

THE AVERAGE NUMBER OF GENERATIONS UNTIL  
EXTINCTION OF AN INDIVIDUAL MUTANT  
GENE IN A FINITE POPULATION

MOTOO KIMURA<sup>1</sup> AND TOMOKO OHTA

*Department of Biology, Princeton University*  
and

*National Institute of Genetics, Mishima, Japan*

Received June 13, 1969

AS pointed out by FISHER (1930), a majority of mutant genes which appear in natural populations are lost by chance within a small number of generations. For example, if the mutant gene is selectively neutral, the probability is about 0.79 that it is lost from the population during the first 7 generations. With one percent selective advantage, this probability becomes about 0.78, namely, it changes very little. In general, the probability of loss in early generations due to random sampling of gametes is very high.

The question which naturally follows is how long does it take, on the average, for a single mutant gene to become lost from the population, if we exclude the cases in which it is eventually fixed (established) in the population.

In the present paper, we will derive some approximation formulas which are useful to answer this question, based on the theory of KIMURA and OHTA (1969). Also, we will report the results of Monte Carlo experiments performed to check the validity of the approximation formulas.

APPROXIMATION FORMULAS BASED ON DIFFUSION MODELS

Let us consider a diploid population, and denote by  $N$  and  $N_e$ , respectively, its actual and effective sizes. The following treatment is based on the diffusion models of population genetics (cf. KIMURA 1964).

As shown by KIMURA and OHTA (1969), if  $p$  is the initial frequency of the mutant gene, the average number of generations until loss of the mutant gene (excluding the cases of its eventual fixation) is

$$\bar{t}_0(p) = \frac{u(p)}{1-u(p)} \int_p^1 \psi(\xi) \{1-u(\xi)\}^2 d\xi + \int_0^p \psi(\xi) \{1-u(\xi)\} u(\xi) d\xi. \quad (1)$$

In this formula,

$$u(p) = \int_0^p G(x) dx \bigg/ \int_0^1 G(x) dx$$

<sup>1</sup> On leave from the National Institute of Genetics, Mishima, Japan. This paper constitutes Contribution number 731 from the National Institute of Genetics, Mishima, Shizuoka-ken, Japan.

is the probability of ultimate fixation of the mutant gene (cf. KIMURA 1957, 1962), and

$$\psi(\xi) = 2 \int_0^1 G(x) dx / \{V_{\delta\xi} G(\xi)\}, \quad (2)$$

where

$$G(x) = \exp \left\{ -2 \int_0^x (M_{\delta x} / V_{\delta x}) dx \right\}, \quad (3)$$

and  $M_{\delta x}$  and  $V_{\delta x}$  are respectively the mean and the variance of the rate of change per generation of the mutant gene frequency  $x$ . In the present paper we will be concerned with mutant genes that are either deleterious or selectively neutral.

Let  $s'$  and  $s'h$  be respectively the selection coefficient against the mutant homozygotes and heterozygotes such that

$$M_{\delta x} = -s'x(1-x)\{h+(1-2h)x\}. \quad (4)$$

Also, we will assume that the sole factor causing random fluctuation in gene frequency is the random sampling of gametes in reproduction so that

$$V_{\delta x} = x(1-x)/(2N_e), \quad (5)$$

where  $N_e$  is the "variance" effective number. Then, if the mutant allele is represented only once at the moment of its appearance, the average number of generations until extinction may be obtained by putting  $p=1/(2N)$  in formula (1) in which  $M_{\delta x}$  and  $V_{\delta x}$  are given respectively by (4) and (5). However, the resulting formula contains integrals that can not be reduced in general to elementary functions. So, we will consider separately three cases, namely, 1) neutral mutations, 2) semidominant deleterious mutations, and 3) completely recessive deleterious mutations. In the following, we will denote by  $\bar{t}_0$  the value of  $\bar{t}_0(p)$  at  $p=1/(2N)$ .

