PROBABILITY OF FIXATION AND MEAN FIXATION TIME OF AN OVERDOMINANT MUTATION*

MASATOSHI NEI AND A. K. ROYCHOUDHURY

Center for Demographic and Population Genetics, University of Texas, Houston, Texas 77025

Manuscript received August 14, 1972
Revised copy received January 11, 1973
Transmitted by R. C. LEWONTIN

ABSTRACT

The probability of fixation of an overdominant mutation in a finite population depends on the equilibrium gene frequency in an infinite population (m) and the product (A) of population size and selection intensity. If $m < 0.5$ (disadvantageous overdominant genes), the probability is generally much lower than that of neutral genes; but if m is close to 0.5 and A is relatively small, it becomes higher. If $m > 0.5$ (advantageous overdominant genes), the probability is largely determined by the fitness of heterozygotes rather than that of mutant homozygotes. Thus, overdominance enhances the probability of fixation of advantageous mutations. The average number of generations until fixation of an overdominant mutation also depends on m and A. This average time is long when m is close to 0.5 but short when m is close to 0 or 1. This dependence on m and A is similar to that of ROBERTSON's retardation factor.

BECAUSE of its evolutionary importance, the probability of fixation of a new mutant gene in a population has been studied by many authors (HALDINE 1927; FISHER 1930; WRIGHT 1942; KIMURA 1957; and others). These authors have worked out rather simple formulae for the probability of fixation of a dominant, semidominant or recessive gene. Very little attention, however, has been paid to an overdominant mutation, perhaps because this type of mutation creates a stable polymorphism in a large population. In nature, however, the effective size of populations is sometimes quite small, so that even overdominant genes may be fixed or lost from the population by chance. This is true also in laboratory experiments such as those conducted by TRACEY (1972). In a study of steady decay of genetic variability, ROBERTSON (1962) showed that in a finite population overdominance may accelerate rather than retard fixation of genes, if the equilibrium gene frequency in an infinite population is outside the range of approximately 0.2 to 0.8. EWENS and THOMSON (1970) and CARR and NASSAR (1970b) obtained a similar result from a study of the average time to fixation or loss. These results suggest that the probability of fixation of an overdominant gene is rather complicated, depending on the relative values of selection coefficients for homozygotes.

* Part of this work was done while the authors were tenured at the Division of Biological and Medical Sciences, Brown University, Providence, R. I.

Theoretically, a general formula for the probability of fixation of a gene was derived by KIMURA (1957, 1962). In the case of overdominant genes, however, numerical integration is required, though in very special cases simplified formulae can be obtained. CARR and NASSAR (1970a) numerically studied the relation between the probability of fixation of overdominant genes and the initial gene frequency. This study appears to be important in relation to animal and plant breeding but apparently is not for evolutionary studies.

In the study of evolution it is also important to know the mean fixation time for an overdominant gene. KIMURA and OHTA (1969) studied this problem for the special case of the equilibrium gene frequency equal to 0.5 and showed that overdominance prolongs the fixation time compared with that of neutral genes. EWENS and THOMSON (1970) and CARR and NASSAR (1970b) studied the expected time until fixation or loss, without separating the events of fixation and loss.

In the present paper we shall study the probability of fixation of a single overdominant mutation that appears in a population of size N. Thus, the initial gene frequency is always $1/(2N)$. We will also investigate the average number of generations until fixation, excluding the event of loss.

