

The age of an allele in a finite population*

By TAKEO MARUYAMA

National Institute of Genetics, Mishima, Japan

(Received 24 April 1973)

SUMMARY

The age of an allele segregating in a finite population may be defined in two ways. They are (1) the age of a mutant gene that has never reached fixation in the population, and (2) the age including any fixation period in the past. Theoretical expressions for these are derived on the assumption that every mutant is unique.

The history of mutant genes in finite populations, such as the fixation probability and the fixation time, has been extensively studied by numerous authors. These theories play an important role in population genetics and evolutionary theory (see Kimura & Ohta, 1971, for a review). This paper is to add a small piece of new knowledge to the theory, the age of an allele whose gene frequency is specified. This problem was first studied by Kimura & Ohta (1973). We assume that every mutant is new to the population. Suppose that the gene frequency of a mutant allele is y . Then there are two possibilities: (1) this allele has never reached fixation, and (2) it has reached fixation previously and the frequency decreased to y at present. We will treat two distinct situations. We first consider the average age under assumption (1) and then the age including both possibilities (1) and (2). We call the former 'the age before fixation' and the latter 'the age including fixation'. Kimura & Ohta (1973) have obtained the age before fixation for a neutral mutant under the assumption of low mutation rate. This paper uses the same method as that of Kimura and Ohta, and obtains general formulas for both definitions of age without restriction of mutation rate.

Let $V_{\delta x}$ and $M_{\delta x}$ be respectively the variance and the mean of the change in the gene frequency x in one generation, and let $p(t, x, y)$ be the transition probability that the gene frequency is x at time 0 and is y at time t . Then it is well known that the $p(t, x, y)$ satisfies

$$\frac{\partial p(t, x, y)}{\partial t} = Lp(t, x, y), \quad (1)$$

in which L is the operator

$$\frac{V_{\delta x}}{2} \frac{d^2}{dx^2} + M_{\delta x} \frac{d}{dx}.$$

The partial differential equation (1) is called the Kolmogorov backward equation. For fixed x and y , let

$$B(x, y) \equiv \int_0^{\infty} t p(t, x, y) dt.$$

* Contribution No. 960 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411 Japan.

$B(x, y)$ gives the average time it took to pass from x to y , times the probability of being at y . Applying the operator L to $B(x, y)$ and using (1), we have

$$LB(x, y) = \int_0^\infty tLp(t, x, y) dt = \int_0^\infty t \frac{\partial p(t, x, y)}{\partial t} dt$$

$$= tp(t, x, y)|_0^\infty - \int_0^\infty p(t, x, y) dt.$$

Note that $tp(t, x, y)|_0^\infty = 0 \quad \int_0^\infty p(t, x, y) dt \equiv \phi(x, y)$

is the sojourn time at frequency y , and $B(x, y)/\phi(x, y)$ is the average interval since the frequency was x . Thus

$$LB(x, y) = -\dot{\phi}(x, y). \tag{2}$$

Applying L to $\phi(x, y)$,

$$L\phi(x, y) = \int_0^\infty Lp(t, x, y) dt = \int_0^\infty \frac{\partial p(t, x, y)}{\partial t} dt = p(t, x, y)|_0^\infty$$

$$= -p(0, x, y) = -\delta(x - y), \tag{3}$$

where $\delta(\cdot)$ is Dirac's delta function. This differential equation can be integrated:

$$\phi(x, y) = -2H(\delta(\eta - y); x) + Ag(0, x), \tag{4}$$

in which

$$G(x) \equiv \exp\left\{-2 \int^x \frac{M_{\delta\xi}}{V_{\delta\xi}} d\xi\right\},$$

$$g(x, y) \equiv \int_x^y G(\xi) d\xi,$$

$$H(f(\eta); x) \equiv \int_0^x G(\xi) \int_0^\xi \frac{f(\eta)}{V_{\delta\eta} G(\eta)} d\eta d\xi$$

for an arbitrary function $f(\cdot)$, and A is a constant to be determined from the boundary condition on $\phi(x, y)$ at $x = 1$. Note that the other boundary condition at $x = 0$ is $\phi(0, y) = 0$, because no reverse mutation occurs and therefore if the gene frequency once becomes 0, it will never become non-zero again. The equation (3) and the solution (4) are essentially due to Wright (1938), Kimura (1964, 1969) and Ewens (1969). Using the $\phi(x, y)$ of (4), the equation (2) can be solved:

$$B(x, y) = -2H(\phi(\eta, y); x) + Ag(0, x), \tag{5}$$

in which A is constant to be determined from the boundary condition at $x = 1$. As in (4), the other boundary is $B(0, y) = 0$.

We can determine the constant A in (4) and (5). For the age before fixation, $\phi(1, y) = 0$ and $B(1, y) = 0$ for all $0 < y < 1$. This is because the process (or a sample path) is stopped as soon as it reaches the boundary, i.e. $x = 1$ is an absorbing state. In this case, assuming $u(x) = g(0, x)/g(0, 1)$ exists,

$$\left. \begin{aligned} \phi_1(x, y) &= \frac{2g(y, 1)u(x)}{V_{\delta y}G(y)} \quad \text{for } y > x \\ &= \frac{2}{V_{\delta y}G(y)} [g(y, 1)u(x) - g(y, x)] \quad \text{for } y < x \end{aligned} \right\} \tag{6}$$

and

$$B_1(x, y) = 2u(x)H(\phi(\eta, y); 1) - 2H(\phi(\eta, y); x),$$

where subscript 1 indicates 'before fixation'.

