

# THE AVERAGE NUMBER OF GENERATIONS UNTIL FIXATION OF A MUTANT GENE IN A FINITE POPULATION<sup>1</sup>

MOTOO KIMURA AND TOMOKO OHTA<sup>2</sup>

*National Institute of Genetics, Mishima, Japan*

Received July 26, 1968

A mutant gene which appeared in a finite population will eventually either be lost from the population or fixed (established) in it. The mean time until either of these alternative events takes place was studied by WATTERSON (1962) and EWENS (1963). They made use of a method previously announced by DARLING and SIEGERT (1953), and, independently by FELLER (1954). Actually, DARLING and SIEGERT refer to its application to genetics.

From the standpoint of population genetics, however, it is much more desirable to determine separately the mean time until fixation and that until loss. Since the gene substitution in a population plays a key role in the evolution of the species, it may be of particular interest to know the mean time for a rare mutant gene to become fixed in a finite population, excluding the cases in which such a gene is lost from the population.

In the present paper, a solution to this problem will be presented together with Monte Carlo experiments to test some of the theoretical results. Throughout this paper, the senior author (M. K.) is responsible for the mathematical treatments, while the junior author (T. O.) is responsible for the numerical treatments based on computers.

## BASIC THEORY

Let us consider a diploid population consisting of  $N$  individuals and having the variance effective number  $N_e$ , which may be different from the actual number (for the definition of  $N_e$ , see KIMURA and CROW 1963). Throughout this paper, we will denote by  $p$  the frequency of a mutant gene ( $A_2$ ), so that  $1-p$  represents the frequency of its allele ( $A_1$ ). Also, we will use the diffusion models (cf. KIMURA 1964) to solve the problem. Let  $u(p, t)$  be the probability that the mutant allele  $A_2$  becomes fixed (i.e., its frequency becomes unity) by the  $t$ th generation, given that its frequency is  $p$  at the start (i.e., at  $t = 0$ ). If we define a quantity  $T_1(p)$  by the relation

$$T_1(p) = \int_0^{\infty} t \frac{\partial u(p, t)}{\partial t} dt, \quad (1)$$

then

$$\bar{t}_1(p) = T_1(p)/u(p) \quad (2)$$

<sup>1</sup> Contribution No. 692 from the National Institute of Genetics, Mishima, Shizuoka-ken, Japan.

<sup>2</sup> Supported by a post-doctoral fellowship from the Japan Soc. for the Promotion of Science.

represents the average number of generations until the mutant allele with initial frequency  $p$  becomes fixed in the population, excluding the cases in which the allele is lost from it. In the above expression,  $u(p)$  stands for the probability of ultimate fixation (KIMURA 1957, 1962), such that

$$u(p) = \lim_{t \rightarrow \infty} u(p, t). \tag{3}$$

If we denote by  $M_{\delta_p}$  and  $V_{\delta_p}$  the mean and the variance of the rate of change in the frequency of  $A_2$  per generation, then as shown by KIMURA (1962),  $u(p, t)$  satisfies the following partial differential equation

$$\frac{\partial u(p, t)}{\partial t} = \frac{1}{2} V_{\delta_p} \frac{\partial^2 u(p, t)}{\partial p^2} + M_{\delta_p} \frac{\partial u(p, t)}{\partial p} \tag{4}$$

Here we assume that the process of change in gene frequency is time homogeneous, that is, both  $M_{\delta_p}$  and  $V_{\delta_p}$  do not depend on time parameter  $t$ .

Differentiating each term of the above equation (4) with respect to  $t$ , multiplying each of the resulting terms by  $t$ , followed by integrating them with respect to  $t$  from 0 to  $\infty$ , we obtain

$$\int_0^\infty t \frac{\partial^2 u(p, t)}{\partial t^2} dt = \frac{1}{2} V_{\delta_p} \frac{\partial^2}{\partial p^2} T_1(p) + M_{\delta_p} \frac{\partial}{\partial p} T_1(p)$$

The left hand side of this equation is reduced to  $-u(p)$ , since

$$\begin{aligned} \int_0^\infty t \frac{\partial^2 u(p, t)}{\partial t^2} dt &= \left[ t \frac{\partial u(p, t)}{\partial t} \right]_0^\infty \\ &\quad - \int_0^\infty \frac{\partial u(p, t)}{\partial t} dt \\ &= -u(p, \infty), \end{aligned}$$

in which we assume that  $t \partial u(p, t) / \partial t$  vanishes at  $t = \infty$ .

