Possibility of extensive neutral evolution under stabilizing selection with special reference to nonrandom usage of synonymous codons

(molecular evolution/quantitative character/natural selection/neutral mutation–random drift hypothesis MODE OF EVOLUTION)

MOTOO KIMURA

National Institute of Genetics, Mishima, 411 Japan

Contributed by Motoo Kimura, June 3, 1981

ABSTRACT The rate of evolution in terms of the number of mutant substitutions in a finite population is investigated assuming a quantitative character subject to stabilizing selection, which is known to be the most prevalent type of natural selection. It is shown that, if a large number of segregating loci (or sites) are involved, the average selection coefficient per mutant under stabilizing selection may be exceedingly small. These mutants are very slightly deleterious but nearly neutral, so that mutant substitutions are mainly controlled by random drift, although the rate of evolution may be lower as compared with the situation in which all the mutations are strictly neutral. This is treated quantitatively by using the diffusion equation method in population genetics. A model of random drift under stabilizing selection is then applied to the problem of “nonrandom” or unequal usage of synonymous codons, and it is shown that such nonrandomness can readily be understood within the framework of the neutral mutation–random drift hypothesis (the neutral theory, for short) of molecular evolution.

It is generally accepted that stabilizing selection is the most prevalent type of natural selection at the phenotypic level (1–4). It eliminates phenotypically extreme individuals and preserves those that are near the population mean (5). It is also called centripetal selection (6) or normalizing selection (7), and many examples have been reported. Probably the best example in human populations is the relationship between the birth weights of babies and their neonatal mortality, as studied by Karn and Penrose (8). These authors found that babies whose weight is very near the mean have the lowest mortality. This optimum weight is slightly heavier than the mean, and mortality increases progressively as the birth weight deviates from this optimum (see also ref. 9). Unlike the type of natural selection that Darwin (10) had in mind when he tried to explain evolution, stabilizing selection acts to keep the status quo rather than to cause a directional change. From this, it might appear that stabilizing selection is antithetical to evolutionary change.

In this note, I intend to show that, under stabilizing phenotypic selection, extensive “neutral evolution” can occur. By neutral evolution, I mean accumulation of mutant genes in the species through random genetic drift (due to finite population size) under mutational pressure. Thus, beneath an unchanged morphology, a great deal of cryptic genetic change may be occurring in natural populations of all organisms, transforming even genes of “living fossils” (11). This will substantiate my neutral mutation–random drift hypothesis (the neutral theory, for short) of molecular evolution (ref. 12; for review, see ref. 13).

I shall also show that this gives a clue to understanding “nonrandom” or unequal usage of synonymous codons (14–16) based on the neutral theory.

Selection intensity at an individual locus when overall phenotypic selection is given

Let us consider a quantitative character, such as height, weight, concentration of some substance, or a more abstract quantity that represents Darwinian fitness in an important way. We assume that the character is determined by a large number of loci (or sites), each with a very small effect in addition to being subjected to environmental effects. We also assume that genes are additive with respect to the character. We follow the method used by Bulmer (17, 18) and Kimura and Crow (19). Let $X$ be the measured phenotypic character with the mean $M$ and the variance $\sigma^2$. We denote by $X_\text{opt}$ the optimum phenotypic value and, unless otherwise stated, we shall take this point as the origin. Let $F(X)$ and $W(X)$ be the relative frequency and the fitness of individuals with character value $X$. Two examples of $W(X)$ are shown in Fig. 1.

Consider a particular locus at which a pair of alleles $A_1$ and $A_2$ are segregating with respective frequencies $1 - p$ and $p$. We assume a random-mating diploid population and let $X_q$ be the average phenotypic value of $A_1A_1$ individuals, where $i = 1$ or 2 and $j = 1$ or 2.

It is often convenient to measure various quantities relating to the character value in units of the standard deviation ($\sigma$). For this purpose, lowercase letters will be used such as $x = (X - X_\text{opt})/\sigma$, and the corresponding frequency and fitness functions will be denoted by $f(x)$ and $w(x)$. We also let $m = (M - X_\text{opt})/\sigma$ and $a_q = (X_q - M)/\sigma$. Note that $a_q$ is the deviation of $A_1A_1$ from the population mean in $\sigma$ units.