On the other hand, for the age including fixation, the process is considered even after it reaches fixation until it eventually reaches extinction. In this case, as the starting point x approaches 1, the fate of the gene frequency becomes the same as that starting $x = 1$. Therefore $\phi_x(1, y) = 0$ and $B_x(1, y) = 0$, where the subscript x indicates the derivative with respect to the first variable. With these boundary conditions,

$$\phi_2(x, y) = \left. \begin{aligned} & \frac{2g(0, x)}{V_{\delta y} G(y)} \quad \text{for } y > x \\ & = \frac{2g(0, y)}{V_{\delta y} G(y)} \quad \text{for } y < x \end{aligned} \right\} \quad (7)$$

and
$$B_2(x, y) = 2g(0, x) \int_0^1 \frac{\phi(\xi, y)}{V_{\delta \xi} G(\xi)} d\xi - 2H(\phi(\eta, y); x),$$

where subscript 2 indicates 'including fixation'. The value of function $\phi(x, y)$ is the total number of generations in which the gene frequency is y if it starts from x , and the $B(x, y)$ gives the average time it took to pass from x to y times the probability of being at y . Therefore if we consider many similar loci and assume that the population is at steady state, $A(1/2N, y) \equiv B(1/2N, y)/\phi(1/2N, y)$ gives the average age of alleles found in such a population.

Let N be the population size and u be mutation rate per gene per generation. Since every mutant is unique, the average number of mutant to occur in each generation is $2Nu$. Therefore $2Nu$ times $\phi_1(1/2N, y)$ or $\phi_2(1/2N, y)$ gives respectively the density of the gene frequency distribution at equilibrium, among those alleles which have never reached fixation and among those including any fixation period. If all alleles are selectively neutral and if $V_{\delta y} = x(1-x)/2N$ and $M_{\delta x} = -ux$ which are usually used,

$$\phi_1\left(\frac{1}{2N}, y\right) = \frac{4N}{yF} \left[1 - \left(1 - \frac{1}{2N}\right)^F \right] \approx \frac{4N}{yF} \frac{F}{2N} = \frac{2}{y},$$

where $F = 1 - 4Nu$. Therefore,

$$2Nu\phi_1\left(\frac{1}{2N}, y\right) \approx \frac{4Nu}{y}. \quad (8)$$

On the other hand, the distribution including fixation period is

$$\phi_2\left(\frac{1}{2N}, y\right) = \frac{4N \left[1 - \left(1 - \frac{1}{2N}\right)^F \right]}{Fy(1-y)^F} \approx \frac{2}{y(1-y)^F}$$

and

$$2Nu\phi_2\left(\frac{1}{2N}, y\right) \approx \frac{4Nu}{y(1-y)^F}. \quad (9)$$

It is interesting to compare formulae (8) and (9). The former is linearly related to mutation rate, whereas formula (9) is a non-linear function of $F = 1 - 4Nu$.

The ratio, $\phi_1(1/2N, y)/\phi_2(1/2N, y)$, gives the probability that if a mutant is known

to have frequency y , it has never been fixed. This probability for a neutral mutant is equal to

$$\frac{\phi_1(1/2N, y)}{\phi_2(1/2N, y)} = (1-y)^F,$$

provided $F = 1 - 4Nu > 0$, which is the condition that temporary fixation is possible. If $4Nu$ tends to 0, the probability becomes $(1-y)$. On the other hand, if $4Nu$ tends to 1, the probability becomes 1 for all y . The latter result is obvious, because as $4Nu$ becomes 1, fixation ceases to occur.

Returning to the main subject of this paper, we can obtain explicit formulae of the average age before fixation and including fixation for some simple cases. As a typical case, let $V_{\delta x} = x(1-x)/2N$ and $M_{\delta x} = -ux$, where N is the population size and u is mutation rate. Then, assuming $4Nu < 1$ and ignoring a term of order $1/2N$, the age before fixation of a neutral mutant allele whose frequency is y , is

$$A_1\left(\frac{1}{2N}, y\right) = \frac{4N}{F} \left[\int_0^y \frac{\{1 - (1-\xi)^F\}}{\xi} d\xi + \frac{\{1 - (1-y)^F\}}{(1-y)^F} \int_y^1 \frac{(1-\xi)^F}{\xi} d\xi \right], \quad (10)$$

where $F = 1 - 4Nu$, and, as $4Nu$ approaches 0, the above formula becomes

$$A_1\left(\frac{1}{2N}, y\right) = \frac{-4Ny \log y}{1-y}. \quad (11)$$

This is the average age of a neutral mutant gene whose frequency is y , provided that it has never reached fixation and that every mutant arises at a homallelic locus, i.e. $4Nu \ll 1$. Formula (10) is the same as (13a) of Kimura & Ohta (1973). However, since we usually do not know whether a mutant gene has previously reached fixation or not, the age including fixation is more important. The average age including fixation of a neutral gene is

$$A_2\left(\frac{1}{2N}, y\right) = \frac{4N}{F} \left[\int_0^y \frac{\{1 - (1-\xi)^F\}}{\xi(1-\xi)^F} d\xi + \{1 - (1-y)^F\} \int_y^1 \frac{d\xi}{\xi(1-\xi)^F} \right], \quad (12)$$

where $F = 1 - 4Nu$.

