

The Effects of Random Selection on Gene Frequency

HENRY C. TUCKWELL*

*Department of Biophysics and Theoretical Biology,
University of Chicago, Chicago, Illinois 60637.*

Communicated by Robert M. May

ABSTRACT

The equations satisfied by the gene frequency for various cases of dominance are converted to the type of stochastic differential equation associated with a diffusion process. Using the physical approach to stochastic integrals, the solutions of the corresponding Fokker-Planck equations are obtained. Conditions for quasifixation are investigated and observed frequencies of the gene *medionigra* of *Panaxia dominula* are considered in relation to the model of random selection without dominance.

1. INTRODUCTION

Stochastic changes in the gene frequencies of populations are attributable to random sampling of gametes or random fluctuations in the pressures of selection and migration. Wright [1] pointed out that in real populations the relative importance of these various sources of randomness may be quite different for different genes.

In this work we wish to study the effects of random selection on gene frequency. In order to do this we will assume that the population size is large enough to make the effects of random sampling of gametes negligible. We will also suppose that the population considered is sufficiently isolated from others of the same species that the effects of migration, whether deterministic or stochastic, may be ignored. Thus the only source of variability will be the fluctuations in the relative fitnesses of the genotypes. Furthermore, because the population size is assumed to be large, the gene frequencies may be studied through their continuous approximations.

The first study of the temporal evolution of gene frequency distributions in the case of randomly varying selection was performed by Kimura [2]. If $X(t)$ denotes an allele frequency at time t , Kimura found an analytic

*Present address: University of British Columbia, Vancouver, B.C., Canada.

expression for its transition probability density function, $\phi(x, t|x_0)$, defined through the relation

$$\phi(x, t|x_0) dx = \Pr\{x < X(t) < x + dx | X(0) = x_0\} \quad (1)$$

for the case of two alleles in a haploid population when the mean value of the selection coefficient was zero. A study of the asymptotic behavior of this function, as t became infinite, revealed that there was eventually an accumulation of probability mass near the fixed classes ($x=0$ and $x=1$) but that these states could not be reached in a finite time. The terms "quasifixation" and "quasiloss" were introduced by Kimura to classify these phenomena, which were shown to occur, under the circumstances mentioned, with probabilities x_0 and $1-x_0$ respectively.

Expressions for the transition density functions could not be obtained, however, for a nonzero mean selection coefficient, nor for any of the cases of dominance in diploid populations. Hence it was not possible to determine the solution of the important problem, raised by Kimura [3], of whether quasifixation was ever likely for an allele which was, on the average, at a selective disadvantage.

Recent progress has been made by Gillespie [4, 5] in finding transition probability densities in the haploid case for both uncorrelated and autocorrelated selection coefficients. Studies of the simultaneous effects of random selection and random sampling have been undertaken by Jensen and Pollak [6], Ohta [7] and Jensen [8]. In these works the diffusion approximation was employed. Ohta used the "classical" infinitesimal moments of the change in gene frequency and calculated the probability of ultimate fixation with the techniques introduced by Kimura [9]. Jensen and Pollak have developed different formulae for the infinitesimal moments which appear in the Fokker-Planck equation satisfied by the transition probability density of the gene frequency.

In a more recent study of the effects of random selection (without any other source of randomness) in haploid populations, Cook and Hartl [10, 11] have constructed a large number of models, divided into the two main classes of discrete, non-overlapping generations (model I) and continuous, overlapping generations (model II). Our approach is different in that we deal directly with continuous random processes, thus avoiding the necessity of carrying out limiting operations to generate them.

Our main object is to determine how the probability that a gene becomes quasifixated in a population depends on its initial frequency and the parameters of the fluctuating selection coefficient. To do this we will model the effects of random selection by the stochastic differential equations which result when the selection coefficient appearing in the deterministic differential equations satisfied by the gene frequency is considered to be a certain

type of random process. We will not, as have previous studies, restrict our attention to the case of haploid populations or diploid populations without dominance, and will consider two cases of complete dominance.

2. EQUATIONS FOR THE GENE FREQUENCY

Consider a locus with two alleles, A_1 and A_2 say, in a diploid population, and let the frequency of A_1 at time t be $X(t)$. With the assumption of random mating, differential equations satisfied by $X(t)$ have been derived for various degrees of dominance, as for example in Crow and Kimura [12]. There are three cases we will consider.