We assume that the background distribution of the character is the same among different genotypes at this locus and that this is given by $f(x)$ with good approximation, because individual gene effects are assumed to be extremely small. Let $w_q$ be the relative fitness of $A_1A_1$, then

$$w_q = \int_{-\infty}^{\infty} w(x)f(x)\,dx,$$  

[1]

as explained in ref. 19. Assuming that $a_q$ is small, we expand $f(x - a_q) \equiv a_qf'(x) + (a_q^2/2)f''(x) - \ldots$,  

[2]

as in ref. 19, but here we retain the second-order term, so that we get, from Eq. 1,

$$w_q = b_0 - a_q b_1 + a_q^2 b_2/2,$$  

[3]

where

$$b_0 = \int_{-\infty}^{\infty} w(x)f(x)\,dx,$$  

$$b_1 = \int_{-\infty}^{\infty} w(x)f'(x)\,dx,$$

and

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.
Here, the prime denotes differentiation.

Let $a$ be the effect of substituting $A_2$ for $A_1$ on the character value $x$. Then, under random mating and assuming an additive gene effect on $x$, we find that $a_{11} = -2ap$, $a_{12} = a(1 - 2p)$, and $a_{22} = 2a(1 - p)$. Then, by using Eq. 3, we can compute the mean fitness $\bar{w} = w_{11}(1 - p)^2 + w_{12}(1 - p)p + w_{22}p^2$ and the average fitness of $A_2$—i.e., $w_s = w_{21}(1 - p) + w_{22}p$—and they turn out to be as follows:

$$\bar{w} = b_0 + a^2p(1 - p)b_2$$ \[4\]

$$w_s = b_0 - a(1 - p)b_1 + a^2(1 - p)b_2/2.$$ \[5\]

The change of the frequency of $A_2$ in one generation is given by $\Delta p = p(w_s - \bar{w})/\bar{w}$ (see p. 180 of ref. 20). Therefore, by substituting Eqs. 4 and 5 in this expression and neglecting terms involving $a^3$ and higher order terms, we obtain

$$\Delta p = \frac{p(1 - p)}{b_0} \left[ - ab_1 + a^2b_2 \left( \frac{1}{2} - p \right) \right].$$ \[6\]

in agreement with Bulmer (17). Then, the selection coefficient, $s$, which represents the selective advantage of $A_2$ over $A_1$ is

$$s = -ab_1/b_0 + a^2b_2(1/2 - p)/b_0.$$ \[7\]

With this selection coefficient $s$, the change of $p$ by selection per generation is $\Delta p = sp(1 - p)$.

In the special, but important, case in which both the frequency and the fitness functions are given by normal distributions,

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp[-(x - m)^2/2],$$ \[8\]

and

$$w(x) = \exp(-sx^2),$$ \[9\]

Eq. 7 reduces to

$$s = -maa + (\lambda^2m^2 - \lambda)(1/2 - p)a^2,$$ \[10\]

where $\lambda = 2k/(1 + 2k)$. An equivalent result was obtained earlier by Bulmer (18).

Note that, if we use the original scale $(X)$ and express the density function of the frequency distribution of the character by $F(X) = (1/\sqrt{2\pi})\exp[-(X - M)^2/(2\sigma^2)]$, the fitness function by $W(X) = \exp(-kX^2)$, and the effect of allele substitution by $A$, then the parameters in Eq. 10 are $m = (M - X_{op})/\sigma$, $\lambda = 2k\sigma^2/(1 + 2k\sigma^2)$, and $a = A/\sigma$. From Eq. 10, it may be seen that, if the deviation of the mean from the optimum is much larger than the effect of the allele substitution $\{m \gg |a|\}$, we have $s/a = -\lambda m$, which agrees with equation 23 of ref. 19. In this case, the situation is similar to truncation selection (see refs. 21 and 22), and natural selection acts very efficiently to change the mean toward the optimum. During this short period of directional selection, extensive shift of gene frequencies is expected to occur at many loci, but this process itself will seldom cause gene substitutions.