The validity of formulae (11) and (12) were verified by computer simulation and results agreed well with the theoretical expectations. We should first note that both functions (10) and (12) are monotone, increasing so that the higher the frequency of an allele the older we expect it to be. As y approaches unity, the value of $A_1(1/2N, y)$ in (11) approaches $4N$, which is in accord with the result of Kimura & Ohta (1969). Upon substitution of $y = 1/2$ in $A_1(1/2N, y)$ of (11), we have $4N \log 2 \approx 2.8N$. Thus, if mutation rate is low, the mutant allele of a polymorphic locus at which the gene frequency is about $1/2$ has been approximately $2.8N$ generations in the population. It is interesting to compare this with $2N$ generations, which is the time required for a mutant gene to reach the gene frequency 0.5 for the first time.

Numerical values of formulae (10) and (12) are tabulated for wide ranges of value of $4Nu$ (Tables 1 and 2).

The tables reveal several biologically interesting facts of the age including fixation. When $4Nu \ll 1$, the age of an allele at high frequency is approximately

Table 1. Numerical values of $A_1(1/2N, y)/4N$ of (10), where $A_1(1/2N, y)$ is the average age before fixation of a neutral gene whose frequency is y , and N is the population size

(The values for $4Nu > 1$ are identical with those in Table 2.)

y	$4Nu$						
	1	0.9	0.7	0.5	0.3	0.1	0
0.001	0.0078	0.0077	0.0074	0.0072	0.0071	0.0069	0.0069
0.01	0.0561	0.0544	0.0521	0.0499	0.0482	0.0456	0.0450
0.03	0.1366	0.1321	0.1248	0.1177	0.1140	0.1086	0.1068
0.1	0.3418	0.3306	0.3074	0.2873	0.2704	0.2593	0.2536
0.2	0.5669	0.5426	0.4987	0.4623	0.4321	0.4105	0.4000
0.3	0.7528	0.7160	0.6529	0.6017	0.5595	0.5283	0.5136
0.5	1.0616	1.0010	0.9025	0.8245	0.7634	0.7129	0.6908
0.7	1.3203	1.2362	1.1059	1.0038	0.9240	0.8585	0.8298
0.9	1.5500	1.4403	1.2803	1.1561	1.0594	0.9804	0.9460
0.999	1.6458	1.5330	1.3591	1.2267	1.1219	1.0366	0.9995

Table 2. Numerical values of $A_2(1/2N, y)/4N$ of (12), where $A_2(1/2N, y)$ is the average age including fixation of a neutral gene whose frequency is y and N is the population size

y	$4Nu$									
	20	10	5	2	1	0.5	0.2	0.1	0.05	0.01
0.001	0.0040	0.0047	0.0054	0.0065	0.0074	0.0089	0.0123	0.0176	0.0278	0.1080
0.01	0.0228	0.0289	0.0357	0.0460	0.0557	0.0697	0.1042	0.1569	0.2588	1.0608
0.03	0.0448	0.0604	0.0787	0.1080	0.1361	0.1778	0.2807	0.4389	0.7446	3.1504
0.1	0.0827	0.1207	0.1697	0.2553	0.3432	0.4785	0.8189	1.3450	2.3635	10.3821
0.2	0.1105	0.1692	0.2500	0.4019	0.5666	0.8297	1.5047	2.5547	4.5903	20.6264
0.3	0.1284	0.2019	0.3074	0.5155	0.7507	1.1364	2.1422	3.7144	6.7661	30.8186
0.5	0.1523	0.2471	0.3904	0.6926	1.0553	1.6755	3.3341	5.9464	11.0277	51.1106
0.7	0.1686	0.2788	0.4512	0.8317	1.3089	2.153	4.4560	8.1039	15.2115	71.3214
0.9	0.1810	0.3033	0.4995	0.9477	1.5298	2.5919	5.5334	10.2126	19.3428	91.4554
0.999	0.1864	0.3140	0.5204	0.9993	1.6302	2.8051	6.0614	11.2478	21.3762	101.4366

