FERTILITY EXCESS NECESSARY FOR GENE SUBSTITUTION
IN REGULATED POPULATIONS

MASATOSHI NEI
Division of Biological and Medical Sciences, Brown University, Providence, Rhode Island 02912
Received May 18, 1970

BIOCHEMICAL studies of certain proteins have recently shown that the rate of evolution or gene substitution is much higher than that previously thought (Kimura 1968; Dayhoff 1969). This high rate of gene substitution has aroused a great deal of controversy about Haldane's (1957; 1960) theory of the cost of natural selection. Haldane postulated that the genetic deaths needed to secure gene substitution by natural selection lower the average fitness of a population, so that there is an upper limit to the number of gene substitutions per unit length of time, depending on the fertility of the species in question. In this case the fitness of a genotype was assumed to be constant. Haldane's theory was accepted by Kimura (1960; 1967) and Crow (1968) but questioned by a number of authors. Among others, Sved (1968) and Maynard Smith (1968) have argued that if the population size is regulated in a density-dependent manner and gene substitution occurs by competitive selection, the mean fitness of the population would not necessarily be reduced. They produced a model of threshold selection in which gene substitutions may occur at a much higher rate than Haldane's computation allows.

Recently, Mather (1969) developed an interesting model of competitive selection and showed that many classical formulae for the change in gene frequency hold true under competitive selection. He, however, argued that the cost of natural selection should be re-examined, taking into account competitive selection and the effect of selection on population size. He thought that the cost of selection would be much less than Haldane's computation suggests. Another type of model for competitive selection was developed by Schutz and Uzanis (1969). In this model, however, neither the cause and process of competition nor the mechanism of population regulation was studied in detail.

The purpose of the present paper is first to develop mathematical models for natural selection in regulated populations and then examine the adequacy of the theory of the cost of natural selection. The cost of natural selection will be looked at from a slightly different angle, and the fertility excess or genetic variance of fitness necessary for a given rate of gene substitution will be studied. It will be shown that competition itself does not lead to threshold selection and Haldane's conclusion is essentially correct.

1 This work was supported in part by U.S. Public Health Service grant FR-07035-04 and also by National Science Foundation grant GB-21289.

Kimura and Crow (1969) developed a mathematical model for genetic selection in haploids, taking into account regulation of population size. Using this model, they showed that the classical formula for the change in gene frequency, which is based on the assumption of constant genotype fitness, holds true even in regulated populations. In this model, however, the genotype that increases in frequency in the population is always the one that has a higher rate of reproduction or a higher resistance to crowding. This does not agree with the experimental observation that there is little correlation between the competitive ability and intrinsic rate of growth or reproduction (Lewontin 1955; Lewontin and Matsuo 1963). In the following we assume that the population size is controlled by two factors, i.e., "intrinsic rate of reproduction" and "competition." Competition may occur through limitations of resources and space, the latter including protective shelters against predation or weather factors such as temperature and humidity. The meaning of competition in this paper is, therefore, broader than the definition by Andrewartha (1961). We further assume that generations are discrete, although a continuous time model can be easily constructed. Discrete generation models are more revealing for the present purpose.

**Haploid model:** Consider a haploid population in which two genotypes, $A$ and $a$, with respect to a locus, are present. Let $n_A$ and $n_a$ be the numbers of adult individuals for genotypes $A$ and $a$, respectively, with $N = n_A + n_a$. The relative frequencies are then $p = n_A/N$ and $q = n_a/N$. In the presence of unlimited resources and space, there will occur no competition, so that the increase of the number of each genotype will be determined by its intrinsic rate of reproduction. In this case, the number of adult individuals for $A$ and $a$ in the next generation are

$$
    n'_A = n_A r_A = n_A k_A v_A 
$$

(1a)

$$
    n'_a = n_a r_a = n_a k_a v_a 
$$

(1b)

respectively. Here $r_A$ and $r_a$ are the intrinsic reproductive values of $A$ and $a$, respectively. The intrinsic reproductive values are constants determined by environmental (physical) conditions and can be written as $kv$'s, where $k$'s and $v$'s are fertility and viability, respectively. In the following we assume for simplicity that $k_A = k_a = k$, and selection occurs through viability, except for a special case. In this connection it is worthwhile to note that most of the fitness models in population genetics are essentially viability models (Bodmer 1965).