If, on the other hand, the mean is at the optimum $(m = 0)$, we have $s = \lambda(1/2 - p)a^2$ from Eq. 10. In this case, $A_2$ is disadvantageous if $p < 1/2$ but advantageous if $p > 1/2$. This selection is frequency dependent, and alleles behave as if negatively overdominant. The change of gene frequency is then given by $\Delta p = A^2a^2p(1 - p)/2p - 1/2a$ in agreement with Robertson (23) and Wright (24). What is pertinent to our evolutionary consideration is that, here, every mutation is deleterious, because $\Delta p < 0$ if $p$ is small. Furthermore, if a large number of loci are segregating, each with a very small effect, $a$ is small so that $a^2$ is extremely small. This applies with particular force if we consider an individual nucleotide site rather than the conventional gene locus, as it has been estimated (25) that the average individual in a large mammalian population is likely to be heterozygous at a million or so nucleotide sites. This substantiates Ohta’s claim (26, 27) that the majority of mutants at the molecular level are nearly neutral but very slightly deleterious. As I shall show below, negatively overdominant alleles are far more susceptible to random genetic drift than unconditionally deleterious alleles having the same magnitude of selection coefficient.

As to the intensity of natural selection involved, we can measure it in terms of load $(L)$—i.e., by the fraction of individuals that are eliminated in each generation by natural selection due to deviation of their phenotypic values from the optimum. For the frequency and fitness functions given as Eqs. 8 and 9, we obtain

$$L = 1 - \sqrt{1 - \lambda \exp(-\lambda m^2/2)}.$$ \[11\]

For $m = 0$, this reduces to $L = 1 - \sqrt{1 - \lambda}$. In the special case in which the fitness function has the same variance as the frequency function—i.e., when $K = 1/(2a^2)$ or $\lambda = k/2$, we get $L \approx 0.293$ or $\approx 30\%$ elimination. In general, $L$ is likely to be small for any single character in mammals. For example, according to Haldane (1), the intensity of selection acting on birth weight of babies through their neonatal mortality is $L = 0.027$. If $L$ is small, we have, with good approximation, $L = \lambda/2$. 

Behavior of mutant alleles in a finite population under stabilizing selection

Probability of Fixation of a Mutant Allele. We denote by $N_e$ the effective size of the population (for the meaning of $N_e$, see ref. 28). Roughly speaking, $N_e$ is equal to the number of breeding individuals in one generation. This number is likely to be much smaller in more cases than the apparent population size, which we denote by $N$. To simplify expressions, we let $\beta_1 = -\lambda a$ and $\beta_2 = \lambda(1 - \lambda m^2a^2)/2$, so that Eq. 10 reduces to

$$s = \beta_1 - \beta_2(1 - 2p).$$ \[12\]

where $p$ is the frequency of $A_2$. Ignoring mutational change for a moment, and denoting the frequency of $A_2$ by $y$, the mean and the variance in the change of $y$ during one generation are

$$M_{\beta y} = [\beta_1 - \beta_2(1 - 2p)]y(1 - y)$$ \[13\]

and

$$V_{\beta y} = y(1 - y)/2N_e.$$ \[14\]
Let \( u(p) \) be the probability that \( A_2 \) becomes eventually fixed in the population (i.e., reaches 100% in frequency), given that its initial frequency is \( p \). Then, \( u(p) \) can be expressed in terms of \( M_{\text{sp}} \) and \( V_{\text{sp}} \) by using a general formula for the probability of fixation obtained by Kimura (ref. 29; see also p. 424 of ref. 20, where \( x \) is used instead of \( y \)). We are particularly interested in the probability of fixation of \( A_2 \) when it is initially singly represented in the population. If we denote this probability by \( u \), then this is given by \( u(p) \) with \( p = 1/(2N) \). We then obtain

\[
u = 1 - \left[ 2N \int_0^1 \exp(-B_1 x + B_2 x(1 - x)) dx \right],
\]

where \( B_1 = 4N_s\beta_1 \) and \( B_2 = 4N_s\beta_2 \).