Table 3. Numerical values of $A_1(1/2N, y)/4N$ of (13), where $A_1(1/2N, y)$ is the average age before fixation of an additive gene whose frequency is y

y	$ 4Ns $							
	100	50	20	10	5	2	1	0
0.001	0.0024	0.0031	0.0040	0.0047	0.0055	0.0062	0.0064	0.0065
0.01	0.0113	0.0158	0.0230	0.0295	0.0368	0.0436	0.0453	0.0460
0.03	0.0191	0.0287	0.0458	0.0625	0.0824	0.1014	0.1061	0.1080
0.1	0.0302	0.0487	0.0866	0.1288	0.1837	0.2372	0.2504	0.2553
0.2	0.0379	0.0634	0.1198	0.1882	0.2820	0.3721	0.3937	0.4019
0.3	0.0432	0.0737	0.1442	0.2344	0.3607	0.4778	0.5053	0.5155
0.5	0.0516	0.0904	0.1857	0.3164	0.4946	0.6461	0.6801	0.6926
0.7	0.0601	0.1078	0.2312	0.4028	0.6147	0.7817	0.8183	0.8317
0.9	0.0744	0.1380	0.3024	0.5018	0.7258	0.8969	0.9341	0.9477
0.999	0.1027	0.1780	0.3513	0.5526	0.7772	0.9484	0.9857	0.9992

equal to the reciprocal of the mutation rate, and the age of a low frequency allele is about y/u , where u is the mutation rate and y is the frequency. On the other hand, when $4Nu \gg 1$, the age tends to be much less. The values of Table 1 show that the age before fixation increases as the mutation rate increases. For example, the fixation time is $4N$ generations if $u = 0$, but it is $1.6 \times 4N$ generations if $4Nu = 1$. When $4Nu > 1$, formula (10) becomes invalid. Analytically the age before fixation and the age including fixation become identical, and they are given by formula (12). The age before fixation therefore increases as $4Nu$ increases to 1 and then decreases as $4Nu$ becomes larger. The age has a maximum at $4Nu = 1$.

The general formula obtained in this paper can be applied to a situation in which one of the alleles at the locus in question is selectively different from the rest. We will obtain the age before fixation for an allele with an additive effect on fitness. Let s be the selection coefficient of the gene and ignore the subsequent mutation of this allele. Then

$$V_{\delta x} = x(1-x)/2N \quad \text{and} \quad M_{\delta x} = sx(1-x),$$

and therefore

$$G(x) = \exp(-4Nsx) \quad \text{and} \quad g(y, x) = \{\exp(-4Nsy) - \exp(-4Nsx)\}/4Ns.$$

Substituting these functions into the general formulae we obtain the average age before fixation of the mutant gene whose frequency is y :

$$A_1\left(\frac{1}{2N}, y\right) = \frac{4N}{S(1-e^{-S})} \int_0^1 \frac{(e^{S\xi} - 1)(e^{-S\xi} - e^{-S})}{\xi(1-\xi)} d\xi - \frac{4N}{S(e^{-Sy} - e^{-S})} \int_y^1 \frac{(1 - e^{-S(1-\xi)})(e^{-Sy} - e^{-S\xi})}{\xi(1-\xi)} d\xi, \quad (13)$$

where $S = 4Ns$. Note that this formula is independent of the sign of $S = 4Ns$. Thus the average age of an additive gene is independent of the direction of selection pressure. Numerical evaluation of formula (13) reveals that if $|S| \leq 1$ the average age is nearly equal to that of a neutral gene. As $|S|$ becomes much greater than unity, the age of a gene whose frequency is not very small decreases almost as $1/|S|$. However, the value for a gene of very small frequency tends to stay closer to that for a neutral gene even for large $|S|$. A few numerical values are given in Table 3. It is interesting to note that the age including fixation of a neutral gene with $4Nu = 2$ is equal to the age before fixation of that with very small $4Ns$ (the column under $4Nu = 2$ of Table 1 and the column under $|4Ns| = 0$ of Table 3).

The above method can be extended to obtain the variance of the ages of alleles whose gene frequency is specified. Let

$$C(x, y) \equiv \int_0^\infty t^2 p(t, x, y) dt.$$

Then applying the operator L to $C(x, y)$, we have

$$LC(x, y) = -2 \int_0^\infty t p(t, x, y) dt = -2B(x, y).$$

Thus integration of this equation yields a formula similar to (5) in which the $\phi(\eta, y)$ is replaced by $2B(\eta, y)$. The boundary condition for the $C(x, y)$ is the same as that

for $B(x, y)$. The $C(x, y)/\phi(x, y)$ is the second moment of $A(x, y)$, and, in particular, $C(1/2N, y)/\phi(1/2N, y)$ is the second moment of the ages of alleles with gene frequency y .

I am greatly indebted to Professor Alan Robertson for many constructive comments, including suggestion of some of the topics discussed in this paper. This study was supported in part by U.S. Public Health Service Research Grants GM 19513 and GM 20293.

REFERENCES

- EWENS, W. J. (1969). *Population Genetics*. London: Methuen.
- KIMURA, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability* **1** 177–232.
- KIMURA, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903.
- KIMURA, M. & OHTA, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771.
- KIMURA, M. & OHTA, T. (1971). *Theoretical Aspects of Population Genetics*. Princeton, New Jersey: Princeton University Press.
- KIMURA, M. & OHTA, T. (1973). The age of a neutral mutant persisting in a finite population. *Genetics* **75**, 199–212.
- WRIGHT, S. (1938). The distribution of gene frequencies under irreversible mutation. *Proceedings of the National Academy of Sciences, U.S.A.* **24**, 253–259.