Thus, we have the following ordinary differential equation for  $T_1(p)$ ,

$$T_1''(p) + a(p) T_1'(p) + b(p) = 0, \tag{5}$$

where  $a(p) = 2M_{\delta_p} / V_{\delta_p}$  and  $b(p) = 2u(p) / V_{\delta_p}$

The above equation can be integrated immediately, and if we determine the two constants involved by the following two boundary conditions,

$$\lim_{p \rightarrow 0} \bar{t}_1(p) = \text{finite} \tag{6}$$

and

$$\bar{t}_1(1) = 0, \tag{7}$$

we obtain

$$\begin{aligned} T_1(p) &= u(p) \int_p^1 \psi(\xi) u(\xi) \{1 - u(\xi)\} d\xi \\ &\quad + \{1 - u(p)\} \int_0^p \psi(\xi) u^2(\xi) d\xi. \end{aligned} \tag{8}$$

In the above formula,

$$u(p) = \int_0^p G(x)dx / \int_0^1 G(x)dx, \quad (\text{KIMURA 1962}), \quad (9)$$

is the probability of ultimate fixation, and  $\psi(x)$  is given by

$$\psi(x) = 2 \int_0^1 G(x)dx / \{V_{\delta x} G(x)\}, \quad (10)$$

where

$$G(x) = \exp \left\{ - \int_0^x \frac{2M_{\delta\xi}}{V_{\delta\xi}} d\xi \right\}, \quad (11)$$

in which  $\exp \{ \cdot \}$  stands for the exponential function.

Of the two boundary conditions, the first, i.e. (6), may need some comments. It reflects the fact that in a finite population a single mutant gene which appeared in the population reaches fixation within a finite time. It is also equivalent to the relation  $\lim_{p \rightarrow 0} T_1(p) = Kp$ , in which  $K$  is a constant.

From (8) and (2), the required solution for our problem is

$$\bar{t}_1(p) = \int_p^1 \psi(\xi)u(\xi)\{1-u(\xi)\}d\xi + \frac{1-u(p)}{u(p)} \int_0^p \psi(\xi)u^2(\xi)d\xi, \quad (12)$$

where  $u$  and  $\psi$  are given respectively by (9) and (10). Similarly, we can derive the average number of generations until the mutant gene is lost from the population, excluding the cases in which the mutant gene is ultimately fixed in the population. This is given by

$$\begin{aligned} \bar{t}_0(p) &= \frac{u(p)}{1-u(p)} \int_p^1 \psi(\xi)\{1-u(\xi)\}^2 d\xi \\ &+ \int_0^p \psi(\xi)\{1-u(\xi)\}u(\xi)d\xi \end{aligned} \quad (13)$$

SOME SPECIAL CASES

In this section we assume that the factor causing random fluctuation in gene frequency is the random sampling of gametes alone so that

$$V_{\delta p} = p(1-p)/(2N_e),$$

where  $N_e$  is the variance effective number.

In the simplest case of neutral mutations, we have

$$M_{\delta p} = 0.$$

Thus from (12), the average number of generations until fixation (excluding the cases of loss) is

$$\bar{t}_1(p) = -\frac{1}{p} \{4N_e(1-p)\log_e(1-p)\}. \quad (14)$$

At the limit of  $p \rightarrow 0$ , we have

$$\bar{t}_1(0) = 4N_e. \quad (15)$$

This shows that an originally rare mutant gene in a population of effective size  $N_e$  takes about  $4N_e$  generations until it spreads to the whole population if we disregard the cases in which such a gene is eventually lost from the population. Similarly, from (13), the number of generations until loss (excluding the cases of fixation) for neutral mutations is

