x$, curve $b$ is the waiting time for the first successful mutant to arise, and $\alpha$ is the convolution of $a$ and $b$. 
THE MEAN AND VARIANCE OF THE FIRST ARRIVAL TIME

Although less informative than the density $a_s(t, 0)$, $a_s(t, 0)$ or $a_s(t, 1)$, the mean and the variance of these first arrival times are also useful. Let

$$f_s(p) = \int_0^\infty t a_s(t, p) dt,$$

and

$$g_s(p) = \int_0^\infty t^2 a_s(t, p) dt.$$

Then, we have

$$\mathcal{L}[f_s(p)] + 1 = 0,$$

and

$$\mathcal{L}[g_s(p)] + 2f_s(p) = 0.$$ (24)

The solutions of (23) are

$$f_s(0) = \frac{4N[1 - (1-x)^f]}{F} \int_0^x \frac{u_s(y)[1 - u_s(y)]}{y(1-y)^f} dy,$$ (25)

and

$$f_s(1) = \frac{4N}{F} \left\{ (1-x)^f \int_x^1 \frac{dy}{y(1-y)^f} + \log x \right\}.$$ (26)

where $u_s(y)$ is given by (12). The second moments are

$$g_s(0) = \frac{8N[1 - (1-x)^f]}{F} \int_0^x \frac{f_s(y)u_s(y)[1 - u_s(y)]}{y(1-y)^f} dy,$$ (27)

and,

$$g_s(1) = \frac{8N}{F} (1-x)^f \int_0^1 \frac{f_s(y)}{y(1-y)^f} dy - \int_0^1 \frac{f_s(y)}{y} dy.$$ (28)

Formulas (25)–(28) can be used readily for numerical calculations.

The mean and variance of the waiting time given by (18) are, respectively, $1/(2N\nu q)$ and $1/(2N\nu q)^2$, where $q = F/2N[1 - (1-x)^f]$. Therefore, the mean of the time required for a new mutant to reach the frequency $x$ is the sum of the two means, $f_s(0)$ of (22) and $1/(2N\nu q)$. Similarly, the variance is the sum of the two variances, i.e., $g_s(0) - f_s(0)^2$ plus $1/(2N\nu q)^2$. Figure 6 presents graphs of the mean time required for a new mutant to reach a specified frequency. It is interesting to note that the mean time is minimum when the value of $4N\nu$ is about unity, and it increases as $4N\nu$ either increases or decreases. For instance, if the frequency $x$ is 0.5, the mean times for cases for $4N\nu = 0.1$, 0.5, 1, 5 and 10 are 21.9N, 5.9N, 3.9N, 4.1N and 23.6N, respectively. Also, note that, if the frequency is less than about 0.2 and if $4N\nu$ is larger than 1, the mean time is approximately equal to $5xN$ generations, where $x$ is the specified
frequency. Table 2 gives the mean and variance of the first arrival time for several cases $4Nv$ and $x$.

The mean first arrival time $f_x(1)$ for various values of the parameter $4Nv$ are given graphically in Figure 7 and in more detail in Table 3. It is worth noting that the mean and the peak of the first arrival distribution can be quite different; this can be seen by a comparison of Figure 4 with Figure 7. For instance, when $4Nv = 1.0$, the mean first arrival time is about $2.5N$ generations, but the peak of the distribution is located at about $0.75N$ generations. The difference is large. Therefore, even though the mean is larger than $2N$ generations, alleles of intermediate frequency will appear after about $0.5N$
TABLE 2
The mean and variance of the time required for a new mutant to arise by mutation and to reach the frequency $x$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$4Nv$ 10</th>
<th>$4Nv$ 5</th>
<th>$4Nv$ 2.0</th>
<th>$4Nv$ 0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.273</td>
<td>0.309</td>
<td>0.427</td>
<td>1.026</td>
</tr>
<tr>
<td></td>
<td>0.0187</td>
<td>0.0252</td>
<td>0.0637</td>
<td>0.688</td>
</tr>
<tr>
<td>0.3</td>
<td>1.650</td>
<td>1.285</td>
<td>1.526</td>
<td>3.283</td>
</tr>
<tr>
<td></td>
<td>1.222</td>
<td>0.539</td>
<td>0.885</td>
<td>6.981</td>
</tr>
<tr>
<td>0.5</td>
<td>23.594</td>
<td>4.124</td>
<td>3.225</td>
<td>5.918</td>
</tr>
<tr>
<td></td>
<td>515.98</td>
<td>9.386</td>
<td>4.500</td>
<td>22.47</td>
</tr>
<tr>
<td>0.7</td>
<td>$2.26 \times 10^3$</td>
<td>$26.062$</td>
<td>$6.601$</td>
<td>$9.198$</td>
</tr>
<tr>
<td></td>
<td>$5.10 \times 10^6$</td>
<td>600.49</td>
<td>23.01</td>
<td>53.65</td>
</tr>
<tr>
<td>0.9</td>
<td>$4.44 \times 10^7$</td>
<td>$2.00 \times 10^3$</td>
<td>$20.974$</td>
<td>$14.05$</td>
</tr>
<tr>
<td></td>
<td>$5.18 \times 10^{15}$</td>
<td>$4.00 \times 10^6$</td>
<td>326.82</td>
<td>122.92</td>
</tr>
</tbody>
</table>