(i) *No dominance:*

$$\frac{dX(t)}{dt} = \frac{1}{2}sX(t)[1 - X(t)]. \quad (2)$$

(ii) *Complete dominance of A_1 , A_1 favored:*

$$\frac{dX(t)}{dt} = sX(t)[1 - X(t)]^2. \quad (3)$$

(iii) *Complete dominance of A_2 , A_1 favored:*

$$\frac{dX(t)}{dt} = sX(t)^2[1 - X(t)]. \quad (4)$$

In these equations, s is the measure of the difference in the fitnesses of the two homozygous classes.

The effects of randomly varying selection may be studied by regarding the selection coefficient as a random process, $s(t)$, whereupon we obtain, in place of Eqs. (2) through (4), corresponding stochastic differential equations. If $s(t)$ is assumed to be a Gaussian, stationary delta-correlated process ("white noise"), then the random processes $X(t)$, representing the gene frequencies, are diffusion processes of the kind studied by Kimura [13]. We note that if $s(t)$ is an autocorrelated process, then the resulting processes $X(t)$ are no longer Markov. For the moment we will let the expectation of $s(t)$ be a constant, \bar{s} , and assume that its covariance kernel is $V_s\delta(t_1 - t_2)$. In the next section we will consider the case where the mean and the variance parameter of $s(t)$ are functions of time.

If we treat the selection coefficient as the above kind of random process, then we should write formally the stochastic version of Eq. (2), for example,

$$dX(t) = \frac{1}{2}\bar{s}X(t)[1 - X(t)]dt + \frac{1}{2}X(t)[1 - X(t)]dW(t), \quad (5)$$

where $W(t)$ is a Wiener process of variance parameter V_s . The solution of this equation can be written

$$X(t) = X(0) + \frac{1}{2}\bar{s} \int_0^t X(t')[1 - X(t')] dt' + \frac{1}{2} \int_0^t X(t')[1 - X(t')] dW(t'), \quad (6)$$

where $X(0)$ is the random variable representing the initial gene frequency, and the second integral is a stochastic integral with respect to the Wiener process [14].

In this last mentioned reference one will find that there are two definitions of this stochastic integral, one being due to Ito [5] and the other due to Stratonovich [16]. The choice of the definition has an important bearing on investigations of stochastic models for the evolution of the genetic structure of populations. We will see in the next section that the results are quite different according to which definition one employs. This point was emphasized in the context of stochastic models of population growth by May [17]. The difference in the two approaches is brought out by noting that the Fokker-Planck equation satisfied by the transition probability density function $\phi(x, t|x_0)$,

$$\frac{\partial \phi}{\partial t} = - \frac{\partial}{\partial x} (K_1(x)\phi) + \frac{1}{2} \frac{\partial^2}{\partial x^2} [K_2(x)\phi], \quad (7)$$

has the same diffusion coefficient $K_2(x)$ for both definitions, but that the drift term $K_1(x)$ depends on which definition is adopted.

To illustrate this, we have for random selection without dominance, as modeled by Eqs. (5) and (6), in both calculi,

$$K_2(x) = \frac{1}{4} V_s x^2 (1-x)^2, \quad (8)$$

whereas the drift term is

$$K_1(x) = \begin{cases} \frac{1}{2} \bar{s} x (1-x) & (9A) \\ \frac{1}{2} \bar{s} x (1-x) + \frac{1}{8} V_s x (1-x) (1-2x), & (9B) \end{cases}$$

where (9A) corresponds to using Ito's integral and (9B) corresponds to using Stratonovich's integral in Eq. (6).

With regard to the non-uniqueness which arises in integrating the stochastic equations for the gene frequency, we note the following.

(i) The infinitesimal moments (drift and diffusion terms) employed in the Fokker-Planck equations of Kimura and Wright were always those

which one would obtain if one employed Ito's definition of stochastic integral.

(ii) It has been fairly well established (see, for example, the comments of Gray and Caughey [18] or Jaswinski [14], but see also Mortensen [19] for an interesting but less conclusive discussion) that when stochastic equations of the kind we are considering are being employed to model a random process in the physical world, then those equations should be integrated according to the calculus of Stratonovich. This conjecture is supported by a convergence theorem of Wong and Zakai [20], and though we will accept it here, one cannot assert that the problem of non-uniqueness is completely resolved.