(a) **Competition at the individual level:** In nature, however, resources and space are limited, and competition will occur between individuals for the limited resources and space. Competition may occur at the individual level as well as at the genotypic level, i.e., between groups of genotypes. Let us first consider competition at the individual level, which is perhaps more important than the latter type of competition. In this case two or more individuals compete for a unit of food or some other resource (including space), and one of them succeeds in getting it. The number of individuals succeeding in a population depends on the number of such units of resource present. Thus, if the level of resource
resent is small compared with the level required by the competing individuals id the same for all generations, the population size as measured by adult individu
als will reach the saturation level and thereafter remain practically constant.
Ve consider competition at the saturation level, where \( kN \) offspring are produced each generation and \( N \) individuals survive to the adult stage. Namely, the a\( g \) rage survival rate is \( 1/k \). Competition may occur between the same genotypes well as between different genotypes. Since we have assumed no fertility ef
ference between genotypes, competition will occur between \( A \) and \( A \) with equency \( p^2 \) \( (p = kn_i/(kN) = n_i/N) \), between \( A \) and \( a \) with frequency \( 2pq \), and between \( a \) and \( a \) with frequency \( q^2 \).

Suppose that \( A \) has a higher competitive ability than \( a \), and when they com
ite, \( A \) wins with probability \( (1 + s)/2 \), while \( a \) wins with probability \( (1 - s)/2 \.

hen competition occurs between two individuals of the same genotype, one of hem wins with probability \( 1/2 \). The probability that either of the two individuals ins is, of course, one. Therefore, we obtain the probability of success of a geno
pe in each competitive event as given in Table 1. Competition may occur once many times during the life of an organism. If we assume that the fitness of an individual is proportional to the probability of success in competition, then the ombers of adult individuals in the next generation under purely competitive ection are given by

\[
\begin{align*}
n' &= n(1 + sq) \\
\end{align*}
\]

\[
\begin{align*}
n' &= n(1 - sp) \\
\end{align*}
\]

hese are equivalent to the formula for the gene frequencies in the next genera
on derived by Mather (1969), though the present derivation is mathematically ore rigorous. In the derivation of the above formulae, we used pairwise compe
ition. It can be shown, however, that the same formulae hold irrespective of the mber of individuals competing for a unit of resource, if each individual eives independently (see Appendix I). Furthermore, the same formulae are pplicable, even if there are several different niches in the habitat of a population see Appendix II).

Let us now consider the intermediate stage between the geometric growth of a population and the saturation level in which only competitive selection occurs. heoretically, it is possible that the population size \( N \) increases geometrically xponentially) until the saturation level is reached. This would occur if there no competition as long as enough resources are available for sustaining all in-

<table>
<thead>
<tr>
<th>Competition between</th>
<th>Frequency</th>
<th>Probability of success</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A : A )</td>
<td>( p^2 )</td>
<td>1</td>
</tr>
<tr>
<td>( A : a )</td>
<td>( 2pq )</td>
<td>( (1 + s)/2 )</td>
</tr>
<tr>
<td>( a : a )</td>
<td>( q^2 )</td>
<td>( (1 - s)/2 )</td>
</tr>
</tbody>
</table>

**TABLE 1**

*Frequencies of competition occurring between the same and different genotypes and probabilities of success of the two genotypes in the haploid model*
individuals of the population. Thus, some animals may move from one location to another to find new resource when the first location is crowded. Ecologists, however, have established that the general pattern of population growth is not exponential but logistic (Lotka 1956). In the present context this suggests that competition occurs even if the population size is below the saturation level and some amount of resource remains unutilized. Perhaps an unequal distribution of resource among individuals causes some of them to compete with each other even if unutilized resource remains in some locations of the habitat. The following formulation is essentially an extension of the logistic equation to the case of genetic selection with discrete generations. It is an approximation to reality just as the logistic equation is.

Suppose that competitive selection occurs with a relative frequency of $c$ and noncompetitive selection occurs with a frequency of $1 - c$ in a generation. Then, we have

\[ n'_1 = n_1 \left[ (1 - c)r_1 + c(1 + sq) \right] \]
\[ n'_2 = n_2 \left[ (1 - c)r_2 + c(1 - sp) \right] \]

where $c$ is a function of $n_1$ and $n_2$. The simplest form of $c$ would be $N/K$, which is identical with the Verhulst-Pearl effect in the logistic equation. In this case $K$ represents the population size at saturation. If $N = K$, gene substitution occurs only through competitive selection. The analogy to the logistic equation will be clearer when we consider the rate of increase of population size [see (5c)]. If the population size increases exponentially until the saturation level is reached, then $c = 0$ for $N' \leq K$ and $c = 1$ for $N = K$.