1) *Selectively neutral mutations*: In this case,  $s' = 0$  and therefore  $M_{\delta x} = 0$ , giving  $G(x) = 1$ ,  $u(p) = p$  and  $\psi(\xi) = 4N_e/\{\xi(1-\xi)\}$ . Thus, formula (1) with  $p = 1/(2N)$  reduces to

$$\bar{t}_0(1/(2N)) = \frac{4N_e}{2N-1} \log_e(2N), \quad (6)$$

as given by KIMURA and OHTA (1969). Writing  $\bar{t}_0$  for  $\bar{t}_0(1/(2N))$  and assuming that  $2N$  is much larger than unity, we obtain, with good approximation,

$$\bar{t}_0 = 2 \left( \frac{N_e}{N} \right) \log_e(2N). \quad (7)$$

2) *Semidominant deleterious mutations*: In this case,  $h=1/2$  and, we will denote by  $s'_1$  the selection coefficient against mutant heterozygotes such that  $M_{\delta x} = -s'_1x(1-x) = -(s'/2)x(1-x)$ . Then, it can be shown that, if  $s'_1 (>0)$  is small but  $4N_e s'_1$  is large, we obtain approximately

$$\bar{t}_0 = 2 \left( \frac{N_e}{N} \right) \left\{ \log_e \left( \frac{N}{2N_e s'_1} \right) + 1 - \gamma \right\}, \quad (8)$$

where  $\gamma$  is Euler's constant,  $0.577 \dots$ . Note that  $\bar{t}_0$  in this case depends mainly

on selective disadvantage ( $s'_1$ ) but not on population numbers as long as the ratio  $N_e/N$  remains constant. This formula may also be expressed as

$$\bar{t}_0 = 2 \left( \frac{N_e}{N} \right) \{ \log_e(2N) - \log_e(2N_e s') + 1 - \gamma \} , \tag{8'}$$

where  $s' = 2s'_1$  ( $2N_e s' \gg 1$ ).

3) *Completely recessive deleterious mutations*: In this case,  $h=0$  and it can be shown that, if  $2N_e s'$  is large, we obtain

$$\bar{t}_0 = 2 \left( \frac{N_e}{N} \right) \left\{ \log_e(2N) - \frac{1}{2} \log_e(2N_e s') + 1 - \frac{\gamma}{2} \right\} \tag{9}$$

as a good approximation, where  $s'$  is the selection coefficient against mutant homozygotes.

EXPERIMENTAL CHECK ON THE APPROXIMATION FORMULAS  
BY THE MONTE CARLO METHOD

In order to check the validity of the approximation formulas (7), (8) and (9), Monte Carlo experiments were performed using TOSBAC 3400 and IBM 360 computers.

These experiments cover the cases of neutral, semidominant deleterious and completely recessive deleterious mutations, using the following scheme. In each generation, the gene frequency was changed deterministically using the formula

$$\Delta x = -s'x(1-x) \{ h + (1-2h)x \} / \bar{w},$$

where  $\bar{w} = 1 - 2hs'x(1-x) - s'x^2$ . Sampling was performed by generating pseudo-random numbers,  $X$  ( $0 < X < 1$ ), using the subroutines RAND in TOSBAC 3400 and RANDU in IBM 360. In one of the two programs, this was done at the gametic stage and in another at the zygotic stage. Each experiment was continued until fixation or extinction of the mutant gene occurred and the number of generations required was recorded.

Table 1 shows the comparison of the results of Monte Carlo experiments and those of the analytical solution for the case of neutral mutations. Two hundred

TABLE 1  
*Comparison of theoretical and Monte Carlo results for the average number of generations until extinction of a neutral mutant gene*

Population size* ( $N_e$ )	Theoretical† result	Monte Carlo‡ result
10	6.3	7.2
20	7.6	6.5
30	8.2	8.0
40	8.9	7.3
50	9.3	11.4
100	10.6	14.1
200	12.0	13.6

\* Actual and effective sizes assumed equal,  $N_e = N$ .

† Computed using text formula (6).

‡ 195-200 replications to obtain each experimental value.

trials were made to get each experimental value, although actual replication ranged from 195 to 200 because mutant genes became fixed rather than lost in a few cases. As seen from the table, agreement between the Monte Carlo results and the theoretical predictions is satisfactory. When  $2N_e$  is large, the variance for the time until extinction becomes large, and the agreement tends to become less satisfactory. This will be discussed in the next section.

In Figure 1, the results for the semidominant deleterious mutations are shown for various degrees of disadvantage. The curve represents the theoretical predictions and the dots represent the results of Monte Carlo experiments. The two agree with each other fairly well. The effective and the actual population numbers are different in these experiments, i.e.,  $N=75$  and  $N_e=50$ . Each experiment was repeated 50 times, except for the case of  $s'=0$ , for which 500 replications were made. This is because a larger variance is expected for neutral mutations than for disadvantageous mutations. Almost all deleterious mutations disappear within a few generations. On the other hand, neutral mutations sometimes increased, persisting in the population for quite a long time until extinction occurs, hence creating a much larger variance.