The age of an allele in a finite population*

By TAKEO MARUYAMA

National Institute of Genetics, Mishima, Japan

(Received 24 April 1973)

SUMMARY

The age of an allele segregating in a finite population may be defined in two ways. They are (1) the age of a mutant gene that has never reached fixation in the population, and (2) the age including any fixation period in the past. Theoretical expressions for these are derived on the assumption that every mutant is unique.

The history of mutant genes in finite populations, such as the fixation probability and the fixation time, has been extensively studied by numerous authors. These theories play an important role in population genetics and evolutionary theory (see Kimura & Ohta, 1971, for a review). This paper is to add a small piece of new knowledge to the theory, the age of an allele whose gene frequency is specified. This problem was first studied by Kimura & Ohta (1973). We assume that every mutant is new to the population. Suppose that the gene frequency of a mutant allele is y . Then there are two possibilities: (1) this allele has never reached fixation, and (2) it has reached fixation previously and the frequency decreased to y at present. We will treat two distinct situations. We first consider the average age under assumption (1) and then the age including both possibilities (1) and (2). We call the former 'the age before fixation' and the latter 'the age including fixation'. Kimura & Ohta (1973) have obtained the age before fixation for a neutral mutant under the assumption of low mutation rate. This paper uses the same method as that of Kimura and Ohta, and obtains general formulas for both definitions of age without restriction of mutation rate.

Let $V_{\delta x}$ and $M_{\delta x}$ be respectively the variance and the mean of the change in the gene frequency x in one generation, and let $p(t, x, y)$ be the transition probability that the gene frequency is x at time 0 and is y at time t . Then it is well known that the $p(t, x, y)$ satisfies

$$\frac{\partial p(t, x, y)}{\partial t} = \mathbf{L}p(t, x, y), \quad (1)$$

in which \mathbf{L} is the operator

$$\frac{V_{\delta x}}{2} \frac{d^2}{dx^2} + M_{\delta x} \frac{d}{dx}.$$

The partial differential equation (1) is called the Kolmogorov backward equation. For fixed x and y , let

$$B(x, y) \equiv \int_0^\infty t p(t, x, y) dt.$$

* Contribution No. 960 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411 Japan.

$B(x, y)$ gives the average time it took to pass from x to y , times the probability of being at y . Applying the operator L to $B(x, y)$ and using (1), we have

$$LB(x, y) = \int_0^\infty tLp(t, x, y) dt = \int_0^\infty t \frac{\partial p(t, x, y)}{\partial t} dt$$

$$= tp(t, x, y)|_0^\infty - \int_0^\infty p(t, x, y) dt.$$

Note that $tp(t, x, y)|_0^\infty = 0 \quad \int_0^\infty p(t, x, y) dt \equiv \phi(x, y)$

is the sojourn time at frequency y , and $B(x, y)/\phi(x, y)$ is the average interval since the frequency was x . Thus

$$LB(x, y) = -\dot{\phi}(x, y). \tag{2}$$

Applying L to $\phi(x, y)$,

$$L\phi(x, y) = \int_0^\infty Lp(t, x, y) dt = \int_0^\infty \frac{\partial p(t, x, y)}{\partial t} dt = p(t, x, y)|_0^\infty$$

$$= -p(0, x, y) = -\delta(x - y), \tag{3}$$

where $\delta(\cdot)$ is Dirac's delta function. This differential equation can be integrated:

$$\phi(x, y) = -2H(\delta(\eta - y); x) + Ag(0, x), \tag{4}$$

in which

$$G(x) \equiv \exp\left\{-2 \int^x \frac{M_{\delta\xi}}{V_{\delta\xi}} d\xi\right\},$$

$$g(x, y) \equiv \int_x^y G(\xi) d\xi,$$

$$H(f(\eta); x) \equiv \int_0^x G(\xi) \int_0^\xi \frac{f(\eta)}{V_{\delta\eta} G(\eta)} d\eta d\xi$$

for an arbitrary function $f(\cdot)$, and A is a constant to be determined from the boundary condition on $\phi(x, y)$ at $x = 1$. Note that the other boundary condition at $x = 0$ is $\phi(0, y) = 0$, because no reverse mutation occurs and therefore if the gene frequency once becomes 0, it will never become non-zero again. The equation (3) and the solution (4) are essentially due to Wright (1938), Kimura (1964, 1969) and Ewens (1969). Using the $\phi(x, y)$ of (4), the equation (2) can be solved:

$$B(x, y) = -2H(\phi(\eta, y); x) + Ag(0, x), \tag{5}$$

in which A is constant to be determined from the boundary condition at $x = 1$. As in (4), the other boundary is $B(0, y) = 0$.

We can determine the constant A in (4) and (5). For the age before fixation, $\phi(1, y) = 0$ and $B(1, y) = 0$ for all $0 < y < 1$. This is because the process (or a sample path) is stopped as soon as it reaches the boundary, i.e. $x = 1$ is an absorbing state. In this case, assuming $u(x) = g(0, x)/g(0, 1)$ exists,

$$\left. \begin{aligned} \phi_1(x, y) &= \frac{2g(y, 1)u(x)}{V_{\delta y}G(y)} \quad \text{for } y > x \\ &= \frac{2}{V_{\delta y}G(y)} [g(y, 1)u(x) - g(y, x)] \quad \text{for } y < x \end{aligned} \right\} \tag{6}$$

and

$$B_1(x, y) = 2u(x)H(\phi(\eta, y); 1) - 2H(\phi(\eta, y); x),$$

where subscript 1 indicates 'before fixation'.