In the above treatment, we have assumed that \( m \) (the deviation of the mean from the optimum) remains unchanged throughout the process. This assumption appears to be unrealistic because, if \( m \neq 0 \), one would expect \( |m| \) to be reduced with time by the directional component of selection. There is an important possibility, however, that this change is opposed by mutational pressure so that \( m \) remains constant under continued stabilizing selection, although \( |m| \) at equilibrium is likely to be smaller. This occurs when the optimum and the mutational equilibrium point do not coincide. We shall elaborate such a case when we discuss the problem of nonrandom synonymous codon usage.

To show that mutants that have negative overdominance (as induced by stabilizing selection) are far more likely to be fixed by random drift than unconditionally deleterious mutants that have comparable selection coefficients, some examples of the probabilities of fixation (\( u \)) for these two cases are listed in Table 1. In the case of stabilizing selection, we let \( m = 0 \) and denote \( \Delta s' = s' \) (selection coefficient for stabilizing selection), so that \( \Delta p = -s'(1 - p)(1 - 2p) \). For the unconditionally deleterious case, we denote the selection coefficient against \( A_2 \) by \(-s' \) (\( s' > 0 \)), so that the probability of fixation is given by

\[
u = S'/[2Ne^{S' - 1}],
\]

where \( S' = 4N_s\Delta s' \) (see p. 426 of ref. 20). In both cases, \( u \) is tabulated taking the probability of fixation of the completely neutral case as the unit—i.e., it is expressed as a multiple of \( u_0 = 1/(2N) \). It is clear from Table 1 that an enormous difference exists between the two cases in fixation probability and that, under stabilizing selection, extensive neutral evolution is possible even when \( 4N_s \Delta s \) is \( \theta \) or more. For \( B = 4N_s \Delta s > 8 \), it can be shown that \( u/u_0 = \sqrt{B}/\pi \exp(-B/4) \).

<table>
<thead>
<tr>
<th>( S )</th>
<th>Negatively overdominant</th>
<th>Unconditionally deleterious</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1.0</td>
<td>0.84</td>
<td>0.58</td>
</tr>
<tr>
<td>8.0</td>
<td>0.23</td>
<td>2.7 \times 10^{-3}</td>
</tr>
<tr>
<td>16.0</td>
<td>0.042</td>
<td>1.8 \times 10^{-5}</td>
</tr>
<tr>
<td>20.0</td>
<td>0.017</td>
<td>4.1 \times 10^{-8}</td>
</tr>
<tr>
<td>30.0</td>
<td>0.0017</td>
<td>2.8 \times 10^{-12}</td>
</tr>
</tbody>
</table>

\( S \) stands for \( 4N_s\Delta s' \) for the negatively overdominant case and \( 4N_s\Delta s \) for the unconditionally deleterious case.

Gene Frequency Distribution. We now incorporate mutational pressure and investigate the probability distribution of allelic frequencies at statistical equilibrium attained under stabilizing selection in a finite population. We shall denote by \( \phi(p) \) the probability density such that \( \phi(p) dp \) represents the probability that the frequency of \( A_2 \) in the population lies in the range \( p - p + dp \), where \( 0 < p < 1 \). We assume reversible mutations between the two alleles and let \( v_1 \) be the mutation rate from \( A_1 \) to \( A_2 \) and \( v_2 \) be the rate in the reverse direction. Then, the mean and the variance in the change of \( A_2 \) in one generation are, respectively,

\[
M_{\text{sp}} = [\beta_1 - \beta_2(1 - 2p)] p(1 - p) - c_1p + v_1(1 - p)
\]

and

\[
V_{\text{sp}} = p(1 - p)/(2N).
\]

By using Wright’s (30) formula for the steady-state gene-frequency distribution (see p. 434 of ref. 20), we obtain

\[
\phi(p) = C e^{BP}/[V_2(2N) V_1].
\]

Similarly, the probability of \( A_1 \) being temporarily fixed in the population (i.e., \( A_2 \) lost) is

\[
f_1 = C/[V_2(2N) V_1].
\]

Then, the ratio of \( f_2 \) to \( f_1 \) is

\[
f_2/f_1 = e^{B}/(V_1/V_2)(2N) V_1/V_2.
\]

In this paper, we shall be mainly concerned with the situation in which both \( V_1 \) and \( V_2 \) are much smaller than unity and alleles are fixed most of the time. This situation is particularly pertinent when we consider individual nucleotide sites rather than conventional gene loci, because the mutation rate per site must be of the order of \( 10^{-6} \) rather than of \( 10^{-9} \).