PROBABILITY OF FIXATION OF A MUTANT GENE

Consider a randomly mating diploid population of size N, in which an overdominant gene a and its allelic gene A are segregating. Let x be the gene frequency of a. We designate the fitnesses of genotypes AA, Aa, and aa by $W_{AA} = 1 - s_1$, $W_{Aa} = 1$, and $W_{aa} = 1 - s_2$, respectively. Then, the mean ($M_{\delta x}$) and variance ($V_{\delta x}$) of the change of gene frequency per generation are given by

$$M_{\delta x} = -(s_1 + s_2)(x - m)x(1 - x) \tag{1}$$

$$V_{\delta x} = x(1 - x)/(2N) \tag{2}$$

where $m = s_1/(s_1 + s_2)$ is the equilibrium gene frequency in an infinitely large population. Putting these quantities into KIMURA's (1962) general formula, the ultimate probability of fixation of gene a becomes

$$u_h(p) = \int_0^p e^{A(x-m)^2}\, dx \Big/ \int_0^1 e^{A(x-m)^2}\, dx \tag{3}$$

where p is the initial gene frequency of a and $A = 2N(s_1 + s_2)$. In the present paper p is $1/(2N)$, unless it is mentioned otherwise. When A is small compared with unity, the above formula can be simplified by expanding the integrand (ROBERTSON 1962). Also, formula (3) can be written as

$$u_h(p) = \frac{e^{Am^2}F(m\sqrt{A}) - e^{A(m-p)^2}F((m-p)\sqrt{A})}{e^{Am^2}F(m\sqrt{A}) + e^{A(1-m)^2}F((1-m)\sqrt{A})} \tag{4}$$

where $p \leq m$ and $F(x) = e^{-x^2}\int_0^x e^{t^2}\, dt$ is Dawson's integral. If $p > m$, then the second term in the numerator has a positive sign. The value of $F(x)$ has been tabulated by ABRAMOWITZ and STEGUN (1964) for $x = 0.00 \sim 2.00$. In general, however, $u_h(p)$ has to be evaluated by numerical methods. We, therefore, used GAUSS's method of numerical integration, except for some special cases. Formula