On the other hand, for the age including fixation, the process is considered even after it reaches fixation until it eventually reaches extinction. In this case, as the starting point x approaches 1, the fate of the gene frequency becomes the same as that starting $x = 1$. Therefore $\phi_x(1, y) = 0$ and $B_x(1, y) = 0$, where the subscript x indicates the derivative with respect to the first variable. With these boundary conditions,

$$\phi_2(x, y) = \left. \begin{aligned} & \frac{2g(0, x)}{V_{\delta y} G(y)} \quad \text{for } y > x \\ & = \frac{2g(0, y)}{V_{\delta y} G(y)} \quad \text{for } y < x \end{aligned} \right\} \quad (7)$$

and

$$B_2(x, y) = 2g(0, x) \int_0^1 \frac{\phi(\xi, y)}{V_{\delta \xi} G(\xi)} d\xi - 2H(\phi(\eta, y); x),$$

where subscript 2 indicates 'including fixation'. The value of function $\phi(x, y)$ is the total number of generations in which the gene frequency is y if it starts from x , and the $B(x, y)$ gives the average time it took to pass from x to y times the probability of being at y . Therefore if we consider many similar loci and assume that the population is at steady state, $A(1/2N, y) \equiv B(1/2N, y)/\phi(1/2N, y)$ gives the average age of alleles found in such a population.

Let N be the population size and u be mutation rate per gene per generation. Since every mutant is unique, the average number of mutant to occur in each generation is $2Nu$. Therefore $2Nu$ times $\phi_1(1/2N, y)$ or $\phi_2(1/2N, y)$ gives respectively the density of the gene frequency distribution at equilibrium, among those alleles which have never reached fixation and among those including any fixation period. If all alleles are selectively neutral and if $V_{\delta y} = x(1-x)/2N$ and $M_{\delta x} = -ux$ which are usually used,

$$\phi_1\left(\frac{1}{2N}, y\right) = \frac{4N}{yF} \left[1 - \left(1 - \frac{1}{2N}\right)^F \right] \approx \frac{4N}{yF} \frac{F}{2N} = \frac{2}{y},$$

where $F = 1 - 4Nu$. Therefore,

$$2Nu\phi_1\left(\frac{1}{2N}, y\right) \approx \frac{4Nu}{y}. \quad (8)$$

On the other hand, the distribution including fixation period is

$$\phi_2\left(\frac{1}{2N}, y\right) = \frac{4N \left[1 - \left(1 - \frac{1}{2N}\right)^F \right]}{Fy(1-y)^F} \approx \frac{2}{y(1-y)^F}$$

and

$$2Nu\phi_2\left(\frac{1}{2N}, y\right) \approx \frac{4Nu}{y(1-y)^F}. \quad (9)$$

It is interesting to compare formulae (8) and (9). The former is linearly related to mutation rate, whereas formula (9) is a non-linear function of $F = 1 - 4Nu$.

The ratio, $\phi_1(1/2N, y)/\phi_2(1/2N, y)$, gives the probability that if a mutant is known

to have frequency y , it has never been fixed. This probability for a neutral mutant is equal to

$$\frac{\phi_1(1/2N, y)}{\phi_2(1/2N, y)} = (1-y)^F,$$

provided $F = 1 - 4Nu > 0$, which is the condition that temporary fixation is possible. If $4Nu$ tends to 0, the probability becomes $(1-y)$. On the other hand, if $4Nu$ tends to 1, the probability becomes 1 for all y . The latter result is obvious, because as $4Nu$ becomes 1, fixation ceases to occur.

Returning to the main subject of this paper, we can obtain explicit formulae of the average age before fixation and including fixation for some simple cases. As a typical case, let $V_{\delta x} = x(1-x)/2N$ and $M_{\delta x} = -ux$, where N is the population size and u is mutation rate. Then, assuming $4Nu < 1$ and ignoring a term of order $1/2N$, the age before fixation of a neutral mutant allele whose frequency is y , is

$$A_1\left(\frac{1}{2N}, y\right) = \frac{4N}{F} \left[\int_0^y \frac{\{1 - (1-\xi)^F\}}{\xi} d\xi + \frac{\{1 - (1-y)^F\}}{(1-y)^F} \int_y^1 \frac{(1-\xi)^F}{\xi} d\xi \right], \quad (10)$$

where $F = 1 - 4Nu$, and, as $4Nu$ approaches 0, the above formula becomes

$$A_1\left(\frac{1}{2N}, y\right) = \frac{-4Ny \log y}{1-y}. \quad (11)$$

This is the average age of a neutral mutant gene whose frequency is y , provided that it has never reached fixation and that every mutant arises at a homallelic locus, i.e. $4Nu \ll 1$. Formula (10) is the same as (13a) of Kimura & Ohta (1973). However, since we usually do not know whether a mutant gene has previously reached fixation or not, the age including fixation is more important. The average age including fixation of a neutral gene is

$$A_2\left(\frac{1}{2N}, y\right) = \frac{4N}{F} \left[\int_0^y \frac{\{1 - (1-\xi)^F\}}{\xi(1-\xi)^F} d\xi + \{1 - (1-y)^F\} \int_y^1 \frac{d\xi}{\xi(1-\xi)^F} \right], \quad (12)$$

where $F = 1 - 4Nu$.

