Proc. Natl. Acad. Sci. USA
Vol. 78, No. 4, pp. 2452–2454, April 1981
Genetics

Evolution of the mutation rate at a heterotic locus

(population genetics/overdominance/modifier theory/genetic drift)

JOHN H. GILLESPIE

Division of Environmental Studies and Department of Genetics, University of California, Davis, California 95616

Communicated by Francisco J. Ayala, December 29, 1980

ABSTRACT **A diffusion model of the modification of mutation rates at a heterotic locus in a finite population is examined. An asymptotic analysis assuming strong selection and weak linkage shows that selection can operate on mutation rates in this setting. There exists a favored mutation rate which is a function only of the equilibrium allele frequency of the heterotic locus and the population size. It is independent of the strength of selection at the heterotic locus. Computer simulations are also provided to show that this form of natural selection can occur.**

The evolution of mutation rates has been actively investigated for many years. Early theoretical studies by Kimura (1, 2) and Levins (3) based their arguments on optimization principles; later studies by Leigh (4, 5), Karlin and McGregor (6), and Gillespie (7) utilized modifier theory. All of these studies tried to capture the idea that mutation has a positive and a negative aspect. The positive aspect is to provide genetic variation for evolutionary responses to a changing environment. The negative aspect is the genetic load caused by the production of deleterious mutations. I have shown (7) that selection can move mutation rates toward a maximal, a minimal, or an intermediate value, depending on the parameters of the model, when the primary locus is subjected to selection in a fluctuating environment.

In this paper, a novel approach to the study of selection of mutation rates is presented which does not involve a changing environment. I postulate that the primary locus is undergoing heterotic selection in a finite population and that the modifier locus affects the mutation rate at the primary locus. In this situation, if there is no mutation at the primary locus it will eventually become monomorphic. This is clearly a suboptimal state for the population. One might expect an allele at the modifier locus which causes mutation at the primary locus to be favored. I will show that this is indeed the case. Furthermore, I will show that selection will always move the mutation rate toward an intermediate value which depends only on the equilibrium frequency of the heterotic locus.

THE DIFFUSION MODEL

Consider a diploid species with two alleles segregating at the primary locus. Let the fitnesses of the genotypes A_1A_1, A_1A_2, and A_2A_2 be $1 - s_1$, 1, and $1 - s_2$, respectively. Define $m = s_2/(s_1 + s_2)$ as the deterministic equilibrium of the heterotic locus and define $s = 2N(s_1 + s_2)$ in which N is the effective population size. Let the mutation process between A_1 and A_2 be symmetric at the rates u_1, $(u_1 + u_2)/2$, and u_2 when in the presence of the M_1M_1, M_1M_2, and M_2M_2 genotypes at the modifier locus. Define $v_i = 2Nu_i$. Let R be the probability of a recombinational event between the A and M loci, and define $r = 2NR$. With these assumptions and definitions, if we fix r, s, m, v_1, and v_2 and allow N to increase, we obtain the following diffusion model for the frequency of the A_1 allele (p_1), the frequency of the M_1 allele (p_2), and the linkage disequilibrium (D):

$$\begin{aligned}
E(dp_1) &= [p_1q_1s(m-p_1) + 2\bar{v}(1/2-p_1) + D(v_2-v_1)]dt \\
E(dp_2) &= [Ds(m-p_1)]dt \\
E(dD) &= \{D[s(q_1-p_1)(m-p_1) - r - 1 \\
&\qquad - (v_1+v_2)] + p_2q_2(1/2-p_1)(v_1-v_2)\}dt \\
E(dp_i^2) &= p_iq_i dt,\ i = 1,2 \\
E(dD^2) &= [p_1p_2q_1q_2 + D(1-2p_1)(1-2p_2) - D^2]dt \\
E(dp_i dD) &= D(1-2p_i)dt \\
E(dp_1dp_2) &= Ddt.
\end{aligned} \qquad [1]$$

In this system, $q_i = 1-p_i$ and $\bar{v} = p_2v_1 + q_2v_2$.

We wish to obtain the fixation probability of the M_1 allele, given an arbitrary initial condition for the diffusion. This probability will indicate the direction of selection for a given set of parameters. Calculation of the fixation probability requires the solution of the backward equation corresponding to [1]. I have been unable to obtain the solution so an asymptotic approximation will be presented for large r and s.

