Proc. Natl. Acad. Sci. USA
Vol. 76, No. 6, pp. 2858–2861, June 1979
Genetics

Fixation of a deleterious allele at one of two "duplicate" loci by mutation pressure and random drift

(population genetics/mutational load/gene duplication/diffusion equation method)

MOTOO KIMURA* AND JACK L. KING†

*National Institute of Genetics, Mishima 411, Japan; and †Department of Biological Science, University of California, Santa Barbara, California 93106

Contributed by Motoo Kimura, April 4, 1979

ABSTRACT **We consider a diploid population and assume two gene loci with two alleles each, *A* and *a* at one locus and *B* and *b* at the second locus. Mutation from wild-type alleles *A* and *B* to deleterious alleles *a* and *b* occurs with mutation rates v_a and v_b, respectively. We assume that alleles are completely recessive and that only the double recessive genotype *aabb* shows a deleterious effect with relative fitness $1 - \epsilon$. Then, it can be shown that if $v_a > v_b$ mutant *a* becomes fixed in the population by mutation pressure and a mutation–selection balance is ultimately attained with respect to the *B*/*b* locus alone. The main aim of this paper is to investigate the situation in which $v_a = v_b$ exactly. In this case a neutral equilibrium is attained and either locus can drift to fixation for the mutant allele. Diffusion models are developed to treat the stochastic process involved whereby the deleterious mutant eventually becomes fixed in one of the two duplicated loci by random sampling drift in finite populations. In particular, the equation for the average time until fixation of mutant *a* or *b* is derived, and this is solved numerically for some combinations of parameters $4N_e v$ and $4N_e\epsilon$, where v is the mutation rate ($v_a = v_b \equiv v$) and N_e is the effective size of the population. Monte Carlo experiments have been performed (using a device termed "pseudo sampling variable") to supplement the numerical analysis.**

It is expected that, in natural populations, deleterious alleles are constantly arising by mutations at every gene locus. However, frequencies of such alleles are kept low by natural selection, unless the deleterious effects are extremely small (smaller than the reciprocal of the effective population size).

It is common sense in population genetics that the mutation pressure can not overcome the barrier of negative selection. An interesting exception to this rule occurs when the environment changes in such a way that the wild-type allele at one locus becomes no longer necessary. Then, amorphic mutations at that locus become selectively neutral and mutation pressure, in conjunction with random genetic drift, can lead one of such alleles to eventual fixation in the population. The loss of vitamin C synthesizing ability in several vertebrates can be explained in the light of such considerations (1). Gene duplications also create conditions that enable random drift to operate much more prominently on recessive deleterious mutants than what was possible before duplication (as we shall show later). This allows fixation of mutants that are slightly deleterious for contemporary conditions but which may have other useful effects for adaptation to a new environment. Here again we find that the paradigm of the neutral theory (2) or "non-Darwinian" view (3) gives adequate explanation for phenomena relating to progressive evolution.

The present paper consists of two parts. In the first part, we discuss a situation in which mutation pressure leads to fixation of a deleterious allele. In the second part, which is the main part of this paper, we present a new treatment for the problem of random fixation of a nonfunctional allele at one of two loci after duplication. Since the present paper was submitted, we discovered that the results of the first part had already been obtained by Christiansen and Frydenberg (4), so in this revised version we present our results of this part only in an abbreviated form in order to serve as an introduction to the second part.

FIXATION BY MUTATION PRESSURE

Let us consider a random mating diploid population. We assume that a pair of alleles *A* and *a* are segregating in the first locus and alleles *B* and *b* are segregating in the second locus. We shall refer to *a* and *b* as mutant alleles. To simplify the situation as much as we can, we make additional assumptions. We assume that the population is large enough so that random fluctuation of gene frequencies can be neglected (this assumption will be removed in the next section). We also assume that genes in two loci combine completely at random. Strictly speaking, under epistatic interaction in fitness, nonrandom association or linkage disequilibrium will develop, particularly if linkage is tight. However, if selection coefficients involved are small and if the linkage is loose, quasilinkage equilibrium (5–7) will be realized, and we can neglect the linkage disequilibrium without serious error. Now, we assume that relative fitnesses of various genotypes are as given in Table 1, in which s is the selection coefficient against a single recessive mutation and ϵ denotes the epistatic effect in fitness ($s \geqq 0$, $\epsilon \geqq 0$). In other words, ϵ represents the excess selection against the double homozygote over that expected with multiplicative interaction. Then the differential equations giving the rates of change of mutant allele frequencies are