generations following a complete loss of variability. It is important to realize that the waiting time for the first arrival of alleles of intermediate frequency can be only poorly inferred from information on the mean and variance of the first arrival times, particularly when $4Nv$ is small.

DISCUSSION

The results presented in this paper are important for several reasons. As theory, the methodology simplifies the effort that must be expended to obtain information on the age of alleles and on the first arrival time. Previous methods were more cumbersome but did not give appreciably better estimates of the statistics. In addition, the distributions of the statistics were not easily obtained using earlier methods. The finding of a lag time, in which there is virtually no chance of observing alleles in various frequency classes is particularly important. The implications of these results for the allele frequency distribution following a bottleneck will be examined in a future study, extending the studies of Li and Nei (1975).

The results are relevant to the interpretation of patterns of variability observable in natural populations. Many species, especially vertebrate species, may periodically experience extreme population restrictions which act to reduce average genetic variability (Nei 1980). In natural populations that have gone through a bottleneck in the recent past, a large number of alleles will not have had sufficient time to move from essentially monomorphic frequencies to more intermediate frequencies. The results that are presented in Figures 4–7 illustrate that about $0.2N$ generations are necessary for the reappearance of alleles with intermediate frequencies, especially when the value of $4Nv$ is small. Even though the mutation rate, $v$, will always be small, the value of $N$ may be substantial, and the value of $0.2N$ generations may not be trivial.
FIGURE 7.—The mean time required for a fixed mutant to reduce its frequency to $x$ by mutation and random drift. a, $4N\mu = 0.1$; b, $4N\mu = 0.5$; c, $4N\mu = 1$; d, $4N\mu = 2$; e, $4N\mu = 5$; f, $4N\mu = 10$, g, $4N\mu = 20$.

Some may question how important such a theoretical finding would be, especially when electrophoretic polymorphism is assumed to be ubiquitous (Selander et al. 1974). Examination of the literature on electrophoretic population surveys [Fuerst, Chakraborty and Nei (1976a); extended by P. A. Fuerst, unpublished results] indicates, however, that reports of little or no
detectable electrophoretic variation are not uncommon, even when the sample of loci or the number of individuals examined are reasonably large. For instance, in mammals, no electrophoretic variability was found in the northern elephant seal *Mirounga angustirostris* (BONNELL and SELANDER 1974), several species or subspecies of primates (NOZAWA *et al.* 1977; HRDY, BARNICOT and ALPER 1975), the polar bear (F. ALLENDORF, F. CHRISTIANSSEN, W. EANES, F. FRYDENBERG, unpublished results), several species of mainland rodents (SELANDER *et al.* 1975; JOHNSON and SELANDER 1971) and in a number of island populations of mice (AVISE *et al.* 1974). Within lower vertebrates, monomorphic species samples have been obtained from island-dwelling Anolis lizards (WEBSTER, SELANDER and YANG 1973), from several different species of freshwater fish (KORNFIELD *et al.* 1979; BUTH 1979) and from estuarine species of saltwater fish (JOHNSON and UTTER 1976). Even in the invertebrates, which often show higher average heterozygosities, monomorphic species have been reported in Hymenoptera (SNYDER 1975) and in the Odonata (KNOFF 1977). In addition, of course, there are a number of species that are monomorphic because of special breeding systems, but these are not relevant to the present considerations. In addition to these monomorphic species and subspecies, there are a large number of samples from natural populations that have variability at only one or two loci out of a much larger sample of loci (FUERST, CHAKRABORTY and NEI 1976a). If we examine natural populations, there is usually good agreement between the expectations of the neutral mutation model and observed statistics of genic variation (cf. FUERST, CHAKRABORTY and NEI 1976a). However, in a number of the species there is an indication that intermediate frequency alleles are slightly less frequent than would be expected
This is also reflected to a small degree in the increase in the variance of heterozygosity which was observed in the studies of Fuerst, Chakraborty and Nei (1976a).