The results for completely recessive deleterious mutations are shown in Figure 2. Theoretical values are given by the curve and the experimental results by dots. Two hundred replications were made to get each experimental value, except for the case of  $s'=0$ , for which the same experimental value as given in Figure 1 was plotted. There are more fluctuations in these experimental results than in the previous case of semidominance, but the agreement is satisfactory. It may be

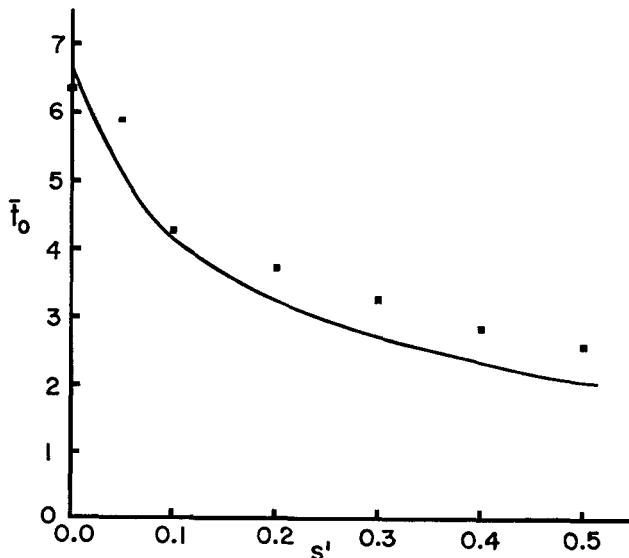


FIGURE 1.—The relationship between the average time until extinction ( $\bar{t}_0$ ) of a single mutant gene and its selective disadvantage ( $s'$ ) for the case of semidominant deleterious mutations. In this figure, the curve represents analytical results (from text formula 8') and the dots represent the results of Monte Carlo experiments, in which  $N_e=50$  and  $N=75$  are assumed.

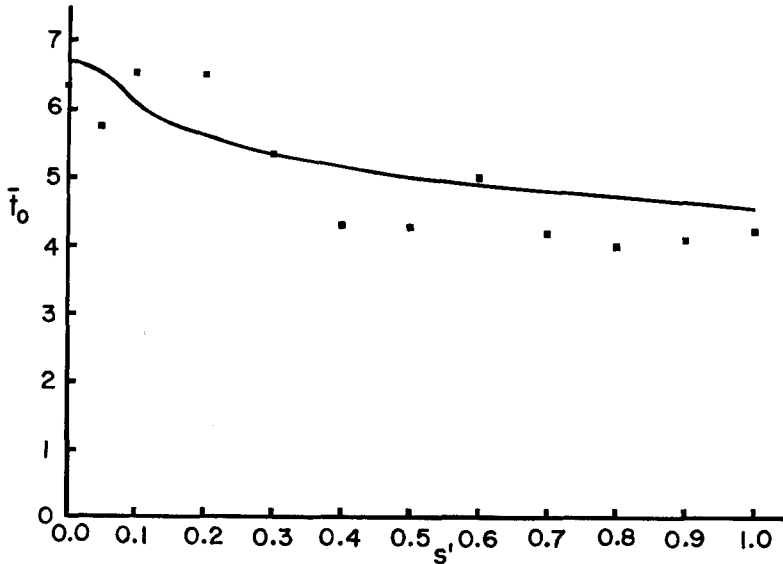


FIGURE 2.—The relationship between the average time until extinction ( $\bar{t}_0$ ) of a single mutant gene and its selective disadvantage ( $s'$ ) for the case of completely recessive deleterious mutations. The theoretical values (from text formula 9) are represented by the curve and the results of Monte Carlo experiments by dots. In the experiments,  $N_e=50$  and  $N=75$ .

noted that for recessive mutations the time until extinction decreases only very slowly with the increase of disadvantage.

#### DISCUSSION

Since a majority of mutations are expected to be either deleterious or at best neutral, it may be easily imagined that the average time until extinction ( $\bar{t}_0$ ), as studied in the previous sections, has an important bearing on the frequency of rare molecular variants in the population. Actually, it can be shown that if a gene (cistron) consists of a large number of codons, we have

$$I_1/N = 2\mu\bar{t}_0, \quad (10)$$

where  $I_1$  is the number of temporarily segregating codons (i.e., number of different kinds of rare molecular variants) in a population of actual size  $N$  and effective size  $N_e$ , and  $\mu$  is the mutation rate per gene per generation. This formula was used to estimate the mutation rate per codon in human hemoglobin genes (KIMURA, unpublished).