$$dp/dt = (1 - p)\{v_a - p^2[s(1 - sq^2) + \epsilon q^2]\} \text{ and}$$
$$dq/dt = (1 - q)\{v_b - q^2[s(1 - sp^2) + \epsilon p^2]\} \qquad [1]$$

where t denotes time in generations. It is assumed here that the selection is weak and linkage is loose so that assumption of complete random combination of genes is essentially valid. From these equations, we can see that if $s > 0$ the mutation–selection balance will be realized at both loci, unless s is extremely small (at least as small as the mutation rates v_a and v_b).

Let us assume, then, that the mutant alleles are completely recessive so that $s = 0$. In this case, mutant alleles are not only completely recessive but are also completely hypostatic to normal alleles at another locus. In other words, only the double recessive *aabb* is deleterious with selection coefficient ϵ. This

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviation: PSV, pseudo sampling variable.

Table 1. Fitness of various genotypes for the diploid model

	$(1-p)^2$ *AA*	$2(1-p)p$ *Aa*	p^2 *aa*	Marginal w
$(1-q)^2$ *BB*	1	1	$1-s$	$1-sp^2$
$2(1-q)q$ *Bb*	1	1	$1-s$	$1-sp^2$
q^2 *bb*	$1-s$	$1-s$	$(1-s)^2-\epsilon$	$(1-s)(1-sp^2)-\epsilon p^2$
Marginal w	$1-sq^2$	$1-sq^2$	$(1-s)(1-sq^2)-\epsilon q^2$	$\bar{w}=(1-sp^2)(1-sq^2)-\epsilon p^2q^2$

The letter w stands for the selective value.

type of epistasis is known as "duplicate" gene in classical genetics; selfing of the double heterozygote *AaBb* (F_1) leads to 15:1 segregation in F_2 (for example, see ref. 8). In this case, Eqs. 1 reduce to

$$dp/dt = (1-p)(v_a - \epsilon p^2q^2) \text{ and}$$

$$dq/dt = (1-q)(v_b - \epsilon p^2q^2). \quad [2]$$

We can then show that, if $v_a > v_b$, the mutant allele *a* increases by mutation pressure to reach fixation ($p = 1$); mutation–selection balance will be reached for the allele *b* ($q = \sqrt{v_b/\epsilon}$). In Fig. 1, courses of the frequencies of mutant alleles in an infinitely large population are illustrated assuming mutation rates $v_a = 2 \times 10^{-5}$, $v_b = 1 \times 10^{-5}$, and selection coefficient $\epsilon = 1 \times 10^{-3}$ and taking $1/\epsilon$ (= 1000 generations) as the unit length of time ($T = t\epsilon$). These are constructed based on the numerical solution of differential Eqs. 2 by the Runge–Kutta method.

In the case of equal mutation rates at both loci ($v_a = v_b \equiv v$), the situation is quite different, and neutral equilibrium will be reached so that $p^2q^2 = v/\epsilon$. Then, mutant frequencies are subject to random drift, and in a finite population the deleterious mutant will eventually become fixed in one of the two loci. We shall investigate the stochastic process involved in the next section.

FIXATION BY RANDOM DRIFT

Let us consider a diploid population of effective size N_e and use the selection model as shown in Table 1 but with $s = 0$ so that only the double mutant homozygote, *aabb*, is deleterious with selective disadvantage ϵ. To treat the process of change of mutant frequencies, we make use of the diffusion equation method (9). Because the analytical solution is difficult to obtain for the present case, we resorted to numerical solution of the partial differential equation involved, for which the senior author (M.K.) is responsible. He is also responsible for Monte Carlo experiments performed to supplement the numerical analysis.