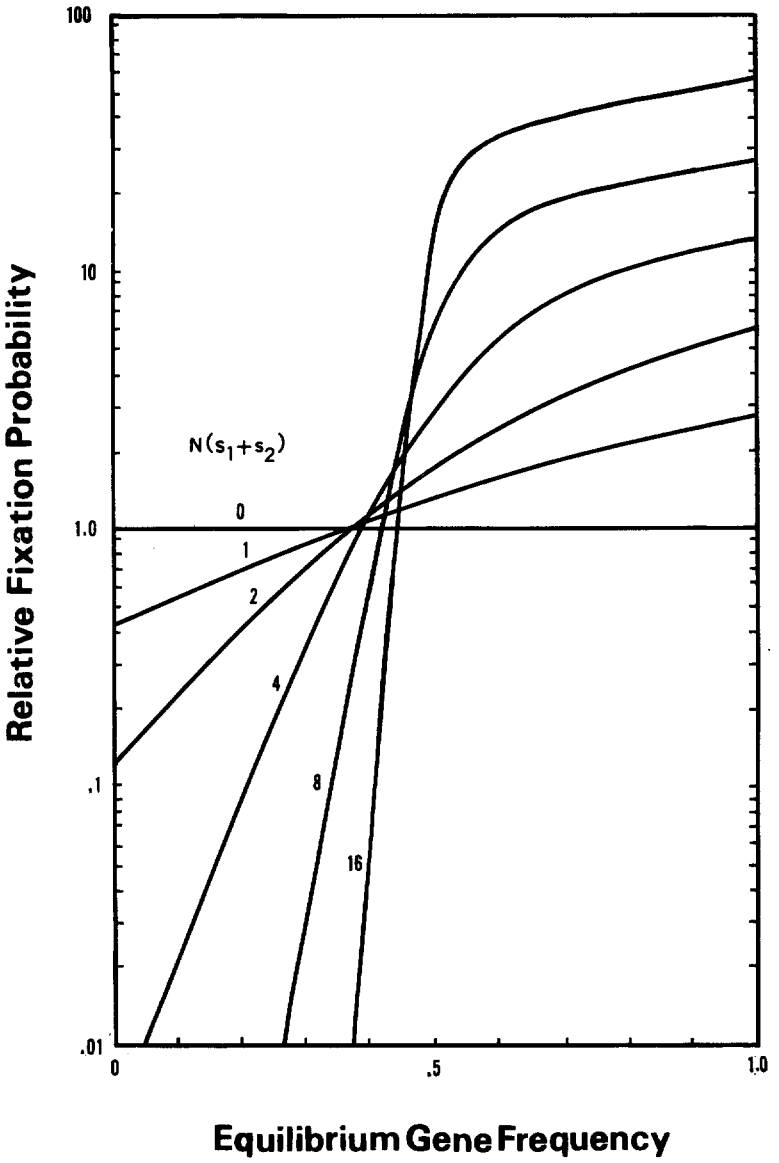

FIGURE 1.—Probability of fixation of an overdominant mutation relative to that of a neutral mutation.

(4) was used to check the numerical computations. Numerical results were also checked by evaluating the Taylor expansions of formula (3).

In Figure 1 the value of $u_h(p)$ relative to the probability of fixation of a neutral gene $(1/2N)$, i.e. the value of $2Nu_h(p)$ is given on the assumption that N is large and s_1 is small compared with unity. In this case $\int_0^p \exp A(x-m)^2\,dx$ is

$(\exp Am^2)/(2N)$ approximately, so that the relative fixation probability is

$$2Nu_h\ (p) = e^{Am^2}/\int_0^1 e^{A(x-m)^2}\ dx$$

approximately. Thus, it depends only on A and m. It is seen that the relative fixation probability is lower than 1 if $m \equiv s_1/(s_1+s_2)$ is much smaller than 0.5. This is expected because in this case the fitness $(1-s_2)$ of mutant homozygote aa is lower than that $(1-s_1)$ of wild-type homozygote AA. The relative fixation probability increases as m increases, and the rate of increase is higher when $N(s_1+s_2)$ is large. Here, an interesting property emerges; the relative fixation probability becomes higher than 1 when m is close to but still smaller than 0.5. Namely, $u_h\ (p)$ is higher than the probability of fixation of a neutral gene even if aa is less fit than AA. This may be explained by the initial advantage of the a gene. When the frequency of the a gene is small, it is mostly in the heterozygous condition, so that it has a selective advantage over the wild-type allele. If m is larger than 0.5, $u_h\ (p)$ is always larger than that of neutral gene.

It is known that for an advantageous mutation to be fixed in the population the fitness of heterozygotes plays an important role. Thus, the probability of fixation of a completely dominant gene ($W_{AA}=1-s_1$, $W_{Aa}=1$, $W_{aa}=1$) is almost the same as that of a semidominant gene ($W_{AA}=1-s_1$, $W_{Aa}=1$, $W_{aa}=1+s_1$). Thus, it is interesting to compare $u_h\ (p)$ with the probability of fixation of a semidominant gene. The probability of fixation of a semidominant gene is given by

$$u_s\ (p) = 4Ns_1p/(1-e^{-4Ns_1})$$

approximately, when s_1 is small compared with unity. Then, the ratio $u_h(p)/u_s(p)$ is

$$\frac{(1-e^{-2Am})e^{Am^2}}{2Am\int_0^1 e^{A(x-m)^2}\,dx}$$

approximately. Note that $4Ns_1$ is equal to $2Am$. This value is again dependent only on A and m. This ratio is given in Figure 2. Clearly, this value is smaller than 1 when the fitness of aa is lower than that of AA $(m<0.5)$. That is, in this case the probability of fixation of an overdominant gene is affected by the fitness of aa as well as by the fitness of Aa. However, if the fitness of aa is higher than that of AA, $u_h(p)$ is largely determined by the fitness of heterozygotes. If $m>0.5$ and $N(s_1+s_2)>16$, $u_h\ (p)$ is almost the same as $u_s\ (p)$.