$$\bar{t}_0(p) = -4N_e \left( \frac{p}{1-p} \right) \log_e p \quad (16)$$

If the mutant gene  $A_2$  has the selective advantage  $s/2$  over its allele  $A_1$  (case of genic selection) such that

$$M_{\delta p} = \frac{s}{2} p(1-p),$$

then, writing  $N_e s = S$ , we have, from (12)

$$\bar{t}_1(p) = J_1 + \frac{1-u(p)}{u(p)} J_2 \quad (17)$$

where

$$J_1 = \frac{2}{s(1-e^{-2S})} \int_p^1 \frac{(e^{2S\xi}-1)(e^{-2S\xi}-e^{-2S})}{\xi(1-\xi)} d\xi,$$

$$u(x) = (1-e^{-2Sx})/(1-e^{-2S})$$

and

$$J_2 = \frac{2}{s(1-e^{-2S})} \int_0^p \frac{(e^{2S\xi}-1)(1-e^{-2S\xi})}{\xi(1-\xi)} d\xi.$$

A more general case of genotypic selection can also be worked out in a similar way using equation (12).

#### NUMERICAL EXAMPLES AND MONTE CARLO EXPERIMENTS

The numerical evaluation of some of the results of the foregoing sections together with Monte Carlo experiments were performed using computers TOSBAC 3400 and IBM 360. The only case which can be evaluated easily from the formula is the one of selectively neutral mutations (formula 14). For other cases, one needs numerical integration. In this section, the numerical examples for the cases of no selection, genic selection (no dominance) and overdominance will be given.

The Monte Carlo experiments were performed for the cases of no selection and genic selection by the following scheme. In each generation, the change of gene frequency by genic selection was carried out deterministically using the formula

$$\Delta p = \frac{s}{2} p(1-p)/(1+sp), \quad (18)$$

where  $s$  is the selection coefficient for the mutant homozygote. Sampling of zygotes was performed by generating pseudo-random numbers,  $X$  ( $0 < X < 1$ ), using the subroutines RAND in TOSBAC 3400 and RANDU in IBM 360. Each experiment was continued until fixation or extinction and the number of generations involved was recorded. 400 replicate trials were done for each set of parameters.

Figure 1 shows the comparison of the results of Monte Carlo simulation (dots) and those of analytical solution (curves) for the cases of  $N_e=10$  (upper curve) and  $N_e=5$  (lower curve). The abscissa represents the initial frequency  $p$ . As  $p$

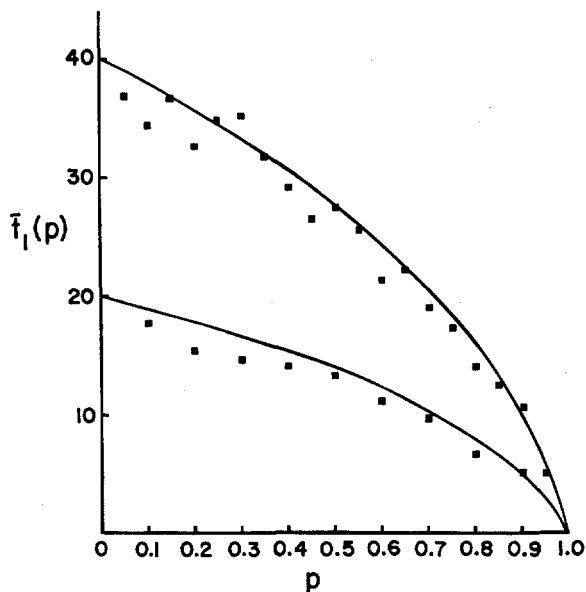


FIGURE 1.—Average number of generations until fixation of a selectively neutral mutant gene as a function of its initial frequency. The theoretical values are represented by curves and those of Monte Carlo simulation by dots.  $2N_e=20$  in the upper curve and  $2N_e=10$  in the lower one.

changes from 0 to 1, the number of generations until fixation changes from 40 to 0 for the case of  $N_e=10$  and from 20 to 0 in the case of  $N_e=5$ . As it is clear from the figure, the agreement between the Monte Carlo results and the theoretical predictions is satisfactory, although the latter seems to overestimate the true value slightly.