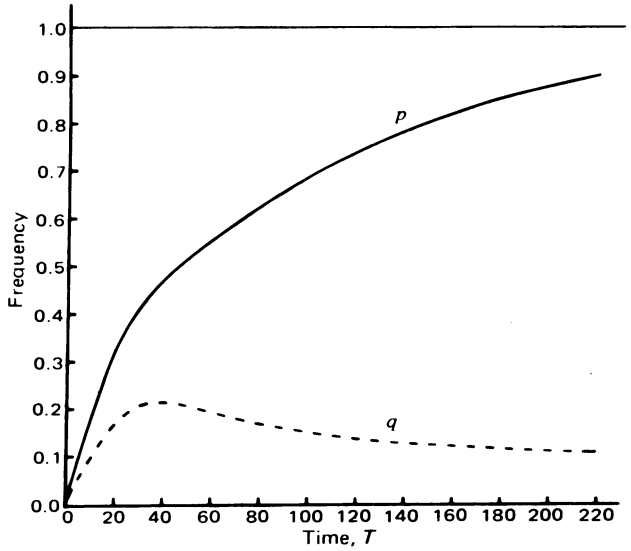

FIG. 1. Courses of change in the frequencies of deleterious alleles by mutation pressure (see text for details). —, For allele *a*; - - -, for allele *b*. Both lines start from $p_0 = q_0 = 0$.

Let $u(p, q; t)$ be the probability of fixation of mutant allele *a* or *b* by the *t*th generation, given that the initial (i.e., at $t = 0$) frequencies are *p* and *q* for *a* and *b*, respectively. Then, assuming random combination of alleles at both loci, we can show that $u(p, q; t)$ satisfies the following Kolmogorov backward equation:

$$\frac{\partial u}{\partial t} = \tfrac{1}{2} V_{\delta p} \frac{\partial^2 u}{\partial_p{}^2} + \tfrac{1}{2} V_{\delta q} \frac{\partial^2 u}{\partial q^2} + M_{\delta p} \frac{\partial u}{\partial p} + M_{\delta q} \frac{\partial u}{\partial q} \quad [3]$$

in which

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

$$M_{\delta p} = (1-p)(v - \epsilon p^2q^2),\ M_{\delta q} = (1-q)(v - \epsilon p^2q^2), \quad [4]$$

and *u* stands for $u(p, q; t)$. For the rationale of Eq. 3, see Crow and Kimura (ref. 10, p. 429). Let $\bar{T}(p, q)$ be the average time until fixation of mutant *a* or *b* in the population such that

$$\bar{T}(p, q) = \int_0^\infty t|\partial u(p, q; t)/\partial t|dt. \quad [5]$$

Then, using Eq. 3, we can show that $\bar{T}(p, q)$ satisfies the following elliptic equation.

$$\frac{1}{2} V_{\delta p} \frac{\partial^2 \bar{T}}{\partial p^2} + \frac{1}{2} V_{\delta q} \frac{\partial^2 \bar{T}}{\partial q^2} + M_{\delta p} \frac{\partial \bar{T}}{\partial p} + M_{\delta q} \frac{\partial \bar{T}}{\partial q} + 1 = 0. \quad [6]$$

Let us denote by $y(p, q)$ the average time until fixation measured with $4N_e$ generations as the unit, so that $\bar{T}(p, q) = 4N_e y(p, q)$. Then *y* satisfies the equation

$$p(1-p)y_{pp} + q(1-q)y_{qq} + (V - Ep^2q^2) \times \{(1-p)y_p + (1-q)y_q\} + 1 = 0 \quad [7]$$

in which $V = 4N_e v$, $E = 4N_e \epsilon$, and y_p, y_{pp}, etc., stand for $\partial y/\partial p$, $\partial^2 y/\partial p^2$, etc. Note that $y(p, q)$ depends only on the products $4N_e v$ and $4N_e \epsilon$ but not on N_e, *v*, and ϵ separately. The appropriate boundary conditions are