In subsequent sections we will, therefore, assume that the physical approach (i.e., Stratonovich integral) is the one to employ. We will then readily find solutions of the Fokker-Planck equations for the stochastic versions of (2) through (4). We point out that in the general case

$$dX(t) = \bar{s}g(X(t))dt + g(X(t))dW(t), \quad (10)$$

the first infinitesimal moment is

$$K_1(x) = \bar{s}g(x) + \frac{1}{2}V_s g(x)g'(x), \quad (11)$$

where the prime denotes differentiation, and the diffusion coefficient is given by

$$K_2(x) = V_s g(x)^2. \quad (12)$$

3. SOLUTIONS OF THE FOKKER-PLANCK EQUATIONS

When a stochastic differential equation is of the form (10), and certain conditions are fulfilled, the solution of the corresponding Fokker-Planck equation can be obtained quite readily. The details of the method and some applications to population growth have been given previously [21]. One makes the transformation

$$Y(X) = \int^X g(X')^{-1} dX' \quad (13)$$

and notes that $Y(t)$ satisfies the equation for a Wiener process with drift:

$$dY(t) = \bar{s} dt + dW(t). \quad (14)$$

Thus, if the initial value of $Y(t)$ is the fixed value y_0 , then its transition probability density is

$$\psi(y, t|y_0) = \frac{1}{\sqrt{2\pi V_s t}} \exp \left[\frac{-(y - y_0 - \bar{s}t)^2}{2V_s t} \right]. \quad (15)$$

It therefore follows that the transition probability density of the original process $X(t)$ is

$$\phi(x, t|x_0) = \frac{|g(x)|^{-1}}{\sqrt{2\pi V_s t}} \exp \left[\frac{-\left(\int_{x_0}^x g(x')^{-1} dx' - \bar{s}t \right)^2}{2V_s t} \right]. \quad (16)$$

The stochastic equations for the gene frequency meet the requirements for the application of this method, so we substitute the appropriate functions $g(\cdot)$ and obtain the following results.

(i) *No dominance:*

$$\phi_1(x, t|x_0) = \frac{2x^{-1}(1-x)^{-1}}{\sqrt{2\pi V_s t}} \exp \left[\frac{-\left(2 \log \left\{ \frac{x(1-x_0)}{x_0(1-x)} \right\} - \bar{s}t \right)^2}{2V_s t} \right]. \quad (17)$$

(ii) *Complete dominance, dominant favored:*

$$\phi_2(x, t|x_0) = \frac{x^{-1}(1-x)^{-2}}{\sqrt{2\pi V_s t}} \exp \left[\frac{-\left(\frac{x-x_0}{(1-x)(1-x_0)} + \log \left\{ \frac{x(1-x_0)}{x_0(1-x)} \right\} - \bar{s}t \right)^2}{2V_s t} \right]. \quad (18)$$

(iii) *Complete dominance, recessive favored:*

$$\phi_3(x, t|x_0) = \frac{x^{-2}(1-x)^{-1}}{\sqrt{2\pi V_s t}} \exp \left[\frac{-\left(\frac{x-x_0}{xx_0} + \log \left\{ \frac{x(1-x_0)}{x_0(1-x)} \right\} - \bar{s}t \right)^2}{2V_s t} \right]. \tag{19}$$

Since the processes $X(t)$ in all these cases are Markov, the transition probability densities (17) through (19), together with the initial distributions, provide complete descriptions of the evolution of gene frequencies with random selection.

It is interesting to compare the transition density (17) with that obtained by Kimura [2] in the case of a zero mean selection coefficient. This will indicate the quantitative difference which arises in the transition densities when the stochastic equation, in the case of no dominance, is integrated according to Ito's or Stratonovich's definition of stochastic integral. Figure 1 shows the computed solutions for the times $t=10$ and $t=100$ when the

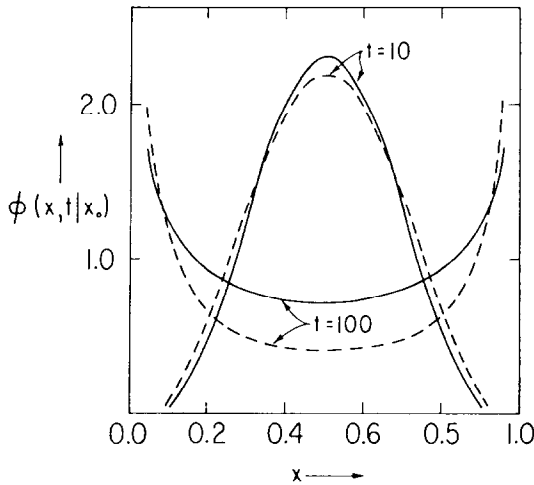


FIG. 1. Transition density functions at various times plotted against gene frequency for random selection in the haploid case. The dashed curves are the results for Ito's stochastic integral and are the same as those computed by Kimura [2]. The solid curves correspond to the use of Stratonovich's integral. In both cases the mean selection coefficient is zero, the initial frequency is 0.5, and $V_s = 0.0483$.

variance parameter V_s has the value 0.0483/unit time, and the initial frequency of A_1 is 0.5. Our solution (Stratonovich calculus) is obtained by adjusting (17) to the haploid case [by omission of the factor $\frac{1}{2}$ in Eq. (2)] in order to make a meaningful comparison.

