Proc. Nat. Acad. Sci. USA
Vol. 71, No. 9, pp. 3377-3379, September 1974

Probability of Gene Fixation in an Expanding Finite Population

(population genetics/diffusion model/fixation probability of mutant/logistic population)

MOTOO KIMURA AND TOMOKO OHTA

National Institute of Genetics, Mishima, Japan

Contributed by Motoo Kimura, June 13, 1974

ABSTRACT **A mathematical theory was developed, based on diffusion models, that enables us to compute the probability of a rare mutant allele eventually spreading through a population when the population size changes with time. In particular, we elaborated the case in which the mutant allele has a definite selective advantage and the population expands following the logistic law. In this case, the probability of ultimate fixation of a single mutant is given by $u = 2s(Z/N)$, where s is the selective advantage and Z/N is a factor by which the probability of fixation is modified through population expansion. Analytical expression was obtained for Z/N, and the validity of the formula for u was checked by Monte Carlo experiments.**

From the standpoint of population genetics, the process of evolution consists of a series of gene substitutions within a species. It is clear, therefore, that the probability of gene fixation, or the probability of an individual mutant allele spreading through the population (reaching 100% in frequency), is essential for our evaluation of the rate of evolution. Following the pioneering works by Haldane (1), Fisher (2), and Wright (3), a general formula for the probability of ultimate fixation of a mutant allele in a finite population was obtained by one of us (4, 5), assuming a constant population size and constant (i.e., time-independent) selection coefficients. More recently, we have obtained (6), based on diffusion models (see ref. 7), the probability of fixation of a mutant gene in a finite population when its selective advantage decreases exponentially with time.

The purpose of the present paper is to present a new theory that enables us to compute the probability by which a rare mutant allele eventually becomes fixed in a finite population when the population size changes with time. In particular, we shall elaborate the case in which the population expands under the logistic law and the mutant allele in it has a definite selective advantage. The following treatments involve approximations that are valid when the initial frequency of the mutant allele is low.

BASIC THEORY

Consider a diploid population and denote by N_t its effective size at time t conveniently measured with one generation as the unit length of time. Let us assume that the population size N_t changes *deterministically* so that we can write

$$\frac{dN_t}{dt} = f(N_t). \qquad [1]$$

Note that for logistic population growth, $f(\cdot)$ is given by

$$f(N) = \frac{rN(K - N)}{K}, \qquad [2]$$

where r is the intrinsic rate of population increase and K is the carrying capacity (corresponding to N_∞ in Eq. 1).

Let $u(p,t)$ be the probability of ultimate fixation of a mutant allele, given that it appeared at time t with initial frequency p. Suppose that the mutant has selective advantage s, which is constant with time. Then, it can be shown that $u = u(p,t)$ satisfies the following diffusion equation, which is a time nonhomogeneous form of the Kolmogorov backward equation (see ref. 8, p. 171):

$$-\frac{\partial u}{\partial t} = \frac{p(1 - p)}{4N_t}\frac{\partial^2 u}{\partial p^2} + sp(1 - p)\frac{\partial u}{\partial p}. \qquad [3]$$

Since we have assumed that the population size changes deterministically with time, we can regard t as a function of N_t provided that the population size changes *monotonically* with time. Noting that

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial N_t}\frac{\partial N_t}{\partial t} = \frac{\partial u}{\partial N_t}f(N_t)$$

and writing N for N_t, we have, from Eq. 3,

$$-f(N)\frac{\partial u}{\partial N} = \frac{p(1 - p)}{4N}\frac{\partial^2 u}{\partial p^2} + sp(1 - p)\frac{\partial u}{\partial p}. \qquad [4]$$

In this equation, u is regarded as a function of p and N. So, in the following treatment we shall denote the probability of fixation by $u(p,N)$. In other words, u in Eq. 4 stands for the probability of ultimate fixation of a mutant allele that appeared with initial frequency p when the population size is N.