$$y(1, q) = y(p, 1) = 0 \quad [8a]$$

and

$$y(0, q) = \text{finite},\ y(p, 0) = \text{finite}. \quad [8b]$$

To apply the numerical method, we cover the domain ($0 \leqq p \leqq 1$, $0 \leqq q \leqq 1$) by $n \times n$ square meshes each with side length *h* ("mesh size"). Let $p = hi$ and $q = hj$, in which *i* and *j* are integers ($i, j = 0, 1, \ldots, n$). Then Eq. 7 may be converted to a

set of finite difference equations in which symbol $y_{i,j}$ is used to represent $y(ih, jh)$. Boundary conditions 8 can be incorporated as follows. Condition 8a is straightforward and we have $y_{n,j} = y_{i,n} = 0$. Condition 8b is more subtle, but we can replace it by $y_{0,j} = y_{1,j} + (y_{1,j} - y_{2,j})$ and $y_{i,0} = y_{i,1} + (y_{i,1} - y_{i,2})$.

For given values of V and E, this set of equations for $y_{i,j}$ can be solved numerically by using a computer (details will be published elsewhere). We chose mesh size $h = 0.1$, and the Gauss–Seidel method was used (for the numerical solution of partial differential equations see, for example, ref. 11, p. 391). The solid curve in Fig. 2 illustrates, for the case $4N_e\upsilon = 2$, the average time until fixation of mutant a or b, starting from $p = q = 0$, as function of $4N_e\epsilon$ for value of $4N_e\epsilon < 50$. In other words, the ordinate represents $\overline{T}(0, 0)$ or $4N_e y(0, 0)$.

In order to supplement these results, Monte Carlo simulation experiments were performed with assumed various population sizes (N_e) ranging from 25 to 2500, $4N_e\upsilon = 2$, and $4N_e\epsilon$ ranging from 1 to 10^4. Note that $\overline{T}/(4N_e)$ depends only on $4N_e\upsilon$ and $4N_e\epsilon$. Starting from $p = q = 0$, the average time until fixation of a or b was investigated. (In nature, newly tetraploid populations presumably are the immediate descendants of one or a few individuals and can be assumed to reach the effective population size N_e quickly before accumulating any mutant alleles.) Each solid circle in Fig. 2 represents the average of 100 replicate trials, except for a few cases ($4N_e\epsilon > 1000$) for which each point is the average of 10–50 replicate trials. The broken line at the tail of the solid curve for $4N_e\upsilon = 2$ represents values of $\overline{T}(0, 0)$ inferred from these simulation experiments for higher values of $4N_e\epsilon$. Thus, the broken-line curve represents crude approximation values only. Fig. 2 also shows the results of simulation for $4N_e\upsilon = 0.4$ (each open circle is the average of 1000 replicate trials). In the simulation experiments, each generation consists of random sampling drift followed by mutation and selection.

To simulate the gene frequency change by random sampling of gametes in one generation, instead of actually sampling gametes $2N_e$ times as is usually done in Monte Carlo experiments in population genetics (for example, see ref. 12), we simply generated a random number (called a "pseudo sampling variable" or PSV) and a realized value of this variable was added to the gene frequency (p) to produce the frequency (p') after sampling drift. The essential point is that it is a uniform random number that has mean = 0 and variance $\sigma^2 = p(1 - p)/(2N_e)$. In other words, if ξ_{PSV} is a PSV, then $\xi_{PSV} = \sqrt{3\sigma^2}\,U_1$, in which U_1 is a random variable that follows a uniform distribution between -1 and $+1$ and is commonly used in Monte Carlo experiments. If p' ($= p + \xi_{PSV}$) becomes negative by chance, which may sometimes happen when p is near zero, then p' is set to zero to continue the experiment. On the other hand, if p' becomes larger than $1 - 1/(2N_e)$, p' is set equal to unity and the run is ended. The reason why PSV can substitute for the actual sampling comes from the nature of the continuous stochastic process—namely, only the mean and the variance (but not the detailed shape of the distribution) of the change per generation determine the process, as long as the higher moments are negligible in magnitude (see ref. 10, p. 374). Note that this scheme of pseudo sampling simulates the diffusion process itself rather than the discrete, binomial sampling process, for which the diffusion model is usually regarded as an approximation.