It can be seen that Kimura's solution spreads more rapidly than ours. This difference can be traced to the drift $\frac{1}{2}V_s(1-x)x(1-2x)$ in our Fokker-Planck equation, which always opposes the change in gene frequency. That is, the drift is positive for small $X(t)$, directed to larger values, whereas for large $X(t)$ the drift is negative.

These quantitative differences are made more precise by noting that the ratio of Kimura's solution to ours, for $x_0 = \frac{1}{2}$ and $\bar{s} = 0$, is

$$R(x, t) = \frac{1}{2}[x(1-x)]^{-1/2} \exp(-\frac{1}{8}V_s t). \quad (20)$$

Hence there will always be two gene frequencies at which the densities are equal, and these frequencies will move towards the ends of the interval (0, 1) as time increases. Also, Kimura's solution is always less than ours between the points of equality, and the difference in this region attains its maximum at $x = \frac{1}{2}$. Asymptotically, as t becomes infinite, Kimura's solution is less than ours throughout the entire interval (0, 1), but qualitatively the asymptotic behaviors of the two solutions are similar in that there is an accumulation of one-half of the entire probability mass at each terminal class.

We now turn our attention to the temporal evolution of the gene frequency transition density in the diploid case when A_1 is dominant and favored. In Fig. 2 we show the results for a mean selection coefficient of zero and an initial frequency of $\frac{1}{2}$. Noteworthy, apart from the clear lack of symmetry at all times (though one-half the probability mass is still found on each side of $x = \frac{1}{2}$) is the vestige of a secondary peak on the curve for $t = 100$ at a very small gene frequency (about 0.04). At $t = 500$ this peak is very much larger than the one at higher gene frequencies. Hence the recessive allele has a small but finite chance of attaining a state of near fixation much earlier than the dominant allele, A_1 .

As mentioned earlier, the mean value and the variance parameter of the selection coefficient may be not constant but explicit functions of time. The allele frequency is still a Markov process, and one would expect this time dependence to be important when environmental conditions, insofar as they affect the fitness of genotypes, undergo periodic fluctuations or long term drifts (e.g., due to major climatological changes). We only wish to point out here that if the mean of $s(t)$ is $\bar{s}(t)$ and its variance parameter is $V_s(t)$, then the transition probability densities for the gene frequency can be found for

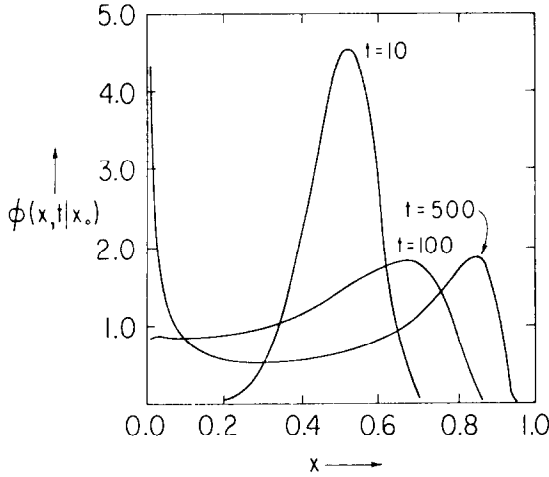


FIG. 2. Transition density functions versus gene frequency for the case of complete dominance, dominant allele favored. The mean selection coefficient is zero, the initial frequency is 0.5 and $V_s = 0.05$. The curves were computed using Eq. (18). Note the vestige of a secondary peak on the curve for $t = 100$ at about $x = 0.04$.

all three cases of dominance from Eqs. (17) to (19) by making the substitutions

$$\bar{s}t \rightarrow \int_0^t \bar{s}(t') dt',$$

$$V_s t \rightarrow \int_0^t V_s(t') dt'.$$

4. QUASIFIXATION

Kimura [2] determined that if $x_1(t)$ and $x_2(t)$ were the abscissae of the left and right hand maxima of the gene frequency transition density at time t (for the case of no dominance and $\bar{s} = 0$) then, given an initial frequency x_0 , the following asymptotic relations prevail for large t :