The Wrightian fitnesses of genotypes $A$ and $a$ are obtained by $w_i = n'_i/n_i$ and $w_s = n'_s/n_s$, respectively. Namely,

\[ w_i = (1 - c)r_i + c(1 + sq) \]
\[ w_s = (1 - c)r_s + c(1 - sp) \]

From these formulae, we can see that the fitness of a genotype under competitive selection is necessarily dependent on the genotype frequency. The increases in numbers of individuals per generation for the two genotypes and the total population are given by

\[ \Delta n_1 = n_1 [a_1 - c(a_1 - sq)] \]
\[ \Delta n_2 = n_2 [a_2 - c(a_2 + sp)] \]
\[ \Delta N = N\bar{a} (1 - c) \]

where $a_1 = r_1 - 1$, $a_2 = r_2 - 1$, and $\bar{a} = pa_1 + qa_2$. If $c = N/K$ and $0 < \bar{a} < 1$, formula (5c) is equivalent to the well-known logistic equation, which is usually expressed in terms of a continuous time scale. Thus $N$ increases logistically until it becomes equal to $K$.

The amount of change in gene frequency of $A$ per generation ($\Delta p$) can be obtained from (5a). It becomes

\[ \Delta p = \frac{pq[(1 - c)(a_1 - a_2) + cs]}{1 + (1 - c)\bar{a}} \]

This formula shows that if an unsaturated population $p$ does not necessarily increase, if the sign of $a_1 - a_2$ is not the same as that of $s$. However, if the population
size reaches the saturation level, where \( c = 1 \), we have \( \Delta p = spq \). This is identical with the formula obtained by Mather (1969). Further, in the special case of \( u_i - a_i = s \), the above formula becomes equal to the equivalent formula obtained by Kimura and Crow (1969). Experimental studies on competitive selection, however, indicate that \( a_i - a_s \) is rarely equal to \( s \), as mentioned earlier.

In some cases the fertility difference between genotypes \( A \) and \( a \) may be important. In these cases it can be seen that \( \Delta p \) in a saturated population is

\[
\Delta p = pq[(k_i - k_s) + sk_i k_s/k]/k
\]

(7)

where \( k = k_i p + k_s q \). If \( k_i < k_s \), an unstable equilibrium can occur. Therefore, gene \( A \) does not necessarily increase in frequency, even if it has a higher competitive ability. Since, however, we are dealing with genes that are primarily concerned with competitive selection in preadult stages, we will not consider this complication any more. When competition occurs at the time of mating, the same formulation as the above can be made, and the result is essentially the same as that for competition in preadult stages.

(b) Competition at the genotypic level: Let us now consider competition at the genotypic level. This type of competition may occur in a number of cases. For example, one genotype may grow or germinate faster than the other, so that it pre-empts a resource, while the other genotype starves or has a lower probability of survival. A similar situation may occur in competitive struggle for limited shelters against predation or weather factors, if one genotype avoids such hazards with a higher probability than the other. Here the total number of adult individuals at the saturation level is again determined by the level of resource, physical condition, predation, etc. If the probabilities of success or the relative fitnesses of genotypes \( A \) and \( a \) when they compete in equal frequencies are \( (1 + s)/2 \) and \( (1 - s)/2 \), respectively, then the absolute fitness of \( A \) and \( a \) under purely competitive selection with the restriction of constant population size becomes

\[
w_i = (1 + s)/[1 + s(1 - 2q)]
\]

(8a)

\[
w_s = (1 - s)/[1 + s(1 - 2q)]
\]

(8b)

respectively. These formulae are obtained by noting \( \bar{w} = pw_i + qw_s = 1 \). Thus the genotype fitnesses are again gene-frequency dependent. Since \( w_i \) and \( w_s \) can be written as \( 1 + 2sq/[1 + s(1 - 2q)] \) and \( 1 - 2sp/[1 + s(1 - 2q)] \), respectively, the formulae equivalent to (3a)–(6) can be obtained by replacing \( s \) by \( 2s/[1 + s(1 - 2q)] \). In particular,

\[
\Delta p = pq[(1 - c)(a_i - a_s) + 2cs/(1 + s(1 - 2q))] / [1 + (1 - c)a_i]
\]

(9)

In saturated populations, therefore, we have \( \Delta p = 2spq/[1 + s(1 - 2q)] \). This is exactly the same as the classical formula, if \( 2s \) is rewritten as \( s \). If \( s \) is small, we have \( \Delta p = 2spq \), approximately. Therefore, competitive selection between groups of genotypes is twice as effective as that between individuals if \( s \) is the same, but the change of gene frequency can be expressed by the same formula.