MEAN FIXATION TIME

KIMURA and OHTA (1969) derived a general formula for the average number of generations until fixation of a mutant gene in the population. It is given by

$$\bar{t}_1(p) = \int_p^1 \psi(x)u(x)\ \{1-u(x)\}dx + \frac{1-u(p)}{u(p)}\int_0^p \psi(x)u^2(x)dx$$

where

$$\psi(x) = 2\int_0^1 G(y)dy/\{V_{\delta x}\,G(x)\}$$

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

in which

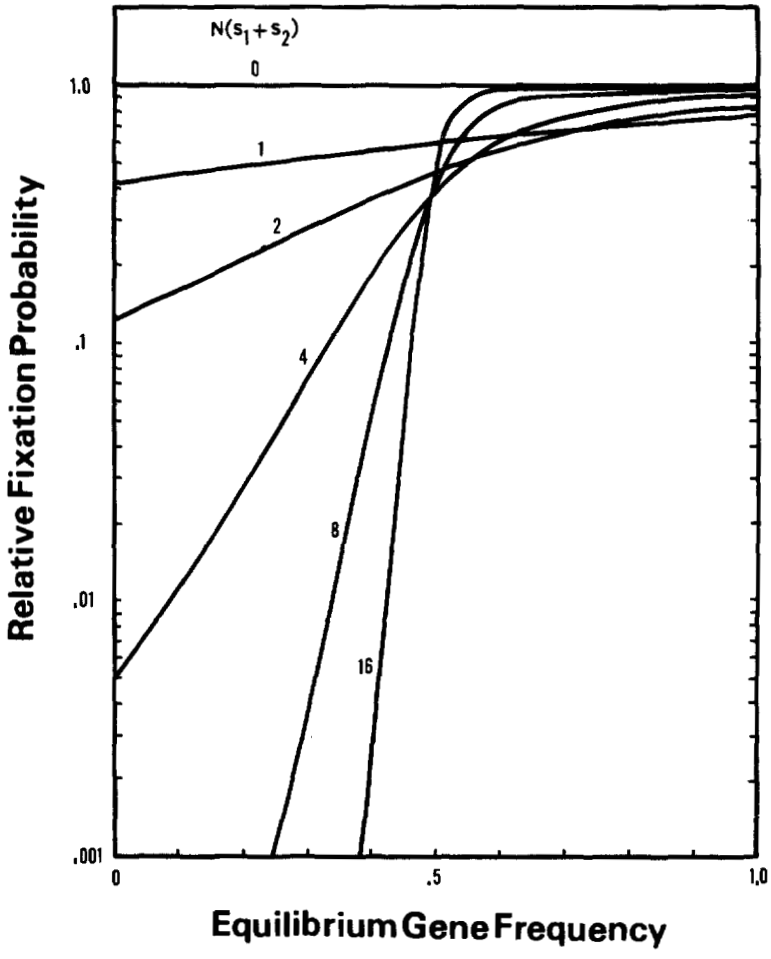

FIGURE 2.—Probability of fixation of an overdominant mutation relative to that of a semi-dominant mutation.

$$G(x) = \exp\{-2\int(M_{\delta x}/V_{\delta x})dx\} .$$

Therefore, using the expressions for $M_{\delta x}$ and $V_{\delta x}$ given in (1) and (2), we can derive the following formula.

$$\bar{t}_1(p) = 4N\int_0^1 \frac{\int_0^y e^{A(x-m)^2}\,dx \int_y^1 e^{A(x-m)^2}\,dx}{K\gamma(1-\gamma)e^{A(y-m)^2}}\,dy$$

$$- 4N\int_0^p \frac{\int_0^y e^{A(x-m)^2}\,dx \int_y^p e^{A(x-m)^2}\,dx}{K'\gamma(1-\gamma)e^{A(y-m)^2}}\,dy \qquad (5)$$

where $K = \int_0^1 \exp A(x-m)^2 dx$ and $K' = \int_0^p \exp A(x-m)^2 dx$.