The results for the case of genic selection are given in Figure 2 with a constant initial frequency ( $p=0.1$ ) and a varying selection intensity. The curve represents the theoretical results by numerical integration, and the dots represent the results of Monte Carlo experiments. The numerical integration was performed by Simpson's rule using the computer IBM 360. Again, the agreement between the two is satisfactory, though the theoretical treatment seems to underestimate the actual value slightly. The slight discrepancy may be due to the existence of the denominator in formula (18) that effectively decreases the value of  $s$  in the numerator as compared with the expression  $(s/2)p(1-p)$  used in the theoretical treatment. As expected, the selective advantage accelerates the fixation of the advantageous allele. So, it may be of some interest to compare such acceleration with the rate of steady decay (KIMURA 1955). It can be shown that, as  $N_e s$  increases, the inverse of the rate of steady decay decreases more rapidly as compared with the shortening of the fixation time (with  $p=0.1$ ). For example, when  $N_e s=5$ , the time until fixation is about half of that of the neutral case, while the rate of steady decay is about 3.5 as large.

In the case of overdominance between a pair of alleles  $A_1$  and  $A_2$ , the formula

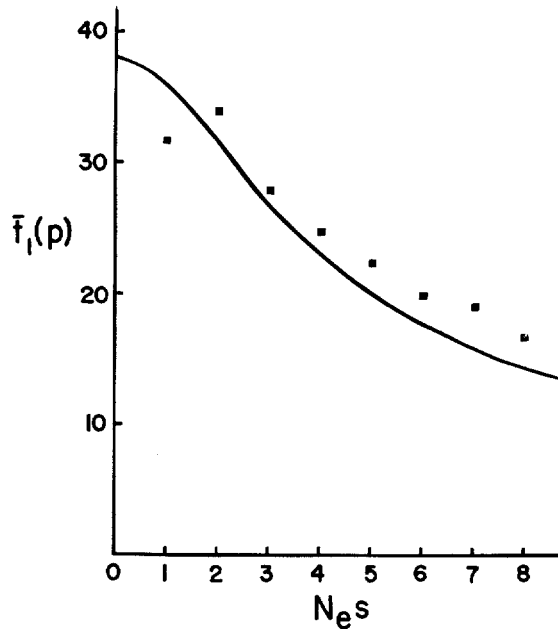


FIGURE 2.—The relationship between time for fixation and selective advantage in the case of genic selection (no dominance). The curve represents the theoretical results by numerical integration and the dots represent the results of Monte Carlo simulation. In this figure, the effective population number  $N_e=10$  and the initial frequency  $p=0.1$ .

(12) contains double integrals. Though no simulation experiments were performed in this case, we studied, using formula (12), the theoretical relationship between the time until fixation and the initial frequency for various values of  $N_e s$  assuming that the fitnesses of the three genotypes  $A_1A_1$ ,  $A_1A_2$  and  $A_2A_2$  are  $1-s$ , 1 and  $1-s$ , respectively. In such a case, the equilibrium gene frequency is 0.5 and the overdominance is most effective for retarding gene fixation (ROBERTSON 1962). Now, the time until fixation is the function of  $N_e$  and  $N_e s$ . So, in Figure 3, the results for  $N_e s = -2, -1, 1, 2$  and 4 are illustrated in comparison with the selectively neutral case ( $N_e s = 0$ ). From the figure, it may be seen, for example, that the overdominance prolongs the time until fixation almost twice as compared with the selectively neutral case when  $N_e s$  is about 2. Let us compare the present results with the rate of steady decay. Again, the increase of fixation time for larger  $N_e s$  is slower as compared with the increase of the inverse of the rate of steady decay which is equivalent to the retardation factor of ROBERTSON. For instance, when  $N_e s = 2$ , the retardation factor is about  $1/0.4 = 2.5$  (MILLER 1962). However, the disagreement is not as large as in the case of the genic selection.