In order to solve Eq. 4, and especially to obtain an approximate solution that is valid for a small p, we try a solution of the form

$$u(p,N) = \frac{1 - e^{-4Zsp}}{1 - e^{-4Zs}}, \qquad [5]$$

where $Z = Z(N)$ is a function of N but independent of p. Note that expression 5 satisfies the necessary boundary conditions;

$$u(0,N) = 0, \text{ and } u(1,N) = 1.$$

Note also that it gives the exact solution for the time homogeneous case [$f(N) = 0$] by putting $Z = N$, since in this case we have

$$u(p,N) = \frac{1 - e^{-4Nsp}}{1 - e^{-4Ns}} \qquad [6]$$

(see ref. 4). In addition, for the neutral case ($s \to 0$), Eq. 5 reduces to the correct formula, i.e., $u(p,N) = p$. Assuming

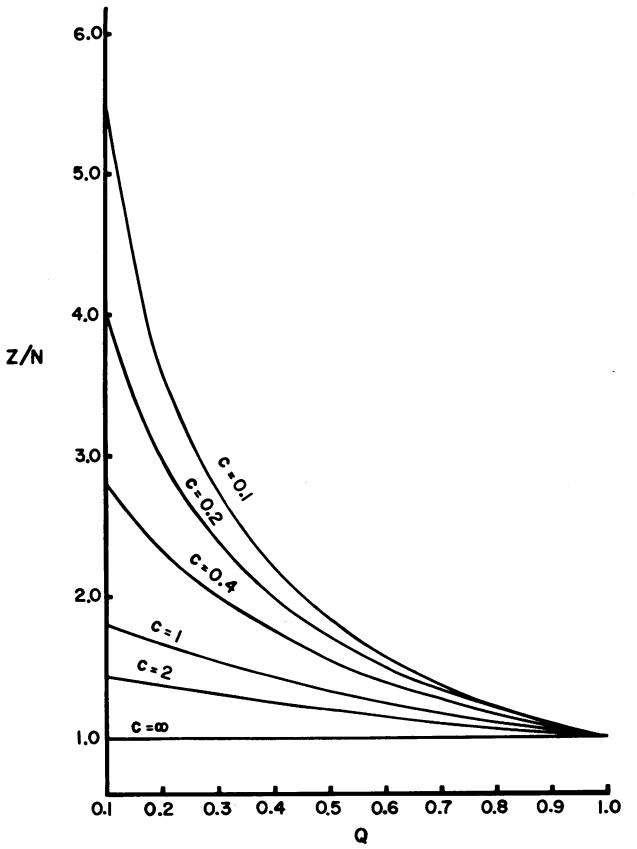

FIG. 1. Relationship between Z/N and Q for various values of c, computed from Eq. 15. Z/N represents the factor by which the probability of ultimate fixation is modified through population expansion and $Q = N/K$, where N is the population size when the mutant allele appeared in the population and K is the carrying capacity.

that p is much smaller than unity, we have, from Eq. **5**, the following approximations;

$$\frac{\partial u}{\partial N} = 4sp\frac{1 - (1 + 4Zs)e^{-4Zs}}{(1 - e^{-4Zs})^2}\frac{dZ}{dN},$$

$$\frac{\partial u}{\partial p} = \frac{4Zs}{1 - e^{-4Zs}},$$

and

$$\frac{\partial^2 u}{\partial p^2} = -\frac{(4Zs)^2}{1 - e^{-4Zs}}.$$

Substituting these in Eq. **4**, and also substituting p for $p(1 - p)$ in the equation (since p is assumed to be very small), we obtain the following ordinary differential equation for Z.

$$\frac{dZ}{dN} = -\frac{(1 - e^{-4Zs})Zs(N - Z)}{1 - (1 + 4Zs)e^{-4Zs}}\cdot\frac{1}{Nf(N)} \quad [7]$$

This equation does not contain p and, therefore, it satisfies the original assumption that Z is a function of N but independent of p.