Analytical solution of (5) is not easy, so that we used GAUSS's method of numerical integration. The results obtained are given in Figure 3. In this figure

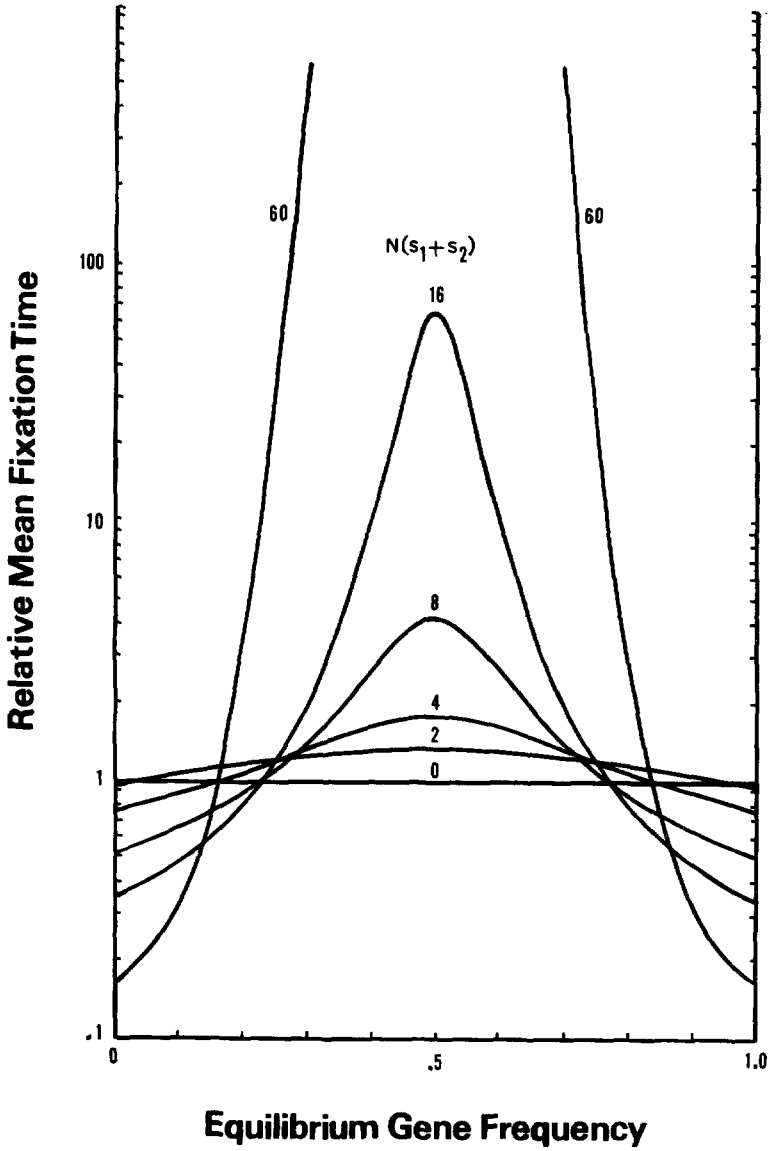

FIGURE 3.—Mean fixation time of an overdominant mutation relative to that of a neutral mutation.

$\bar{t}_1(p)$ is given relative to the fixation time of a neutral mutation, i.e. $4N$ (KIMURA and OHTA 1969). Figure 3 shows that the relative mean fixation time depends markedly on the equilibrium gene frequency, m. As expected, if m is close to 0.5, the fixation time is much longer than that for neutral genes when $N(s_1 + s_2)$ is large. However, if m is outside the range of approximately 0.2 to 0.8, the fixation time of overdominant mutations is shorter than that of neutral mutations, depending on the value of $N(s_1 + s_2)$. A continued increase in this quantity gradually widens the range of m for prolonged mean fixation time. It is seen that

the relative fixation time is virtually symmetric around $m = 0.5$. The relation between the relative mean fixation time and m is superficially similar to that between ROBERTSON's (1962) retardation factor and m, though the absolute values are considerably different. For example, when $N(s_1 + s_2) = 16$ and $m = 0.5$, the relative fixation time is 68.3, while the retardation factor is slightly more than 100.