The validity of formulae (11) and (12) were verified by computer simulation and results agreed well with the theoretical expectations. We should first note that both functions (10) and (12) are monotone, increasing so that the higher the frequency of an allele the older we expect it to be. As y approaches unity, the value of $A_1(1/2N, y)$ in (11) approaches $4N$, which is in accord with the result of Kimura & Ohta (1969). Upon substitution of $y = 1/2$ in $A_1(1/2N, y)$ of (11), we have $4N \log 2 \approx 2.8N$. Thus, if mutation rate is low, the mutant allele of a polymorphic locus at which the gene frequency is about $1/2$ has been approximately $2.8N$ generations in the population. It is interesting to compare this with $2N$ generations, which is the time required for a mutant gene to reach the gene frequency 0.5 for the first time.

Numerical values of formulae (10) and (12) are tabulated for wide ranges of value of $4Nu$ (Tables 1 and 2).

The tables reveal several biologically interesting facts of the age including fixation. When $4Nu \ll 1$, the age of an allele at high frequency is approximately

Table 1. Numerical values of $A_1(1/2N, y)/4N$ of (10), where $A_1(1/2N, y)$ is the average age before fixation of a neutral gene whose frequency is y , and N is the population size

(The values for $4Nu > 1$ are identical with those in Table 2.)

y	$4Nu$						
	1	0.9	0.7	0.5	0.3	0.1	0
0.001	0.0078	0.0077	0.0074	0.0072	0.0071	0.0069	0.0069
0.01	0.0561	0.0544	0.0521	0.0499	0.0482	0.0456	0.0450
0.03	0.1366	0.1321	0.1248	0.1177	0.1140	0.1086	0.1068
0.1	0.3418	0.3306	0.3074	0.2873	0.2704	0.2593	0.2536
0.2	0.5669	0.5426	0.4987	0.4623	0.4321	0.4105	0.4000
0.3	0.7528	0.7160	0.6529	0.6017	0.5595	0.5283	0.5136
0.5	1.0616	1.0010	0.9025	0.8245	0.7634	0.7129	0.6908
0.7	1.3203	1.2362	1.1059	1.0038	0.9240	0.8585	0.8298
0.9	1.5500	1.4403	1.2803	1.1561	1.0594	0.9804	0.9460
0.999	1.6458	1.5330	1.3591	1.2267	1.1219	1.0366	0.9995

Table 2. Numerical values of $A_2(1/2N, y)/4N$ of (12), where $A_2(1/2N, y)$ is the average age including fixation of a neutral gene whose frequency is y and N is the population size

y	$4Nu$									
	20	10	5	2	1	0.5	0.2	0.1	0.05	0.01
0.001	0.0040	0.0047	0.0054	0.0065	0.0074	0.0089	0.0123	0.0176	0.0278	0.1080
0.01	0.0228	0.0289	0.0357	0.0460	0.0557	0.0697	0.1042	0.1569	0.2588	1.0608
0.03	0.0448	0.0604	0.0787	0.1080	0.1361	0.1778	0.2807	0.4389	0.7446	3.1504
0.1	0.0827	0.1207	0.1697	0.2553	0.3432	0.4785	0.8189	1.3450	2.3635	10.3821
0.2	0.1105	0.1692	0.2500	0.4019	0.5666	0.8297	1.5047	2.5547	4.5903	20.6264
0.3	0.1284	0.2019	0.3074	0.5155	0.7507	1.1364	2.1422	3.7144	6.7661	30.8186
0.5	0.1523	0.2471	0.3904	0.6926	1.0553	1.6755	3.3341	5.9464	11.0277	51.1106
0.7	0.1686	0.2788	0.4512	0.8317	1.3089	2.153	4.4560	8.1039	15.2115	71.3214
0.9	0.1810	0.3033	0.4995	0.9477	1.5298	2.5919	5.5334	10.2126	19.3428	91.4554
0.999	0.1864	0.3140	0.5204	0.9993	1.6302	2.8051	6.0614	11.2478	21.3762	101.4366