We now assume that the population expands following the logistic law: $dN/dt = f(N) = rN(K - N)/K$, so that the above equation becomes

$$\frac{dZ}{dN} = -\frac{(1 - e^{-4Zs})Zs(N - Z)K}{\{1 - (1 + 4Zs)e^{-4Zs}\}rN^2(K - N)}. \quad [8]$$

Also, we shall consider the situation in which the mutant allele has a definite selective advantage so that $4Zse^{-4Zs}$ as well as e^{-4Zs} are negligibly small as compared with unity. Since Z is expected to be larger than N in an expanding population, as we shall see later,

$$4Zse^{-4Zs} \ll 1 \quad [9]$$

is a realistic assumption for definitely advantageous mutations.

Under these assumptions, Eq. **8** becomes

$$\frac{dZ}{dN} = -\frac{Zs(N - Z)K}{rN^2(K - N)}. \quad [10]$$

Substituting $Z = NY$ in this equation, we obtain

$$\frac{dY}{dN} - \frac{K(1 + c) - N}{N(K - N)}Y + \frac{cK}{N(K - N)} = 0, \quad [11]$$

where

$$c = s/r \quad [12]$$

is the ratio between the selective advantage of the mutant and the intrinsic growth rate of the population. Differential Eq. **11** can readily be integrated, and, if we impose the boundary condition

$$\lim_{N\to K} Y = 1, \quad [13]$$

we obtain

$$Y = cKN^{1+c}(K - N)^{-c}\int_N^K \nu^{-2-c}(K - \nu)^{c-1}d\nu \quad [14]$$

as the pertinent solution for $Y = N/Z$ (assuming $K \geqq N$). Condition **13** is based on the consideration that if the population has already expanded to its maximum size coinciding to the carrying capacity, that is $N = K$, then the probability of ultimate fixation is given by Eq. **5** with $Z = K$; in other words, $Y = K/K = 1$.

From the way by which quantity Z is introduced, and especially by comparing expressions **5** and **6**, we note that Z serves as a single representative effective population size that is applicable throughout the process of gene fixation starting from the time when the population size is N until the process is over at $t = \infty$. Thus, the ratio Z/N will be of interest as a factor by which the probability of gene fixation is modified through subsequent population expansion (see also Eq. **17** below). From Eq. **14**, we have

$$\frac{Z}{N} = \frac{Q^{-1-c}(1 - Q)^c}{c\int_Q^1 q^{-2-c}(1 - q)^{c-1}dq}, \quad [15]$$

where $Q = N/K$ is the population size (N) expressed as a fraction of the carrying capacity.

The probability of ultimate fixation of a single mutant gene that has definite selective advantage and that appears when the population size is N can be obtained from Eq. **5** by putting $p = 1/(2N)$ in the numerator, while neglecting the term e^{-4Zs} in the denominator. Thus, we have

$$u = 1 - e^{-2s(Z/N)} \quad [16]$$

If, in addition, the selection coefficient s, itself, is much smaller than unity even if $4Ns$ is much larger than unity, the above

TABLE 1. *Results of Monte Carlo experiments on the probability of fixation of an advantageous mutant gene in an expanding population**

	$Q = 0.1$		$Q = 0.25$	
c	Monte Carlo	Theoretical	Monte Carlo	Theoretical
0.1	0.356 ± 0.015	0.356	0.222 ± 0.013	0.221
0.2	0.244 ± 0.014	0.274	0.169 ± 0.012	0.190
0.4	0.211 ± 0.013	0.202	0.152 ± 0.011	0.159
1.0	0.120 ± 0.010	0.135	0.100 ± 0.009	0.124

* The corresponding theoretical values obtained from Eq. 5 are also given for comparison.

formula for the probability of fixation reduces to

$$u = 2s\left(\frac{Z}{N}\right), \qquad [17]$$

where Z/N is given by Eq. 15. Fig. 1 illustrates (based on the numerical integration of the denominator of Eq. 15) the relationship between Z/N and Q for various values of $c = s/r$. Note that at the limit $c = \infty$ (corresponding to no population expansion $r = 0$) we have $Z/N = 1$, so that Eq. 17 reduces to $u = 2s$, which is the well-known result first obtained by Haldane (1) for a very large stationary population. Note also that in an interesting special case in which $c = 1$ or $s = r$, Eq. 15 reduces to

$$\frac{Z}{N} = \frac{2}{1+Q}. \qquad [18]$$

In this case the probability of fixation of a single mutant gene with definite selective advantage is given by

$$u = 4s/(1+Q) \qquad [19]$$

assuming that it appears when the population size (N) is KQ. In deriving formula 18 we assumed that the population is expanding so that N is not larger than K. However, it can be shown that Eq. 18 also holds if $N > K$ or $Q > 1$, that is, if the population size is larger than the carrying capacity when the mutant first appeared, and that the population subsequently decreases toward the ultimate size K.