The estimation of $4Nv$ from natural populations has caused some controversy. Ewens (1972) has cogently made the point that the best unbiased estimator of $4Nv$ is the number of alleles in a sample. This will be true, however, only for equilibrium populations. Chakraborty and Fuerst (1979) and others have argued that, because of the sensitivity of number of alleles to nonequilibrium conditions, estimation of $4Nv$ from the average heterozygosity in a population might be a preferred alternative strategy. Again, however, the results of Nei, Maruyama and Chakraborty (1975) indicate that heterozygosity will also be affected by nonequilibrium factors such as population bottlenecks. If we consider the information that is available from natural populations concerning both average heterozygosity and number of alleles, we are led to believe that the value of $4Nv$ characterizing most natural populations will be fairly small, often of the order of 0.5 or much less. If such estimates are indicative of the true value of $4Nv$, then the results which we are presenting here indicate that most populations will require very long periods in recovering intermediate alleles that have been lost during bottlenecks.

For the transient period following a monomorphic state, the results given in this paper suggest that a longer period of time will be required before a pair of populations will show the constant, statistically stable, rate of increase in genetic distance predicted by the theory of Nei (1972). This accords with the results presented by Chakraborty and Nei (1977), which showed that a bottleneck will cause a rapid increase in genetic distance, followed by very little change in genetic distance for a substantial period of time. Our results indicate that this substantial static period results from the long waiting time until intermediate alleles reappear in the population. In addition, because alleles that are made monomorphic following a bottleneck require an extended period of time to be lost from the population (see Figure 7), there should be considerable sharing of alleles in natural populations between related species, a sharing that has little to do with the adaptive value of the alleles to the species involved.

This type of pattern would appear to be consistent with the results reported by Koehn and Eanes (1978), who showed that a group of loci (presumably those with small $4Nv$ values) showed substantial sharing of alleles between species, whereas another group of loci with higher average heterozygositites and presumably higher $4Nv$ values showed substantial differentiation. The difference in first arrival time caused by differences in $4Nv$ values would also contribute to the negative relationship between heterozygosity and genetic distance which has been noted by Fuerst, Chakraborty and Nei (1976b) and Prakash (1977). It is important to keep in mind that the present results were obtained using a model that assumed that all loci can be treated as replicates of the same $4Nv$ value. To make truly valid comparisons with data from natural populations for many loci, some results incorporating the assumptions of the varying mutation model of Nei, Chakraborty and Fuerst (1976) are called for. However, this latter model is considerably more complex.
than the model considered here, and its analysis has been put off to a future study.

The theory presented in this paper is based on a model that assumes that there was complete monomorphism in the initial population. The results of NEI, MARUYAMA and CHAKRABORTY (1975) for bottlenecked populations indicate that this will not be true of every population which goes through a bottleneck. However, the analysis of data from natural populations suggests that many, or even most, loci in natural populations will be essentially monomorphic (FUERST, CHAKRABORTY and NEI 1976a). This is especially true of vertebrate species. The patterns of variability seen in populations that have repeatedly gone through population bottlenecks, such as introduced species, are in agreement with this assertion (HUETTEL et al. 1980). Because the majority of loci will be monomorphic, we believe that the results presented here are important for natural populations.

One additional assumption of the theory was that the population was returned to a fixed population size immediately following the event that caused monomorphism. This would not necessarily be true of populations that have actually undergone a bottleneck. Rather, they may gradually increase in size, until they reach their new equilibrium size. However, if the period of time required for a reestablishment of population size is small relative to the mutation rate, which would be true of most situations, the assumption of immediate increase in population size would have a trivial effect on the results given here. It should be noted that a similar approach was used by Li and NEI (1975) in their initial studies of the allele frequency distribution under transient conditions. The effects caused by the assumption, if any, will be studied in future work.

This work was partially supported by research grant DEB-8110220 from the United States National Science Foundation and by grant A-439017 from the Japanese Ministry of Education, Science and Culture.

LITERATURE CITED


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


Corresponding editor: B. S. WEIR