- (a) $\Pr\{0 < X(t) < x_1(t) | X(0) = x_0\} \rightarrow 0.$
- (b) $\Pr\{x_1(t) < X(t) < \mathfrak{E}_1 | X(0) = x_0\} \rightarrow 1 - x_0.$
- (c) $\Pr\{\mathfrak{E}_1 < X(t) < x_2(t) | X(0) = x_0\} \rightarrow x_0.$
- (d) $\Pr\{x_2(t) < X(t) < 1 | X(0) = x_0\} \rightarrow 0.$

Here \mathfrak{E}_1 and \mathfrak{E}_2 are arbitrary frequencies less than and greater than $x_2(t)$

and $x_1(t)$ respectively and lying within $(0, 1)$. The accumulation of probability mass near $x=0$ was termed quasiloss, and that near $x=1$ termed quasifixation. Recall that these results are those obtained when one uses Ito's calculus.

One finds that it is not necessary, however, to study the asymptotic probabilities in relation to the maxima in order to exhibit the phenomenon of quasifixation. One may ask instead for

$$Q_0(x_0, \bar{s}, V_s) = \lim_{t \rightarrow \infty} \Pr\{0 < X(t) < \bar{\epsilon} | X(0) = x_0\}, \quad (21)$$

$$Q_1(x_0, \bar{s}, V_s) = \lim_{t \rightarrow \infty} \Pr\{\bar{\epsilon} < X(t) < 1 | X(0) = x_0\}, \quad (22)$$

where $\bar{\epsilon} \in (0, 1)$. One finds no dependence on $\bar{\epsilon}$ for these quantities, as none of the processes considered possess stationary (equilibrium) distributions in the limit of large t . This is clear because they are transformed Wiener processes on the whole real line, for which $+\infty$ and $-\infty$ are natural boundaries (see Feller [22] for boundary classification).

For the case of no dominance and $\bar{s}=0$ we obtain $Q_0(x_0)=1-x_0$ and $Q_1(x_0)=x_0$ when we use Kimura's transition density. If, on the other hand, we use the physical approach to stochastic integrals, we obtain the value $\frac{1}{2}$ for both of these asymptotic probabilities. For this approach the same limiting probabilities are in fact obtained, for $\bar{s}=0$, for both cases of dominance. Hence we see a striking difference in that with the physical approach, the probabilities of quasifixation and quasiloss are independent of the initial gene frequency.

Calculating Q_0 and Q_1 from (17), (18) or (19) for nonzero values of \bar{s} yields the following results:

$$Q_0 = \begin{cases} 1 & \text{if } \bar{s} < 0, \\ 0 & \text{if } \bar{s} > 0, \end{cases} \quad (23)$$

$$Q_1 = 1 - Q_0. \quad (24)$$

These relations are obtained simply by noting that, in the general case, for the transition density of Eq. (16),

$$\begin{aligned} \Pr\{a < X(t) < b | X(0) = x_0\} &= \int_a^b \phi(x, t | x_0) dx \\ &= \mathcal{N}\left(\frac{y(b) - y(x_0) - \bar{s}t}{\sqrt{V_s t}}\right) - \mathcal{N}\left(\frac{y(a) - y(x_0) - \bar{s}t}{\sqrt{V_s t}}\right), \end{aligned} \quad (25)$$

where $y(x)$ is the function defined by the transformation (13), and $\mathcal{N}(\cdot)$ is the normal distribution function. A similar set of limiting results was obtained in the haploid case by Gillespie [4] and for changes of fitness at regular intervals by Cook and Hartl [10]. We may summarize the results for diploid populations thus: *Under conditions of randomly varying selection, quasifixation for an allele which has a mean selective advantage will occur with probability one, no matter what its initial frequency in the population nor how large the fluctuations in selection might be.* This statement is true for all the cases of dominance considered, and is valid provided $x_0 \neq 0$, V_s is finite and the selection coefficient is uncorrelated from generation to generation.