Table 3. Numerical values of $A_1(1/2N, y)/4N$ of (13), where $A_1(1/2N, y)$ is the average age before fixation of an additive gene whose frequency is y

y	$ 4Ns $							
	100	50	20	10	5	2	1	0
0.001	0.0024	0.0031	0.0040	0.0047	0.0055	0.0062	0.0064	0.0065
0.01	0.0113	0.0158	0.0230	0.0295	0.0368	0.0436	0.0453	0.0460
0.03	0.0191	0.0287	0.0458	0.0625	0.0824	0.1014	0.1061	0.1080
0.1	0.0302	0.0487	0.0866	0.1288	0.1837	0.2372	0.2504	0.2553
0.2	0.0379	0.0634	0.1198	0.1882	0.2820	0.3721	0.3937	0.4019
0.3	0.0432	0.0737	0.1442	0.2344	0.3607	0.4778	0.5053	0.5155
0.5	0.0516	0.0904	0.1857	0.3164	0.4946	0.6461	0.6801	0.6926
0.7	0.0601	0.1078	0.2312	0.4028	0.6147	0.7817	0.8183	0.8317
0.9	0.0744	0.1380	0.3024	0.5018	0.7258	0.8969	0.9341	0.9477
0.999	0.1027	0.1780	0.3513	0.5526	0.7772	0.9484	0.9857	0.9992

equal to the reciprocal of the mutation rate, and the age of a low frequency allele is about y/u , where u is the mutation rate and y is the frequency. On the other hand, when $4Nu \gg 1$, the age tends to be much less. The values of Table 1 show that the age before fixation increases as the mutation rate increases. For example, the fixation time is $4N$ generations if $u = 0$, but it is $1.6 \times 4N$ generations if $4Nu = 1$. When $4Nu > 1$, formula (10) becomes invalid. Analytically the age before fixation and the age including fixation become identical, and they are given by formula (12). The age before fixation therefore increases as $4Nu$ increases to 1 and then decreases as $4Nu$ becomes larger. The age has a maximum at $4Nu = 1$.

The general formula obtained in this paper can be applied to a situation in which one of the alleles at the locus in question is selectively different from the rest. We will obtain the age before fixation for an allele with an additive effect on fitness. Let s be the selection coefficient of the gene and ignore the subsequent mutation of this allele. Then

$$V_{\delta x} = x(1-x)/2N \quad \text{and} \quad M_{\delta x} = sx(1-x),$$

and therefore

$$G(x) = \exp(-4Nsx) \quad \text{and} \quad g(y, x) = \{\exp(-4Nsy) - \exp(-4Nsx)\}/4Ns.$$

Substituting these functions into the general formulae we obtain the average age before fixation of the mutant gene whose frequency is y :

$$A_1\left(\frac{1}{2N}, y\right) = \frac{4N}{S(1-e^{-S})} \int_0^1 \frac{(e^{S\xi} - 1)(e^{-S\xi} - e^{-S})}{\xi(1-\xi)} d\xi - \frac{4N}{S(e^{-Sy} - e^{-S})} \int_y^1 \frac{(1 - e^{-S(1-\xi)})(e^{-Sy} - e^{-S\xi})}{\xi(1-\xi)} d\xi, \quad (13)$$

where $S = 4Ns$. Note that this formula is independent of the sign of $S = 4Ns$. Thus the average age of an additive gene is independent of the direction of selection pressure. Numerical evaluation of formula (13) reveals that if $|S| \leq 1$ the average age is nearly equal to that of a neutral gene. As $|S|$ becomes much greater than unity, the age of a gene whose frequency is not very small decreases almost as $1/|S|$. However, the value for a gene of very small frequency tends to stay closer to that for a neutral gene even for large $|S|$. A few numerical values are given in Table 3. It is interesting to note that the age including fixation of a neutral gene with $4Nu = 2$ is equal to the age before fixation of that with very small $4Ns$ (the column under $4Nu = 2$ of Table 1 and the column under $|4Ns| = 0$ of Table 3).

The above method can be extended to obtain the variance of the ages of alleles whose gene frequency is specified. Let

$$C(x, y) \equiv \int_0^\infty t^2 p(t, x, y) dt.$$

Then applying the operator L to $C(x, y)$, we have

$$LC(x, y) = -2 \int_0^\infty t p(t, x, y) dt = -2B(x, y).$$

Thus integration of this equation yields a formula similar to (5) in which the $\phi(\eta, y)$ is replaced by $2B(\eta, y)$. The boundary condition for the $C(x, y)$ is the same as that

for $B(x, y)$. The $C(x, y)/\phi(x, y)$ is the second moment of $A(x, y)$, and, in particular, $C(1/2N, y)/\phi(1/2N, y)$ is the second moment of the ages of alleles with gene frequency y .

I am greatly indebted to Professor Alan Robertson for many constructive comments, including suggestion of some of the topics discussed in this paper. This study was supported in part by U.S. Public Health Service Research Grants GM 19513 and GM 20293.

REFERENCES

- EWENS, W. J. (1969). *Population Genetics*. London: Methuen.
- KIMURA, M. (1964). Diffusion models in population genetics. *Journal of Applied Probability* **1**, 177–232.
- KIMURA, M. (1969). The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics* **61**, 893–903.
- KIMURA, M. & OHTA, T. (1969). The average number of generations until fixation of a mutant gene in a finite population. *Genetics* **61**, 763–771.
- KIMURA, M. & OHTA, T. (1971). *Theoretical Aspects of Population Genetics*. Princeton, New Jersey: Princeton University Press.
- KIMURA, M. & OHTA, T. (1973). The age of a neutral mutant persisting in a finite population. *Genetics* **75**, 199–212.
- WRIGHT, S. (1938). The distribution of gene frequencies under irreversible mutation. *Proceedings of the National Academy of Sciences, U.S.A.* **24**, 253–259.