Diploid model: Consider the three possible genotypes, \( AA \), \( Aa \), and \( aa \), for a pair of alleles at a locus. Let \( n_1 \), \( n_2 \), and \( n_3 \) be the numbers of adult individuals for \( AA \), \( Aa \), and \( aa \), respectively, with \( n_1 + n_2 + n_3 = N \). The relative frequencies
are, therefore, \( P = n_1/N, \) \( Q = n_2/N, \) and \( R = n_3/N. \) We again assume that selection occurs only through viability and there are no genetic differences in fertility. We denote by \( v_1, v_2, \) and \( v_3, \) the viabilities of \( AA, Aa, \) and \( aa, \) respectively, in the presence of unlimited resources and space, the fertility being \( k \) for all genotypes. Note that \( P, Q, \) and \( R \) do not necessarily follow the Hardy-Weinberg proportions, but the genotype frequencies before selection do. In the presence of unlimited resources and space, the numbers of individuals of \( AA, Aa, \) and \( aa \) in the next generation will be given by

\[
\begin{align*}
    n'_1 &= Np^2kv_1 \\
    n'_2 &= 2Npqkv_2 \\
    n'_3 &= Nq^2kv_3
\end{align*}
\]

respectively, where \( p = P + Q/2 \) is the gene frequency of \( A \) and \( q = 1 - p. \)

(a) **Competition at the individual level**: The numbers of the three genotypes under purely competitive selection can be obtained from Table 2, where the probabilities of success of the three genotypes are given. They become

\[
\begin{align*}
    n'_1 &= Np^2(1 + 2pqsv_1 + q^2sv_2) \\
    n'_2 &= 2Npq(1 - p^2sv_1 + q^2sv_2) \\
    n'_3 &= Nq^2(1 - p^2sv_2 - 2pqsv_2)
\end{align*}
\]

Therefore, the genotype fitnesses of \( AA, Aa, \) and \( aa \) under purely competitive selection are \( w_1 = (1 + 2pqsv_1 + q^2sv_2), w_2 = (1 - p^2sv_1 + q^2sv_2), \) and \( w_3 = (1 - p^2sv_2 - 2pqsv_2), \) respectively, which are again frequency dependent.

The recurrence equations for \( n's \) when both competitive and noncompetitive selections operate are rather complicated. But the changes in the numbers of genes \( A \) and \( a \) (\( n_A = 2Np \) and \( n_a = 2Nq, \) respectively) and the total population size per generation can be written in the same form as those for the haploid model. That is,

\[
\begin{align*}
    \Delta n_A &= n_A [a_A - c(a_A - \bar{s}q)] \\
    \Delta n_a &= n_a [a_a - c(a_a + \bar{s}q)] \\
    \Delta N &= Na(1 - c)
\end{align*}
\]

where \( a_A = k(pv_1 + qv_2) - 1, \) \( a_a = k(pv_1 + qv_2) - 1, \) \( \bar{a} = pa_A + qa_a, \) and \( \bar{s} = p^2sv_1 + pqsv_2 + q^2sv_2, \) respectively. Therefore, the formula for the amount of change in gene frequency also takes the same form as (6) with the parameters defined here. In this case, however, \( a_A, a_a, \) and \( \bar{s} \) are not constant but a function of gene

---

**Table 2**

<table>
<thead>
<tr>
<th>Competition between</th>
<th>Frequency</th>
<th>( AA ) ( \text{Probability of success} )</th>
<th>( Aa ) ( \text{Probability of success} )</th>
<th>( aa ) ( \text{Probability of success} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( AA : AA )</td>
<td>( p^4 )</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( AA : Aa )</td>
<td>( 4p^3q )</td>
<td>( (1 + s_1)/2 ) (1 - ( s_1 ))/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( AA : aa )</td>
<td>( 2p^2q^2 )</td>
<td>( (1 + s_2)/2 ) (1 - ( s_2 ))/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Aa : Aa )</td>
<td>( 4p^2q^2 )</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Aa : aa )</td>
<td>( 4pq^3 )</td>
<td>( (1 + s_2)/2 ) (1 - ( s_2 ))/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( aa : aa )</td>
<td>( q^4 )</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
frequencies. So the change in gene frequency in unsaturated populations could be more complicated than that for a haploid organism.