Recently, there is growing evidence for suggesting that random fixation of selectively neutral mutations is playing an important role in molecular evolution (KIMURA 1968, 1969; CROW 1969; KING and JUKES 1969). So it may be appropriate here to consider in some detail the behavior of selectively neutral mutants in a finite population.

As shown in the present paper, as well as in our previous paper (KIMURA and

TABLE 2

*Variance of the number of generations until extinction for neutral mutations†  
with various population sizes\**

Population size* ( $N_e$ )	Variance	
	From text formula (11)	From Monte Carlo experiment
10	120	192
20	262	167
30	413	320
40	561	237
50	714	1207
100	1490	2966
200	3060	2327

\*  $N_e = N$  is assumed.

† A single mutant gene is assumed at the start.

OHTA 1969), the time until extinction ( $t_0$ ) is fairly short for neutral mutants in the sense that it has a small mean value. Namely, the average number of generations until extinction is about  $2(N_e/N)\log_e(2N)$ . On the other hand, it has a large variance relative to the mean. In fact, as shown in the APPENDIX, the variance of the number of generations until extinction (excluding the cases of eventual fixation) is

$$\text{Var}(t_0) = 16(N_e^2/N) - [2(N_e/N)\log_e 2N]^2. \quad (11)$$

If both  $N$  and  $N_e$  are very large, this gives the approximate standard deviation

$$\sigma(t_0) \approx 4N_e/\sqrt{N} \quad (12)$$

which is much larger than the mean.

Table 2 shows some results of Monte Carlo experiments to check formula (11). The agreement between theoretical predictions and the experimental results is fairly good. In Table 3 a frequency distribution of the length of time until extinction is presented, based on Monte Carlo experiments.

When we treat problems of molecular evolution, however, the time until fixation is more important than that until extinction. It was shown in our previous paper (KIMURA and OHTA 1969) that it takes about  $4N_e$  generations for an individual mutant to reach fixation (excluding the cases of extinction), if it is selectively neutral. We show in the APPENDIX of this paper that the variance of the number of generations until fixation is approximately

$$\text{Var}(t_1) \approx 4.58 N_e^2 \quad (13)$$

so that the standard deviation is about  $(2.14)N_e$ . Thus, the standard deviation is roughly half of the mean.

We would like to thank Dr. ROBERT H. MAC ARTHUR and Dr. HENRY E. SCHAFER for reading the manuscript and correcting the English.

TABLE 3

*Frequency distribution of the number of generations until extinction of a single neutral mutant gene that appeared in a population of  $N_e = N = 200$*

Number of generations	Frequency (percent) (based on 200 replicates)
1	42
2	12.5
3	11.5
4	5
5	2.5
6	3
7	1.5
8	3
9	1.5
10	0.5
11-20	7
21-30	2
31-40	1.5
41-50	1.5
51-60	0
61-70	0.5
71-80	0
81-90	1
91-100	1
161-170	0.5
201-210	1
241-250	0.5
521-530	0.5
	100.0

## SUMMARY

The average number of generations until extinction of an individual mutant gene in a finite population (actual size  $N$ , effective size  $N_e$ ) was studied using diffusion models; approximation formulas were given for neutral, semi-dominant and recessive deleterious mutations. Monte Carlo experiments were performed to check the validity of these formulas.—In addition, the variance of the time until extinction was studied for neutral mutations. It was shown that for such mutants, the standard deviation of the number of generations until extinction (excluding the cases of fixation) is roughly  $4N_e/\sqrt{N}$ , and this is much larger than the mean which is  $2(N_e/N)\log_e 2N$ . On the other hand, the number of generations until fixation of a neutral mutant (excluding the cases of extinction) has a mean of approximately  $4N_e$  and a standard deviation of roughly  $2N_e$ .

## LITERATURE CITED

- Crow, J. F., 1969 Molecular genetics and population genetics. Proc. 12th Intern. Congr. Genet. **3**: 105-113.
- Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. The Clarendon Press, Oxford.

KIMURA, M., 1964 Diffusion models in population genetics. *J. Appl. Probab.* **1**: 177-232. —, 1968 Evolutionary rate at the molecular level. *Nature* **217**: 624-626. —, 1969 The rate of molecular evolution considered from the standpoint of population genetics. *Proc. Natl. Acad. Sci. U.S.* **63**: 1181-1188.

KIMURA, M. and T. OHTA, 1969 The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**: 763-771.

KING, J. L. and T. H. JUKES, 1969 Non-Darwinian evolution. *Science* **164**: 788-798.