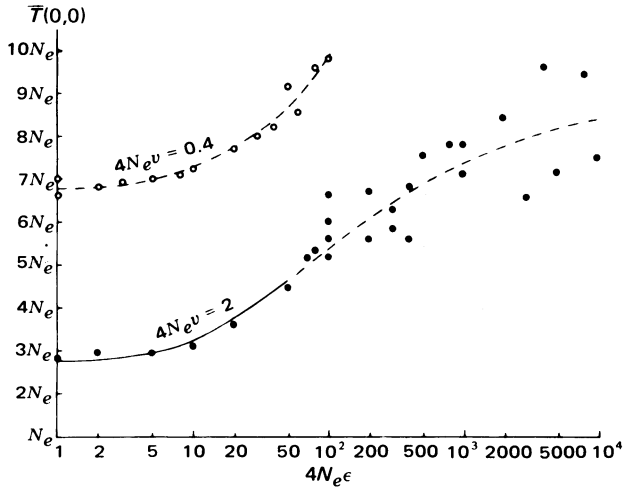

FIG. 2. Average time until fixation of a deleterious mutant at one of two "duplicate loci." Relationship between the average time, $\overline{T}$ (0, 0), and $4N_e\epsilon$ is illustrated for two cases, $4N_e\upsilon = 2$ and $4N_e\upsilon = 0.4$. N_e, effective population size; ϵ, selection coefficient against the double mutant individual; υ, mutation rate per locus (see text for details).

Finally, we should remark that, because the time until fixation of a or b has a large standard deviation around its mean ($\overline{T}$), this mean time, in very rough sense, may represent the time by which fixation occurs in half of the cases—i.e., $T_{0.5} \approx \overline{T}$. Actually, the distribution of fixation time has a positive skewness of roughly unity, so that $T_{0.5}$ is somewhat smaller than $\overline{T}$.

DISCUSSION

It is clear that the remarkable phenomenon of a deleterious allele reaching fixation in one of the two loci is possible only when the "duplicate" type epistasis is complete in fitness. In addition, mutation rates (υ_a and υ_b) must be significantly different at the two loci for the mutation pressure to control the process deterministically. Under these conditions, mutant a at the first locus goes to fixation if $\upsilon_a > \upsilon_b$ and b reaches mutation–selection balance, and vice versa. This confirms the results obtained earlier by Christiansen and Frydenberg (4). The rate of increase of a is roughly proportional to the difference in mutation rates, $\upsilon_a - \upsilon_b$. In other words, a is pushed by the pressure that comes from the excess mutation rate.

From the standpoint of evolution, the most likely origin of "duplicate" type epistasis is gene duplication, especially as it occurs in the formation of allotetraploids. Considering the prevalence of gene duplication in evolution, it might be expected that "duplicate" genes are commonly found in plants and, to a lesser extent, in animals. Duplicate genes are indeed not uncommon, but loss of duplicate genes and reversion to functional diploidy are certainly common and perhaps more usual (13).

Where duplicate genes persist, it is of course possible that complete hypostasis and recessivity of deleterious mutants may be lacking. Slightly deleterious mutants in *Drosophila* usually show considerable dominance, so that they are mainly selected against in the heterozygous condition (14, 15). For "null" mutants at enzyme loci, on the other hand, it is possible that the heterozygotes with the wild-type (active) allele are so nearly normal that we can regard the mutant alleles, for practical purposes, as being completely recessive. Presumably, such mutants at duplicate enzyme loci might also be completely hypostatic. However, unless we see evidence to the contrary, it seems likely that mutation rates stay the same in duplicated loci, and the effect of unequal mutation rates at duplicate loci may seldom be important in and of itself.