#### DISCUSSION

It now appears (KIMURA 1968) that mutation and random genetic drift play a more important role in determining the genetic structure of Mendelian populations than previously considered, especially when molecular mutations are

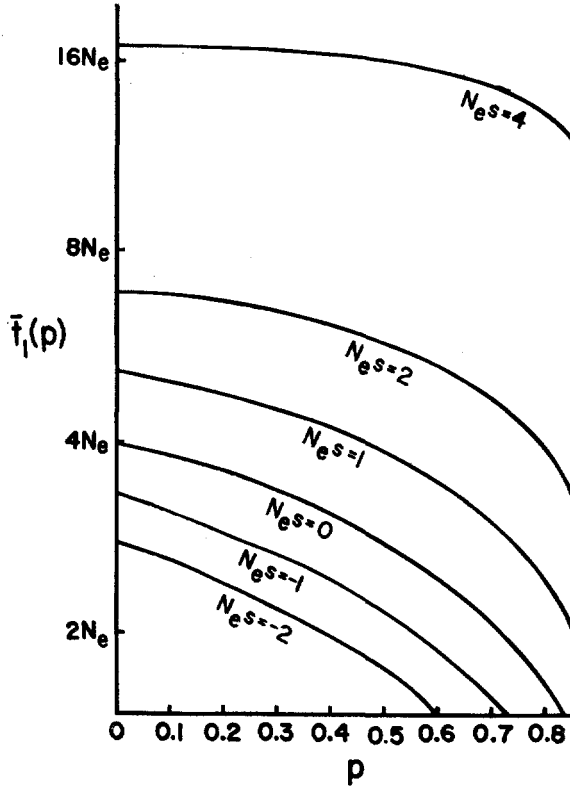


FIGURE 3.—Time for fixation of an overdominant mutant gene as a function of its initial frequency with various intensities of selection. Only the theoretical results obtained by numerical integration are shown. No Monte Carlo experiments were performed. The selectively neutral case is also shown for comparison. The ordinate gives the time for fixation by the logarithmic scale.

taken into account. Furthermore, the majority of such mutations appear to be almost neutral for natural selection. Thus the results obtained in the previous sections assuming  $M_{\delta p} = 0$  should be applicable to such mutations.

Now, in a population consisting of  $N$  individuals, if we assume that each mutant gene is represented only once at the moment of its occurrence,  $p = 1/(2N)$ , and from formula (14), the average number of generations until fixation of a neutral mutation becomes

$$\bar{t}_1\left(\frac{1}{2N}\right) = -8NN_e\left(1 - \frac{1}{2N}\right) \log_e\left(1 - \frac{1}{2N}\right) \tag{19}$$

If  $N$  is large, this is very close to  $4N_e$ , the value given as  $\bar{t}_1(0)$  in (15). Since the condition (6), i.e.  $\bar{t}_1(0) = \text{finite}$  is a crucial one in deriving the formula (8) and therefore formula (12), it may be of some interest to examine the value of  $\bar{t}_1(0)$  more in detail. It was shown by KIMURA (1955) that for a neutral mutation

$$u(p,t) = p + \sum_{i=1}^{\infty} (-1)^i \frac{2i+1}{2i(i+1)} (1-r^2) T_{i-1}^{-1}(r) e^{-\lambda_i t}, \tag{20}$$

where  $r=1-2p$ ,  $\lambda_i=i(i+1)/(4N_e)$  and  $T_{i-1}^{-1}(r)$  represents the Gegenbauer polynomial. The above formula enables us to calculate  $\bar{t}_1(p)$  directly from (1) and (2), giving

$$\bar{t}_1(p)=4N_e(1+r)\sum_{i=1}^{\infty}(-1)^{i-1}\frac{2i+1}{i^2(i+1)^2}T_{i-1}^{-1}(r). \quad (21)$$

At the limit of  $p \rightarrow 0$  ( $r \rightarrow 1$ ), if we use the relation  $T_{i-1}^{-1}(1)=i(i+1)/2$ , we obtain

$$\bar{t}_1(0)=4N_e\sum_{i=1}^{\infty}(-1)^{i-1}\frac{2i+1}{i(i+1)}=4N_e, \quad (22)$$

thus confirming the result given in (15). The above results show that a single mutant gene, if it is neutral, takes about  $4N_e$  generations until fixation if we disregard the cases in which it is eventually lost. In this connection, it is interesting to note FISHER'S (1930) inference that in the absence of favorable selection the number of individuals having a gene derived from a single mutation cannot greatly exceed the number of generations since its occurrence.