## APPENDIX

*Variance of the number of generations until extinction, with special reference to selectively neutral mutations*

Let  $u_0(p, t)$  be the probability that a mutant allele becomes lost from the population by the  $t$ th generation, given that its initial frequency is  $p$ . Also, let

$$T_0(p) = \int_0^{\infty} t \{ \partial u_0(p, t) / \partial t \} dt$$

and

$$S_0(p) = \int_0^{\infty} t^2 \{ \partial u_0(p, t) / \partial t \} dt.$$

Then

$$\bar{t}_0(p) = T_0(p) / u_0(p)$$

and

$$\overline{t_0^2(p)} = S_0(p) / u_0(p) \tag{A1}$$

respectively, represent the mean and the mean square of the number of generations until extinction of a mutant allele whose initial frequency is  $p$ . In the above formulas,  $u_0(p) \equiv u_0(p, \infty)$  stands for the probability of eventual loss.

For a selectively neutral allele, it can be shown that  $u_0(p, t)$  satisfies the following partial differential equation

$$\frac{\partial u}{\partial t} = \frac{p(1-p)}{4N_e} \frac{\partial^2 u}{\partial p^2}, \tag{A2}$$

where  $N_e$  is the "variance" effective number of the population.

Differentiating both sides of this equation with respect to  $t$ , multiplying the resulting terms by  $t^2$ , and then integrating both sides with respect to  $t$  from 0 to  $\infty$ , we obtain

$$\int_0^{\infty} t^2 \frac{\partial^2 u}{\partial t^2} dt = \frac{p(1-p)}{4N_e} \frac{\partial^2}{\partial p^2} S_0(p).$$

Assuming that  $t^2 \partial u(p, t) / \partial t$  vanishes at  $t = \infty$ , the left hand side of this equation is reduced to  $-2T_0(p)$ , which is equal to  $8N_e p \log_e p$ , according to KIMURA and OHTA (1969). Thus we have

$$\frac{d^2}{dp^2} S_0(p) = (32 N_e^2) \frac{\log p}{1-p} \tag{A3}$$

The solution of this differential equation which satisfies the boundary conditions,

$$S(0) = 0 \tag{A4}$$

and

$$\lim_{p \rightarrow 1} \overline{t_0^2(p)} = \text{finite}, \tag{A5}$$

turns out to be as follows:

$$S_0(p) = 32 N_e^2 \left\{ p \log_e p - (1-p) \int_0^p \frac{\log x}{1-x} dx \right\}. \tag{A6}$$

(For the meaning of these boundary conditions, readers may refer to KIMURA and OHTA 1969). Thus the mean square time until extinction is

$$\overline{t_0^2(p)} = 32 N_e^2 \left[ \frac{p \log p}{1-p} - \int_0^p \frac{\log x}{1-x} dx \right], \tag{A7}$$

since  $u_0(p) = 1-p$  for a neutral allele.



The variance of the number of generations until extinction may then be obtained from

$$\text{Var} \{t_0(p)\} = \overline{t_0^2(p)} - \{\bar{t}_0(p)\}^2,$$

in which  $\bar{t}_0(p) = -4N_e(p \log_e p)/(1-p)$ .

If the mutant allele is represented only once at the moment of its appearance, we may put

$$p = 1/(2N)$$

in the above formulas. Writing  $t_0$  for  $t_0(1/2N)$ , we obtain

$$\bar{t}_0 = 2(N_e/N) \log_e 2N \tag{A8}$$

and

$$\overline{t_0^2} = 16 N_e^2/N \tag{A9}$$

approximately.

Formula (A7) also gives the mean square time until fixation of a neutral allele, if we note that

$$t_0(1-p) = t_1(p),$$

i.e., loss of one allele is the same event as fixation of the alternative allele, where  $t_1(p)$  is the time until fixation of an allele whose initial frequency is  $p$ .

Since

$$\int_0^1 \frac{\log x}{1-x} dx = -\frac{\pi^2}{6}$$

and

$$\lim_{\epsilon \rightarrow 0} \bar{t}_0(1-\epsilon) = 4N_e, \tag{A10}$$

we obtain

$$\lim_{\epsilon \rightarrow 0} \overline{t_0^2(1-\epsilon)} = 32 N_e^2 \left(\frac{\pi^2}{6} - 1\right), \tag{A11}$$

and

$$\text{Var}\{t_1\} = \overline{t_0^2(1)} - \bar{t}_0^2(1) \approx 4.58 N_e^2, \tag{A12}$$

where  $t_1$  is the number of generations until fixation (excluding the cases of extinction) of an individual mutant gene.