The above results are those one obtains when the physical approach to stochastic integrals is used. In order to compare the calculi further, we studied the long term behavior of the transition density satisfying the Fokker-Planck equation which one obtains (for $\bar{s} \neq 0$ and no dominance) with the Ito integral (Wright-Kimura approach). Analytic asymptotic solutions were obtained as $x \rightarrow 0$ and $x \rightarrow 1$, for $x_0 \approx 0$ and $x_0 \approx 1$ respectively. The derivations of the expressions are too lengthy to be given here, so we merely quote the results.

$$Q_1(x_0, \bar{s}, V_s) = \begin{cases} x_0 & \text{if } \bar{s} > -\frac{1}{2}V_s, \\ \frac{1}{2}x_0 & \text{if } \bar{s} = -\frac{1}{2}V_s, \\ 0 & \text{if } \bar{s} < -\frac{1}{2}V_s, \end{cases} \quad (26)$$

$$Q_0(x_0, \bar{s}, V_s) = \begin{cases} 1 - x_0 & \text{if } \bar{s} < \frac{1}{2}V_s, \\ \frac{1}{2} - \frac{1}{2}x_0 & \text{if } \bar{s} = \frac{1}{2}V_s, \\ 0 & \text{if } \bar{s} > \frac{1}{2}V_s. \end{cases} \quad (27)$$

Hence, for example, using Ito's calculus, we find that an allele has a nonzero probability of becoming quasifixed even though it is on average at a selective disadvantage, provided the variance of the selection coefficient is greater than twice the magnitude of its mean ($V_s > 2|\bar{s}|$). In contrast to the physical approach, one then finds that fluctuations in selection may carry an allele to fixation which in the absence of "noise" would have been destined to complete loss in the population. This result seems to favor the use of the Stratonovich integral when integrating the stochastic differential equations for the allele frequencies.

5. ANALYSIS OF THE DATA FOR MEDIONIGRA

Fisher and Ford [23] reported the annual variations in the frequency of the *medionigra* gene in the moth *Panaxia dominula*. This gene has the effect of reducing or eliminating one of the yellow spots on the forewings. The

population studied was a wild one near Oxford, England, the numbers being large enough and the colony being sufficiently isolated to enable the conditions of our model of random selection to be met. Fisher and Ford concluded, in fact, that the fluctuations in frequency, for the years 1939–1946, could not be attributed to the random sampling of gametes and must have arisen from fluctuations in selection. Our aim is to examine that claim in a quantitative fashion.

Wright [1] estimated the variance parameter of the selection coefficient as 0.0483 per generation and stated that it was likely that the heterozygote, $+/m$, where m denotes *medionigra*, had a slight mean selective advantage over the homozygotes m/m and $+/+$. Wright went on to examine stationary (equilibrium) distributions under a variety of conditions, but no comparison of the data was made, or has subsequently been made, with the results from the time dependent theory of random selection. It was therefore decided to determine the transition densities for diploid random selection without dominance, using Wright's estimate of V_s and the values 0 and -0.129 for the mean selection coefficient. The latter value of \bar{s} was estimated from the data in the following way.

Let x be the frequency of the m -allele, and let $X(t_i)$, $i = 1, \dots, 8$, be the observed frequencies of *medionigra* for the years 1939–1946. Further, let $\Delta X_i = X(t_{i+1}) - X(t_i)$, $i = 1, \dots, 7$, be the increments in gene frequency. The term $K_1(x)$ in the Fokker-Planck equation represents the mean rate of change of gene frequency per unit time.

$$K_1(x) = \lim_{\Delta t \downarrow 0} \frac{E[X(t + \Delta t) - X(t) | X(t) = x]}{\Delta t} \quad (28)$$

If the selective advantage of m/m is $2s$ relative to $+/+$, we have

$$K_1(x) = \frac{1}{2}V_s x(1-x)(1-2x) + \bar{s}x(1-x). \quad (29)$$

Using this formula we can compute, from the given value of V_s , the expected change in frequency for each time interval, with $\Delta t = 1$, in terms of \bar{s} . We then equate the sum of the expected changes to the observed net change in the frequency of m for the years 1939–1946. Our estimate is then

$$\bar{s}_{\text{est}} = \frac{\sum_1^7 \Delta X_i - \frac{1}{2}V_s \sum_1^7 X(t_i)[1 - X(t_i)][1 - 2X(t_i)]}{\sum_1^7 X(t_i)[1 - X(t_i)]}. \quad (30)$$

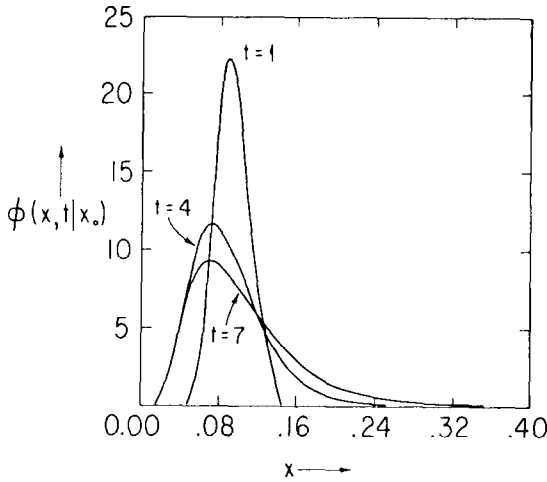


FIG. 3. Showing the spread of the transition density for random selection in the diploid case without dominance using parameters for the gene *medionigra* of *Panaxia dominula*. Here $\bar{s} = 0$ and $V_s = 0.0483$. The initial frequency (x_0 at $t = 0$) is 0.092 as given by Fisher and Ford [23] for the year 1939.