As r and s increase, $p_1 \to m$ and $D \to 0$ in probability. It is reasonable, therefore, to approximate their dynamics with a two-dimensional Ornstein–Uhlenbeck process obtained by linearizing (p_1,D) around (m,o). Because the autocovariance of both D and p_1 will approach zero for a fixed time lag as s and r increase while that of p_2 remains essentially unchanged, we can view p_2 as a parameter in the Ornstein–Uhlenbeck approximation to (p_1,D). The approximation may be written

$$\begin{aligned}
E(dp_1) &= \{sm^2(1-m) + \bar{v} \\
&\qquad - [sm(1-m) + 2\bar{v}]p_1 + (v_2-v_1)D\}dt \\
E(dD) &= \{p_2q_2(v_1-v_2)/2 \\
&\qquad - (r+1+v_1+v_2)D - p_2q_2(v_1-v_2)p_1\}dt \\
E(dp_1^2) &= m(1-m)dt \\
E(dD^2) &= p_2q_2m(1-m)dt \\
E(dp_1dD) &= 0.
\end{aligned} \qquad [2]$$

We could, of course, obtain the complete solution to this linearized process. However, we only require the expectation of $sD(m-p_1)$ with respect to the stationary distribution of the linearized system. This is because the dynamics of D and p_1 are

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.

enough faster than those of p_2 that p_2 experiences D and p_1 only through their mean values. A straightforward analysis of system 2 yields

$$E[sD(m-p_1)] \simeq \frac{p_2q_2}{r}(v_1-v_2)\left[\frac{1}{2}-\frac{2\bar{v}(1/2-m)^2}{m(1-m)}\right]dt. \quad [3]$$

The diffusion for p_2 thus becomes, as r and s get larger,

$$E(dp_2) \sim \frac{p_2q_2}{r}(v_1-v_2)\left[\frac{1}{2}-\frac{2\bar{v}(1/2-m)^2}{m(1-m)}\right]dt \quad [4]$$

$$E(dp_2^2) = p_2q_2dt.$$

This diffusion has a small drift coefficient (of order $1/r$), so we can approximate the fixation probability given an initial frequency for the M_1 allele of p_2 as

$$\Pi(p_2) = p_2 + \frac{p_2q_2}{r}(v_1-v_2)[1/2 - \alpha(v_1+v_2+\bar{v})] \quad [5]$$

where

$$\alpha = 2(1/2-m)^2/[m(1-m)]$$

$$\bar{v} = p_2v_1 + q_2v_2.$$

This is the main result.

DISCUSSION

The asymptotic analysis yields an expression for the fixation probability of the M_1 allele. If selection could not operate on the modifier locus, we would expect $\Pi(p_2) = p_2$ as for a neutral allele. This is clearly not the case. The main property of this model is that there exists a mutation rate toward which selection always moves. This rate is

$$v^* = m(1-m)/4(1/2-m)^2, \quad [6]$$

or, in terms of the actual, rather than the scaled, mutation rates,

$$u^* = \frac{1}{2N}\cdot\frac{m(1-m)}{4(1/2-m)^2}. \quad [7]$$

This is easily verified by setting $v_2 = v^*$ in Eq. 5 and noticing that $\Pi(p_2)$ is always less than p_2. Thus, any allele with a mutation rate different from v^* is selected against. This favored mutation rate is highest when $m \simeq 1/2$. When $m = 1/2$ the favored rate is the highest achievable rate. This follows from the fact that the mutation process is symmetric so that increasing the rate pushes the allele frequency of the primary locus toward 1/2. As m approaches 0 or 1, the favored mutation rate decreases. In this instance we are approaching the case of mutation-selection balance where we know (6) that selection will always decrease the mutation rate.

Diffusion process 4 and the fixation probability 5 are both independent of the strength of selection, s, on the primary locus. This appears to be because of two conflicting influences of s on the process. On the one hand, s is going to affect the total selection in the system, so increases in s should lead to increased selection at both the primary and modifier loci. However, as s increases, the allele frequency p_1 is held more closely to m so that the term (p_1-m) in $E(dp_2)$ becomes smaller. It just so happens that this term is exactly of order $1/s$, so s times this term becomes independent of s.

Notice that the favored (unscaled) mutation rate u^* is proportional to $1/2N$. As population size increases, the favored mutation rate becomes quite small. Because selection on mutation rates in the case of mutation-selection balance always tries to minimize mutation rates (6) and because in the present model we have $u^* = 0$ $(1/2N)$ (except for the special case $m = 1/2$) we might expect mutation rates in nature to be of order $1/2N$ or smaller if most of the selection were in response to heterotic and deleterious alleles. It would be impossible to check this prediction accurately without a fairly exact estimate of effective population sizes and the relative amounts of heterotic versus deleterious allele selection. Nonetheless, it does seem implausible that mutation rates would be adjusted to values that are independent of population size as assumed by various versions of the neutral allele theory.