Although the relative fixation time and retardation factor are superficially similar, the explanations must be entirely different. The retardation factor is the rate of decay of gene frequency distribution at steady state relative to that of neutral genes. Mathematically, the rate of decay is given by $\lambda = -\frac{1}{\phi(x;t)} \frac{\partial \phi(x;t)}{\partial t}$, in which $\phi(x;t)$ is the gene frequency distribution at steady state. On the other hand, the average fixation time is defined as

$$\bar{t}_1(p) = \int_0^\infty t \frac{u(p,t)}{\partial t} dt/u(p),$$

where $u(p,t)$ is the probability that a gene whose initial frequency is p is fixed in the population by generation t (KIMURA and OHTA 1969).

From the above definitions it is clear that the retardation factor $2N\lambda$ (the rate of decay for neutral genes $= 1/(2N)$) is symmetric around $m = 0.5$, since both the loss and fixation of genes are taken into account. The reason for the symmetry of $\bar{t}_1(p)$ is less obvious, but it is provided by studying the conditional sojourn time (MARUYAMA and KIMURA 1971; EWENS 1972). These authors have shown that the mean time the frequency of mutant gene a spends in the interval $(x, x + dx)$, given that eventually a is fixed, is $t^*(x)dx$, where

$$t^*(x) = 2u(x)\{V_{\delta x} G(x)\}^{-1} \int_x^1 G(y)dy, \quad p \leq x \leq 1$$

$$t^*(x) = 2\{1 - u(p)\}u(x)\ \{u(p)V_{\delta x} G(x)\}^{-1} \int_0^x G(y)dy. \quad 0 \leq x \leq p$$

In the present case $p = 1/(2N)$ is very close to 0 when N is large, so that we can neglect the range $0 \leq x \leq p$. Then, the conditional sojourn time, $t^*(x)$, is

$$t^*(x) \equiv t^*(x,m) = 4N \int_0^x e^{A(y-m)^2} dy \int_x^1 e^{A(y-m)^2} dy / \{Kx(1-x)e^{A(x-m)^2}\}$$

where

$$K = \int_0^1 \exp A(y-m)^2 dy.$$

It is now clear that $t^*(x,m)$ is equal to $t^*(1-x, 1-m)$. Therefore, the average fixation time, which may be obtained by $\int_0^1 t^*(x,m)dx$, becomes symmetric around $m = 0.5$ when $p \to 0$. Note that the first term in (5) is the same as $\int_0^1 t^*(x,m)dx$. The second term of this equation appears because actually p is not 0, but this term, which causes a slight asymmetry of $\bar{t}_1(p)$, is negligibly small compared with the first term.

In order to check our results, which were obtained by the diffusion method, we made a computer simulation by using the method of Markov chains. In this simulation $N = 10$ and $s_1 + s_2 = 0.4$ were used, and the average number of generations

TABLE 1

Average numbers of generations until fixation and loss of an overdominant gene with $N = 10$ *and* $s_1 + s_2 = 0.4$.
These results were obtained by the Markov chain method

m	Probability of fixation	Fixation time (generations)	Loss time (generations)
0.1	.00055	29.98	6.51
0.2	.00272	37.29	8.38
0.8	.4894	37.31	3.51
0.9	.5621	29.99	2.70