In general, for any set of values of \( B_1, B_2, V_1, V_2 \), and \( V_0 \), we can compute the mean frequency \( \bar{p} \) and the mean heterozygosity \( H_{A_2} \) per locus through numerical integration by using \( \bar{p} = E(p) = \int_0^1 \phi(p) dp \) and \( H_{A_2} = E[p(1 - p)] = \int_0^1 2p(1 - p) \phi(p) dp \). If the phenotype is determined by \( n \) equivalent loci in addition to environmental effects, we have \( M = 2naP \) and \( \sigma^2 = nE[p^2(1 - p)]/\bar{p}^2 = nA^2H_{A_2}/\bar{p}^2 \), where \( A \) is the effect of substituting \( A_2 \) for \( A_1 \) on the character (\( A = a\sigma \)) and \( \bar{p} \) is the fraction of phenotypic variance due to gene segregation—i.e., broad sense heritability. Furthermore, if mutation rates are equal in both directions (\( c_1 = c_2 = c \)) and the phenotypic mean coincides with the optimum phenotype (i.e., \( m = 0 \) or \( B_1 = 0 \), \( B_2 = 2N_s\Lambda_s^2 \)), the distribution formula is much simplified, and the values of \( E[p(1 - p)] \) are tabulated by Bulmer (18) for some combinations of values of \( 4N_s \) and \( \Lambda_s^2 \). One interesting property of the frequency distribution of alleles under stabilizing selection is that it is more \( U \) shaped than the strictly neutral case having the same mutation rate. It may sometimes be convenient to take as the standard the situation in which the minus alleles are fixed at all loci. Then, the range of \( X \) lies between 0 and \( 2aP \), and if we assume that no loci are involved and the effect of allele substitution is the same at all the loci. Let

\[
X_{\text{op}} = 2nAQ_{\text{op}},
\]

where \( Q_{\text{op}} \) is the position of the optimum when the total range of \( X \) is rescaled so that it lies in the interval \([0, 1]\). If \( A_1 \) is the minus allele and \( A_2 \) is the plus allele, so that \( A > 0 \), then the optimum is less than the mean if \( Q_{\text{op}} < \bar{p} \) and more than the mean if \( Q_{\text{op}} > \bar{p} \).
As shown above, the distribution function \( p(x) \) (Eq. 19), when \( A_2 \) is fixed (\( A_2 \) is lost). Then, from Eq. (19), dropping the subscript \( s \) for \( f_2/f_1 = 0.7/0.3. \) as we observe at the third position of the codons in globin and other mammalian mRNAs, we get \( u/w_0 = 0.89. \) In other words, the evolution is retarded by \( \approx 10\% \) from what is expected under complete selective neutrality. Under the more extreme condition \( f_2/f_1 = 0.9/0.1, \) we get \( u/w_0 = 0.49, \) which means \( \approx 50\% \) retardation.

In actual situations, however, there are four possible "alleles" (bases) per nucleotide site rather than two and, together with other complications due to differences in the speed of translation among different types of genes, etc. (16), we need more careful and detailed analysis to arrive at a more accurate figure for the retardation.

**Discussion**

During its lifetime, an individual is subject to natural selection through a large number of quantitative characters, many of which are mutually correlated. Let us assume, to simplify the treatment, that we can choose a certain number, say \( n_c, \) of independent characteristics that collectively represent, to a first approximation, the total pattern of selection. Various parameters pertaining to the \( i \)th character will be expressed by subscript \( i \) (\( i = 1, 2, \ldots, n_c \)).