A simple calculation shows that the favored mutation rate v^* is also the mutation rate that maximizes the mean value of the population mean fitness to the same order of approximation as the other results. It should be cautioned that several of these qualitative results may be due to the nature of the approximations or the symmetry of the mutation process.

Finally, it should be noted that this form of selection is very weak. However, if the modifier locus affects not one but a large number of primary loci, we might expect selection on the modifier locus to be cumulative and thus to be much stronger than in the single primary locus example.

COMPUTER SIMULATIONS

Some effort has been made to document this form of selection as well as to check the approximations presented in this paper by using computer simulations. Simulating this model turns out to be quite time-consuming for a number of reasons. The strength of selection is quite small, so a large number of replications are needed to see any effect at all. Because the aim is, in part, to check the approximate analysis, simulations with large values of r and s but with large enough population sizes so that the parameters s_i and R are small enough to expect reasonable agreement with a diffusion model are required. The time required to complete these simulations increases not as N but as N^2 because it takes on the order of N generations to reach fixation, starting from a moderate allele frequency. We cannot expect to get better agreement with the approximations by increasing r and s without also increasing N. Increasing N beyond about 250 makes the simulations intolerably slow. Nonetheless, some results are presented which illustrate that the qualitative features of the approximate analysis are certainly borne out although the quantitative agreement is not always as good as we would hope, for the reasons cited above.

The simulations were performed on a Z-80, S-100 based microcomputer with a 4-MHz clock. The parts of the program concerned with selection and drift were written in assembly language to achieve maximal speed. Furthermore, a hardware pseudorandom number generator was constructed based on a 32-stage maximal-length sequence-generator using TTL shift registers to provide pseudorandom numbers as quickly as the

Table 1. Simulation results for fixation probabilities

m	r	s	N	Observed	Expected
0.5	10	20	128	0.537 ± 0.015	0.550
0.5	20	10	128	0.518 ± 0.015	0.525
0.5	20	20	128	0.525 ± 0.015	0.525
0.5	20	30	128	0.538 ± 0.015	0.525
0.3	20	50	128	0.515 ± 0.015	0.501
0.3	20	50	256	0.507 ± 0.015	0.501
0.1	20	100	128	0.480 ± 0.015	0.341
0.1	20	100	256	0.459 ± 0.015	0.341

For each case, $v_1 = 4.0$ and $v_2 = 0.01$ with 4000 replicates. The initial frequency of the modifier allele was $p_2 = 0.5$.

microprocessor could utilize them. Because the individuals were represented as locations in memory, no floating point operations were required and the simulations proceeded at a speed as fast as all but the largest computers. Nonetheless, a single simulation of a population of size 256 with 4000 replicates took about 80 hr to complete.

In all of the simulations, the initial conditions were $p_1 = p_2 = 0.5$ and $D = 0$. The values for the parameters which are not listed in Table 1 are $v_1 = 4.0$ and $v_2 = 0.01$. Each simulation involved 4000 replicates. The 95% confidence limits of the estimated fixation probabilities are provided. Under these conditions if $\Pi(0.5) > 0.5$, selection tends to increase the mutation rate whereas if $\Pi(0.5) < 0.5$, selection tends to decrease the rate.

As Table 1 illustrates, the selection is quite weak but the agreement with the approximate fixation probabilities is acceptable except for the very asymmetric case $m = 0.1$ where the agreement is not very good. The problem for this case is probably that we require $sm(1-m)$ to be large to get agreement with the approximations yet, when we make $sm(1-m)$ large, the absolute values of the selection coefficients are too large to provide good agreement with the diffusion equation. Notice that the agreement with the predictions improves in both $m = 0.1$ and $=0.3$ cases when the population size is doubled. The simulations illustrate that selection shifts from increasing the mutation rate to decreasing the rate as m shifts from 0.5 to 0.1.

I am very grateful to Gordon Ellis for his help in programming the Z-80 and for assistance in the construction of the hardware random number generator.

1. Kimura, M. (1960) *J. Genet.* **57**, 21–34.
2. Kimura, M. (1967) *Genet. Res. Camb.* **9**, 23–34.
3. Levins, R. (1967) *Genetics* **56**, 163–178.
4. Leigh, E. (1970) *Am. Nat.* **104**, 301–305.
5. Leigh, E. (1973) *Genet. Suppl.* **73**, 1–18.
6. Karlin, S. & McGregor, J. (1974) *Theor. Pop. Biol.* **5**, 59–105.
7. Gillespie, J. H. (1981) *Evolution*, in press.