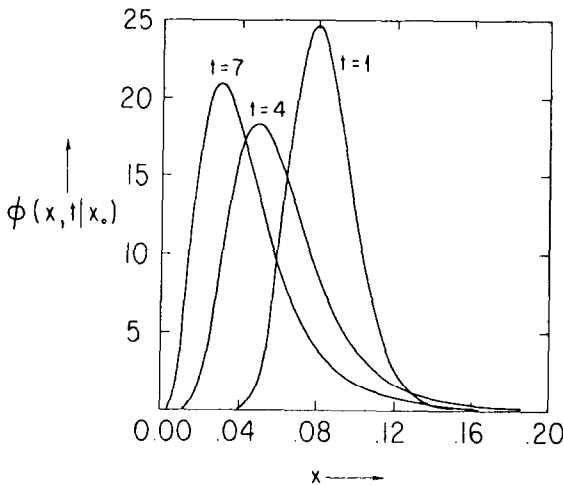


FIG. 4. Computed densities at various times for *medionigra* using $\bar{s} = -0.129$, the remaining parameters being as for Fig. 3. Note that this and the last figure have different gene frequency scales.

In Figs. 3 and 4 are shown the computed transition densities for the above values of \bar{s} , the initial frequency being the 1939 value of 0.092. It can be seen that when $\bar{s}=0$ the density spreads appreciably within a few generations to frequencies of 0.25 and greater. If our estimated value of \bar{s} is used, so that m has a mean selective disadvantage, then the density tends to accumulate at frequencies less than 0.1.

Using the above values for the parameters, the expectation and variance of *medionigra* were computed by numerical integration, using our analytic solutions for the transition densities. These calculations were done for the years 1940–1946, and the results, together with the observed gene frequencies are collected in Table 1. The last column of this table contains the number of standard deviations (σ_i for time t_i) contained in the absolute value of the difference between the observed frequencies and the calculated mean frequencies. The reason for computing this quantity is to see how likely it is that the data could be a sample from the random process modeling the gene frequency. The underlying premise is that one expects the data to lie within one or two standard deviations of the mean if the model is a reasonable approximation to the process occurring in nature. From the calculated and experimental quantities, we make the following observations:

(i) When $\bar{s}=0$, the calculated mean frequency of *medionigra* increases at approximately 0.0015 per annum. One can compare this with the rate of decrease of about 0.006 per year when $\bar{s} - \bar{s}_{est} = -0.129$. These rates of

TABLE 1
Observed Frequencies $X(i)$ of *medionigra* for the Years 1939–1946, ($i=1, \dots, 8$)
from Fisher and Ford [23], Together with Means \bar{X}_i and Standard
Deviations σ_i Calculated for Random Selection Without Dominance for Two
Values of the Mean Selection Coefficient \bar{s}

i	Observed Frequency $X(i)$	Standard					
		Calculated Mean, \bar{X}_i		Deviation σ_i		$ X(i) - \bar{X}_i /\sigma_i$	
		$\bar{s}=0$	$\bar{s}=-0.129$	$\bar{s}=0$	$\bar{s}=-0.129$	$\bar{s}=0$	$\bar{s}=-0.129$
1	.092	—	—	—	—	—	—
2	.111	.094	.083	.0187	.0169	0.909	1.657
3	.068	.095	.075	.0270	.0219	1.000	0.320
4	.054	.097	.068	.0336	.0245	1.280	0.571
5	.056	.099	.062	.0395	.0258	1.089	0.233
6	.045	.100	.056	.0448	.0263	1.228	0.418
7	.065	.102	.050	.0497	.0264	0.744	0.568
8	.043	.103	.045	.0541	.0260	1.109	0.077

change are, of course, only relevant to the limited time period under consideration.

(ii) When $\bar{s}=0$, all the observed frequencies fall within about one standard deviation of the calculated means. For all years except 1940 the observed frequencies are less than the calculated means.