Because the total selection intensity is limited, the selection intensity, as measured by \( L_i \) at each component character is expected to be small if the number of characters involved is large. Let \( L_T \) be the total selection intensity, then \( 1 - L_T = \Pi(1 - L_i) \) so that \( L_T = 1 - \exp(-\Sigma L_i). \) To simplify the treatment still further, let us suppose that the \( L_i \)s are all equal among component characters, so that \( L_T = -1/n_c \ln(1 - L_i). \) The selection coefficient per site is then \( -\lambda(1/2 - p)\alpha_2. \) and, noting that \( L_i = \lambda_i/2 \) and dropping the subscript \( s, \) we have \( s(i) = \ln(1 - L_i)1/2)\alpha_2^2/n_c \) (approximately). On the right-hand side of this formula, we note that \( n_c^2\sigma_2^2 = \alpha_2^2/n_c^2, \) where \( n_c, \sigma_2^2 \) is the variance of the total phenotype. Let \( h \) be the average heterozygosity per site and, if we denote by \( n_{ue} \) the total number of nucleotide sites concerned, then \( n_{ue}h = n_c\sigma_2^2, \) where \( \sigma_2 \) is the effect of substituting one nucleotide on a component phenotype and \( 1/2 \) is second order heritability. Thus, the coefficient for stabilizing selection \( s_s \) as defined by the relationship \( s = -s_s(1 - 2p), \) turns out to be \( s_s = -\ln(1 - L_T)/\sigma_2^2/n_{ue}h \) (approximately). This represents the selection intensity involved in nucleotide substitution under stabilizing selection (assuming \( m = 0). \)

Let us assume that the average heterozygosity per enzyme locus with respect to electrophoretically detectable alleles is 0.1 and (rather conservatively) that there is twice as much heterozygosity with respect to silent alleles. Then, if the average number of nucleotide sites that comprises a locus is \( 10^4, \) we get \( h = 3.3 \times 10^{-7}. \) Extrapolating this to the total genome of a mammal that has \( 3.5 \times 10^9 \) nucleotide sites, the average number of heterozygous nucleotide sites per individual is \( n_{ue}h = 1.05 \times 10^6. \) As typical values of genetic load and heritability for a mammal (such as the human species), let us suppose that \( L_T = 0.5 \) and \( 1/2 = 0.5, \) then, we obtain \( s_s = 3.3 \times 10^{-7}. \) This is a very small selection coefficient for stabilizing selection and shows that the majority of mutations at the molecular level are nearly neutral but very slightly deleterious. This agrees with Ohta's hypothesis of very slightly deleterious mutations (26, 27). However, the fitness of the species does not drift downward in this view as it does in Ohta's hypothesis. Also, in this view, those genes that are substituted by random drift and those that are responsible for phenotypic variability of quantitative traits belong to the same class. It is possible that many, and even most, of the mu-
tants affecting a quantitative trait are regulatory rather than structural. DNA outside the coding region may be more important from this standpoint than translated DNA. The present analysis agrees with Lande (35), who suggests that many polygenic changes can accumulate by random drift because they have little or no net phenotypic effect.

Needless to say, some sites produce much larger phenotypic effects than others and therefore are subject to stronger selection. On the other hand, a certain fraction of sites (presumably a large fraction) produce no phenotypic effects at all and therefore are completely neutral with respect to natural selection.

The picture of evolution that emerges from the present analysis is as follows. From time to time, the position of the optimum shifts due to changes in environment and the species tracks such changes rapidly by altering its mean. But, most of the time, stabilizing selection predominates. Under this selection, neutral evolution (random fixation of alleles by sampling drift) occurs extensively, transforming all genes, including those of living fossils, profoundly at the molecular level.

I thank Drs. Tomoko Ohta and R. Milkman for stimulating discussions in composing the manuscript. Thanks are also due to Drs. J. Maynard Smith, J. F. Crow, and K. Aoki for helpful suggestions and useful comments to improve the presentation. This is contribution no. 1364 from the National Institute of Genetics, Mishima, Shizuoka-ken, 411 Japan.