Next, let us consider the number of generations until a neutral mutant gene is lost from the population disregarding the cases in which it is eventually fixed. This is given by formula (16) by putting  $p=1/(2N)$ . Namely,

$$\bar{t}_0\left(\frac{1}{2N}\right)=\frac{4N_e}{2N-1}\log_e(2N)\approx 2\left(\frac{N_e}{N}\right)\log_e(2N) \quad (23)$$

Since the ratio  $N_e/N$  is around 0.8 in man (CROW 1954), a single mutant gene which appeared in a human population will be lost from the population on the average in about  $1.6\log_e 2N$  generations. If  $N=10^4$ , this amounts to about 16 generations. These results show that a great majority (fraction  $1-\frac{1}{2N}$ ) of neutral or nearly neutral mutant genes which appeared in a finite population are lost from the population within a few generations, while the remaining minority (fraction  $\frac{1}{2N}$ ) spread over the entire population (i.e. reach fixation) taking a very large number of generations.

In the present paper, we have studied the average (i.e., the first moment) of the length of time until fixation (and, separately, until loss), but the present method can immediately be adapted to obtain the  $n$ th moment of the length of time until fixation in terms of  $(n-1)$ th moment, thus enabling us to obtain the higher moments step by step starting from the first moment.

#### SUMMARY

In a finite population, a mutant gene is either fixed in the population or lost from it within a finite length of time. A theory was presented which enables us to obtain the average number of generations until fixation, and separately, that until loss, based on the method of diffusion equations. Also, Monte Carlo experiments were performed to test some of the theoretical results.—It was shown that a single mutant gene, if it is selectively neutral, takes about  $4N_e$  generations until fixation in a population of effective size  $N_e$ .



## LITERATURE CITED

- CROW, J. F., 1954 Breeding structure of populations. II. Effective population number. pp. 543-556. In: *Statistics and Mathematics in Biology*. Iowa State College Press, Ames, Iowa.
- DARLING, D. A., and A. J. F. SIEGERT, 1953 The first passage problem for a continuous Markov process. *Ann. Math. Statist.* **24**: 624-639.
- EWENS, W. J., 1963 The mean time for absorption in a process of genetic type. *J. Australian Math. Soc.* **3**: 375-383.
- FELLER, W., 1954 Diffusion processes in one dimension. *Trans. Am. Math. Soc.* **77**: 1-31.
- FISHER, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- KIMURA, M., 1955 Solution of a process of random genetic drift with a continuous model. *Proc. Natl. Acad. Sci. U.S.A.* **41**: 144-150. ——— 1957 Some problems of stochastic processes in genetics. *Ann. Math. Statist.* **28**: 882-901. ——— 1962 On the probability of fixation of mutant genes in a population. *Genetics* **47**: 713-719. ——— 1964 Diffusion models in population genetics. *J. Appl. Probability* **1**: 177-232. ——— 1968 Evolutionary rate at the molecular level. *Nature* **217**: 624-626.
- KIMURA, M., and J. F. CROW, 1963 The measurement of effective population number. *Evolution* **17**: 279-288.
- MILLER, G. G., 1962 The evaluation of eigenvalues of a differential equation arising in a problem in genetics. *Proc. Cambridge Phil. Soc.* **58**: 588-593.
- ROBERTSON, A., 1962 Selection for heterozygotes in small populations. *Genetics* **47**: 1291-1300.
- WATTERSON, G. A., 1962 Some theoretical aspects of diffusion theory in population genetics. *Ann. Math. Statist.* **33**: 939-957.