The situation is less clear for unequal epistasis—e.g., where the B allele is completely epistatic but the A allele is not. If multiple alleles are considered, it seems likely that slightly hypomorphic alleles might occur and increase at one or the

other locus, through mutation and drift in the absence of strong selection. Once one duplicate locus had such a slightly deleterious allele either fixed or in a significantly high frequency, the symmetry of selection would be ended and additional, more severe, mutant alleles should also increase in frequency at the same locus.

Loss of gene expression in one or the other of duplicated loci has recently been reported in some groups of fish. In this case, fixation of "null" alleles by mutation and drift appears to be the most likely explanation. According to Allendorf (13), species in both salmonid and catostomid fish have lost approximately 50% of the gene duplication produced by tetraploidy. We can show that the time needed for such an evolutionary loss ($T_{0.5}$) depends much on the mutation rate, assuming that the population starts from the state having no null mutants. This assumption may be realistic because the tetraploidization must have started from a single individual or a few related individuals free of null alleles. As shown in Fig. 2, if $2N_e v = 1$ (i.e., one new mutation appears in each generation), it takes roughly $9N_e$ generations if $4N_e\epsilon = 10{,}000$. If the mutation rate for appropriate null alleles is much lower, the time needed for 50% loss must be longer. Our rough estimation based on Monte Carlo experiments suggests that, if $4N_e v = 0.1$ and $4N_e\epsilon = 1000$, the average time ($\bar{T}$) until fixation of a null allele in one of the loci is about $20N_e$ generations which must be somewhat larger than the time needed for 50% loss. These results do not seem to agree with those of Bailey *et al.* (16) who claim that $T_{0.5} \approx 15N_e + v^{-3/4}$. Clearly, $T_{0.5}$ can be much shorter than $15N_e$ if $4N_e v$ is larger than unity. A more detailed study of the problem including the situation in which epistasis is not complete has been done by N. Takahata and T. Maruyama (personal communication), and the results will be published in the near future. In particular, they found that $T_{0.5}$ becomes much larger if null alleles show small deleterious effects in combinations other than the double null homozygote.

We thank Dr. F. W. Allendorf for calling our attention to the paper of Christiansen and Frydenberg (ref. 4). Thanks are also due to Drs. T. Maruyama and N. Takahata for stimulating discussions. This is contribution no. 1247 from the National Institute of Genetics, Mishima, Shizuoka-ken 411, Japan.

1. Jukes, T. H. & King, J. L. (1975) *J. Hum. Evol.* **4**, 85–88.
2. Kimura, M. (1968) *Nature (London)* **217**, 624–626.
3. King, J. L. & Jukes, T. H. (1969) *Science* **164**, 788–798.
4. Christiansen, F. B. & Frydenberg, O. (1977) *Am. J. Hum. Genet.* **29**, 195–207.
5. Kimura, M. (1965) *Genetics* **52**, 875–890.
6. Wright, S. (1967) *Proc. Natl. Acad. Sci. USA* **58**, 165–172.
7. Nagylaki, T. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 526–530.
8. Darlington, C. D. & Mather, K. (1949) *The Elements of Genetics* (Allen & Unwin, London).
9. Kimura, M. (1964) *J. Appl. Probab.* **1**, 177–232.
10. Crow, J. F. & Kimura, M. (1970) *An Introduction to Population Genetics Theory* (Harper & Row, New York).
11. Todd, J., ed. (1962) *Survey of Numerical Analysis* (McGraw-Hill, New York).
12. Kimura, M. & Ohta, T. (1969) *Genetics* **61**, 763–771.
13. Allendorf, F. W. (1978) *Nature (London)* **272**, 76–78.
14. Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. (1972) *Genetics* **72**, 335–355.
15. Simmons, M. J. & Crow, J. F. (1977) *Annu. Rev. Genet.* **11**, 49–78.
16. Bailey, G. S., Poulter, R. T. M. & Stockwell, P. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5575–5579.