until fixation and loss were separately recorded. Some of the results obtained are given in Table 1. It is clear that even in a population of size 10 the symmetry about fixation time is observed. On the other hand, there is no such symmetry about the average loss time. We also studied the probability distribution of fixation time, i.e. $\Delta u(p,t)/u(p)$, where $\Delta u(p,t)$ is the absolute probability of fixation at generation t. The values for selected generations are given in Table 2. It is seen that the probability distributions of fixation times for mutant genes with m and with $1 - m$ are virtually the same except for the first generation, of which the values are extremely small. While we have not fully understood why the two probability distributions should be virtually the same, this provides another explanation for the symmetry of average fixation time around $m = 0.5$. In this connection, it is worthwhile to note that in the four cases given in Table 2 the steady state distribution of gene frequencies (with the accuracy of the second significant value) was attained approximately at the $50th$ generation, by which time the mutant gene had been lost or fixed with probability $0.90 \sim .99$.

One of the interesting conclusions which can be made from this study is that

TABLE 2

Probabilities of fixation of an overdominant gene at selected generations $(\Delta u(p,t)/u(p))$ *with* $N = 10$ *and* $s_1 + s_2 = 0.4$.
These results were obtained by the Markov chain method

Generation	$m = 0.1$	$m = 0.2$	$m = 0.8$	$m = 0.9$
1	$.258 \times 10^{-22}$	$.116 \times 10^{-22}$	$.165 \times 10^{-22}$	$.423 \times 10^{-22}$
2	$.448 \times 10^{-10}$	$.208 \times 10^{-10}$	$.203 \times 10^{-10}$	$.434 \times 10^{-10}$
3	$.181 \times 10^{-6}$	$.889 \times 10^{-7}$	$.871 \times 10^{-6}$	$.176 \times 10^{-6}$
4	$.118 \times 10^{-4}$	$.601 \times 10^{-5}$	$.594 \times 10^{-5}$	$.116 \times 10^{-4}$
5	$.142 \times 10^{-3}$	$.750 \times 10^{-4}$	$.744 \times 10^{-4}$	$.140 \times 10^{-3}$
10	.01224	.00742	.00741	.01222
20	.03525	.02711	.02711	.03526
30	.02412	.02293	.02294	.02413
40	.01331	.01540	.01541	.01331
50	.00705	.00986	.00987	.00705
100	$.279 \times 10^{-3}$	$.995 \times 10^{-3}$	$.996 \times 10^{-3}$	$.279 \times 10^{-3}$
150	$.110 \times 10^{-4}$	$.100 \times 10^{-3}$	$.100 \times 10^{-3}$	$.110 \times 10^{-4}$
$u(p)$	.00055	.00272	.4894	.5621

the fixation time of a completely dominant mutation ($s_2 = 0$) is the same as that of a completely recessive mutation ($s_1 = 0$) when $N(s_1 + s_2)$ is the same for the two cases.

DISCUSSION

We have seen that disadvantageous overdominant genes ($W_{AA} > W_{aa}$) are fixed in a population generally with a low probability. However, if m is close to 0.5, the probability of fixation is higher than that for neutral genes in relatively small populations. If such a gene is fixed, the fitness of the population is expected to decline by the amount $W_{AA} - W_{aa}$, compared with the fitness before the mutant gene is introduced. On the other hand, advantageous overdominant genes ($W_{AA} < W_{aa}$) always have a higher probability of fixation than neutral genes. In this case overdominance enhances the probability of fixation considerably. In large populations the probability is determined not by $W_{aa} - W_{AA}$ but by $W_{Aa} - W_{AA}$. If $4Ns_1 >> 1$, it is roughly $2s_1 \cong 2(W_{Aa} - W_{aA})$. This has an evolutionary implication. If an advantageous mutation has a slight overdominant effect in the heterozygous condition, it has a higher probability of fixation than a completely dominant mutation when $W_{aa} - W_{AA}$ remains the same.