(iii) When $\bar{s} = -0.129$, all the observed values except the 1940 reading fall within about one-half standard deviation of the calculated means. Two of the observations lie above and the rest below the expected frequencies. Note that the 1940 reading was based on the smallest sample.

It is concluded from (ii) and (iii) that random fluctuations in selection could quite easily have given rise to variations in the frequency of *medionigra* of the order of magnitude of those observed. This conclusion is based on the computed results for either $\bar{s}=0$ or $\bar{s} = -0.129$. The latter value is admittedly of large magnitude (though compare it with that estimated for the Bar gene of *Drosophila melanogaster* by Wright and Kerr [24]), but it is apparent that smaller in magnitude yet negative values of \bar{s} will result in calculated means which lie closer to the observed frequencies than those for the case $\bar{s}=0$. It seems reasonable to postulate that *medionigra* did have an associated mean selective disadvantage for the time period considered.

The author is grateful for helpful discussions with Professor L. M. Ricciardi, comments on the manuscript by Professor M. Slatkin and assistance from Dr. S. Coyne.

REFERENCES

- 1 S. Wright, On the roles of directed and random changes in the genetics of populations, *Evolution* **2**, 279–295 (1948).
- 2 M. Kimura, Process leading to quasifixation of genes in natural populations due to random fluctuations of selection intensities, *Genetics* **39**, 280–295 (1954).
- 3 M. Kimura, Stochastic processes and distribution of gene frequencies under natural selection, *Cold Spring Harbor Symp. Quant. Biol.* **20**, 33–53 (1955).
- 4 J. H. Gillespie, The effects of stochastic environments on allele frequencies in natural populations, *Theor. Popul. Biol.* **3**, 241–248 (1972).
- 5 J. H. Gillespie, Natural selection with varying selection coefficients—a haploid model, *Genet. Res.* **21**, 115–120 (1973).
- 6 L. Jensen and E. Pollack, Random selective advantages of a gene in a finite population, *J. Appl. Prob.* **6**, 19–37 (1969).
- 7 T. Ohta, Fixation probability of a mutant influenced by random fluctuation of selection intensity, *Genet. Res.* **19**, 33–38 (1972).
- 8 L. Jensen, Random selective advantages of genes and their probabilities of fixation, *Genet. Res.* **21**, 215–219 (1973).
- 9 M. Kimura, On the probability of fixation of mutant genes in a population, *Genetics* **47**, 713–719 (1962).

- 10 R. D. Cook and D. L. Hartl, Uncorrelated random environments and their effects on gene frequency, *Evolution* **28**, 265–274 (1974).
- 11 D. L. Hartl and R. D. Cook, Autocorrelated random environments and their effects on gene frequency, *Evolution* **28**, 275–280 (1974).
- 12 J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory*, Harper and Row, New York, 1970.
- 13 M. Kimura, Diffusion models in population genetics, *J. Appl. Prob.* **1**, 177–232 (1964).
- 14 A. H. Jaswinski, *Stochastic Processes and Filtering Theory*, Academic, New York, 1970.
- 15 K. Ito, On stochastic differential equations, *Mem. Am. Math. Soc.*, No. 4, 1951.
- 16 R. L. Stratonovich, A new representation for stochastic integrals and equations, *SIAM J. Control* **4**, 362–371 (1966).
- 17 R. M. May, Stability in model ecosystems, *Proc. Ecol. Soc. Aust.* **6**, 18–56 (1971).
- 18 A. H. Gray and T. K. Caughey, A controversy in problems involving random parameter excitation, *J. Math. and Phys.* **44**, 288–296 (1965).
- 19 R. E. Mortensen, Mathematical problems of modeling stochastic nonlinear dynamic systems, *J. Stat. Phys.* **1**, 271–296 (1969).
- 20 E. Wong and M. Zakai, On the convergence of ordinary integrals to stochastic integrals, *Ann. Math. Stat.* **36**, 1560–1564 (1965).
- 21 H. C. Tuckwell, A study of some diffusion models of population growth, *Theor. Pop. Biol.* **5**, 345–357 (1974).
- 22 W. Feller, The parabolic differential equations and the associated semigroups of transformations, *Ann. Math.* **55**, 468–519 (1952).
- 23 R. A. Fisher and E. B. Ford, The spread of a gene in natural conditions in a colony of the moth *Panaxia dominula* L., *Heredity* **1**, 143–174 (1947).
- 24 S. Wright and W. E. Kerr, Experimental studies of the distribution of gene frequencies in very small populations of *Drosophila melanogaster* II. Bar., *Evolution* **8**, 225–240 (1954).