MONTE CARLO EXPERIMENTS

In order to check the validity of the above analytical treatments, we performed extensive Monte Carlo experiments. In each experiment, the population size was allowed to increase following the logistic equation: $N_t = K/(1 + \beta e^{-rt})$, where $\beta = (K - N_0)/N_0$, and N_0 corresponds to N in the above treatments. At the start of each run, a single mutant gene was assumed. As soon as the mutant allele was lost or fixed in the population, a new run was started.

Each generation consists of selection and sampling. Selection was carried out deterministically. The sampling scheme was a simple one: if a uniform random number happens to be equal to or less than the frequency of the mutant allele, a gamete with the mutant allele was sampled, otherwise a gamete with the wild-type allele was sampled. Sampling of gametes was repeated until the required number of gametes was obtained to form the next generation. By use of such a scheme, two sets of experiments were conducted. In one set, the population size N_0 was 10% of the carrying capacity when a single mutant appeared in the population ($Q = N_0/K = 0.1$), where N_0 consisted of 100 gametes (or 50 diploid individuals). In another set, $N_0 = 0.25\,K$ or $Q = N_0/K = 0.25$, where N_0 consisted of 125 gametes. In all the experiments, selection coefficient $s = 0.04$ was assumed.

Table 1 lists experimental outcomes together with theoretical values computed from Eq. 5. Each experimental value is the average of 1000 runs, and the standard errors were computed from $\sqrt{P(1-P)/1000}$, where P is the proportion fixed. As seen from the table, satisfactory agreements were found between the experimental results and the corresponding theoretical predictions.

DISCUSSION

The present treatments offer a theoretical basis to quantify the prediction (2) that a mutant gene has a higher chance of avoiding extinction in an increasing population than in a declining population. In the present paper we have elaborated the case in which the mutant allele has a definite selective advantage and the total population number is regulated by the logistic law. In particular, when the selective advantage (s) and the intrinsic rate of population growth (r) are equal in magnitude ($c = 1$), a simple approximation formula, $u = 4s/(1 + Q)$, was obtained for the probability of ultimate fixation of a mutant gene that appeared when the population size (N) is the fraction Q of the carrying capacity (K).

For a mutant allele having a small selective advantage or disadvantage as compared with the intrinsic rate of population growth, assumption 9 is not valid, and more careful treatment of the differential equations will be required. In other words, Eq. 8 cannot be approximated by Eq. 10, so that it will be necessary to undertake numerical solution of Eq. 8 under the condition

$$\lim_{N \to K} Z = K. \qquad [20]$$

It is likely that a large fraction of mutations that can only be detected at the molecular level are of this type, and it will be desirable to undertake such analysis to elucidate their role in evolution.

We thank Dr. L. L. Cavalli-Sforza and Miss Diane Wagener for stimulating discussions. Contribution no. 997 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411, Japan.

1. Haldane, J. B. S. (1927) *Proc. Camb. Phil. Soc.* **23**, 838–844.
2. Fisher, R. A. (1930) *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford).
3. Wright, S. (1931) *Genetics* **16**, 97–159.
4. Kimura, M. (1957) *Ann. Math. Stat.* **28**, 882–901.
5. Kimura, M. (1962) *Genetics* **47**, 713–719.
6. Kimura, M. & Ohta, T. (1970) *Genetics* **65**, 525–534.
7. Kimura, M. (1964) *J. Appl. Probabil.* **1**, 177–232.
8. Kimura, M. & Ohta, T. (1971) *Theoretical Aspects of Population Genetics* (Princeton University Press, Princeton, N.J.).