One of the interesting results in this study is that the mean fixation time of a mutant gene is highly dependent on the value of m. If m is within the range of approximately 0.2 to 0.8, the fixation time is longer than that of neutral genes in relatively small populations, while if m is outside the range, it is shorter. This suggests that the contribution of the first group of genes ($0.2 < m < 0.8$) to the genetic variability of a population is much larger than that of the second group of genes ($m < 0.2$ or $m > 0.8$). Thus, it is likely that the majority of overdominant genes found in natural populations is of the first group. A similar conclusion has been derived by ROBERTSON (1962) in his study on the rate of decay of genetic variability at steady state.

The mean fixation times of a mutant gene with $m < 0.2$ and of a gene with $m > 0.8$ are both shorter than that of neutral genes in relatively small populations. However, the former is fixed in the population with a very small probability, so that it is unimportant in practice. On the other hand, the probability of fixation of advantageous genes with $m > 0.8$ is enhanced by overdominance, as mentioned above. It is interesting to note that the fixation of such genes is not retarded but accelerated by overdominance.

EWENS and THOMSON (1970) studied the mean time to loss or fixation of an overdominant allele in finite populations. This mean time ($\bar{t}(p)$) is related to the mean fixation time ($\bar{t}_1(p)$) by $\bar{t}(p) = u_h(p)\bar{t}_1(p) + \{1 - u_h(p)\}\bar{t}_0(p)$, where $\bar{t}_0(p)$ is the mean loss time (KIMURA 1971). When $p = m$, they found that the dependence of $t(m)$ on m and $N(s_1 + s_2)$ is very similar to that of ROBERTSON's retardation factor. This similarity, however, can be easily explained, since the phenomena studied by ROBERTSON (1962) and EWENS and THOMSON (1970) are essentially the same. The dependence of our mean fixation time ($\bar{t}_1(p)$) on m and $N(s_1 + s_2)$ is also similar to that of retardation factor; but, as we have already

discussed, the explanation seems to be quite different from that of retardation factor.

This study was supported by NSF grant GB-21224, PHS grants GM-17719 and GM-20293. We thank DR. MOTOO KIMURA for his valuable comments.

LITERATURE CITED

ABRAMOWITZ, M. and A. STEGUN, 1964 *Handbook of Mathematical Functions.* National Bureau of Standards, Washington, D. C.

CARR, R. N. and R. F. NASSAR, 1970a Effects of selection on drift on the dynamics of finite populations. I. Ultimate probability of fixation of a favorable allele. Biometrics **26:** 41–49. ——, 1970b II. Expected time to fixation or loss of an allele. Biometrics **26:** 221–227.

Ewens, W. J., 1972 Concepts of substitutional load in finite populations. Theoret. Population Biol. **3:** 153–161.

EWENS, W. J. and C. THOMSON, 1970 Heterozygote selective advantage. Ann. Hum. Genet. **33:** 365–376.

FISHER, R. A., 1930 *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

HALDANE, J. B. S., 1927 A mathematical theory of natural and artificial selection. Part V. Selection and mutation. Proc. Camb. Phil. Soc. **28:** 838–844.

KIMURA, M., 1957 Some problems of stochastic processes in genetics. Ann. Math. Stat. **28:** 882–902. ——, 1962 On the probability of fixation of mutant genes in a population. Genetics **47:** 713–719. ——, 1971 Theoretical foundation of population genetics at the molecular level. Theoret. Pop. Biol. **2:** 174–178.

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.

MARUYAMA, T. and M. KIMURA, 1971 Some methods for treating continuous stochastic processes in population genetics. Japan J. Genet. **46:** 407–410.

ROBERTSON, A., 1962 Selection for heterozygotes in small populations. Genetics **47:** 1291–1300.

TRACEY, M. L., 1972 Sex chromosome translocations in the evolution of reproductive isolation. Genetics **72:** 317–333.

WRIGHT, S., 1942 Statistical genetics and evolution. Bull. Amer. Math. Soc. **48:** 223–246.