In saturated populations \( \Delta p \) can be written as
\[
\Delta p = pq(p^2s'_1 + pqs'_2 + q^2s'_3)
\]
which is the same as the formula obtained by Mather (1969) on a slightly different assumption. In the case of genic selection \( s'_2/s'_1 = s''_2 = s''_1 = s''_3 = s' \). Therefore, \( \Delta p = spq \), which is approximately equal to the classical formula, if \( s \) is small. If \( A \) is completely dominant over \( a, s_i = 0 \) and \( s_a = s''_2 = s''_3 = s''_1 = s \), giving \( \Delta p = pq^2s \), which is again similar to the classical formula. In the case of overdominance, however, we get \( \Delta p = pq(-p^2s'_1 + pqs'_2 + q^2s'_3) \) where \( s'_i = -s_i \). At equilibrium \( \Delta p = 0 \), so that the nontrivial equilibrium gene frequency is
\[
\hat{p} = \frac{s_2 - 2s_a + \sqrt{s'^2_2 + 4s'_1s'_3}}{2(s'_1 + s'_2 - s)} \quad 0 < \hat{p} < 1 \tag{14}
\]
The above equilibrium is stable, since
\[
-1 < \left[ \frac{d\Delta p}{dp} \right] \frac{\hat{p}}{1 - \hat{p}} \frac{s'_2 + 4s'_1s'_3}{s} < 0 \tag{15}
\]
Formula (14) does not hold when \( s_i = s'_i + s \). In this case \( \Delta p = pq(-p^2s'_1 + q^2s'_3) \), so that \( \hat{p} = s_i / (s'_i + s) \), which is identical with the classical formula. Therefore, only in the special case of \( s_i = -s'_i + s \) the classical formula holds. The latter formula has been obtained by Mather (1969).

(b) Competition at the genotypic level: When competition occurs at the genotypic level, we denote by \( S_i, S_{s'_i} \) and \( S \) the relative fitnesses of \( AA, Aa, \) and \( aa \) when they compete in equal frequencies. The absolute fitnesses of these three genotypes under purely competitive selection are given by \( w_i = S_i/S, w_s = S_s/S, \) and \( w_a = S_a/S, \) respectively, where \( S = p^2S_i + 2pqS_a + q^2S_s \). The mean fitness \( \bar{w} = p^2w_i + 2pqw_s + q^2w_a \) is, of course, unity. The genotype fitnesses are obviously frequency dependent. This model is essentially the same as that developed by Wright (1969b op. cit. p. 125). It can be shown that the changes in the numbers of genes \( A \) and \( a \) and the total population size per generation are expressed in the same form as (12a)–(12c), if we redefine \( s \) as \( [p(S_i - S_a) + q(S_s - S_a)]/S \).

In saturated populations \( \Delta p \) can be written as
\[
\Delta p = pq[p(w_i - w_s) + q(w_s - w_a)] \tag{16}
\]
If we note \( w_i = S_i/S, w_s = S_s/S, \) and \( w_a = S_a/S, \) it is seen that the above formula is exactly the same as the classical formula, though interpretation is entirely different. With the model of constant fitness the mean fitness always increases, while in the present model \( \bar{w} \) does not change in saturated populations. The equilibrium gene frequency which arises from overdominant selection or mutation-selection balance is the same as that for the model of constant fitness.

FERTILITY EXCESS AND GENETIC VARIANCE NECESSARY FOR GENE SUBSTITUTIONS

In this section we consider only the haploid model, since the diploid model gives essentially the same result as that for the haploid model. The haploid model
can be applied to diploid organisms, if there is no dominance (genic selection). We also restrict our discussion to competitive selection at the individual level for the same reason.

Let us consider gene substitution in a saturated population, since most natural populations are considered to be more or less near the saturation level. In this case \( \Delta p = pq s \). For a gene substitution to proceed with this rate, the fitness of genotype \( A \) should be \( 1 + sq \). Namely, the fertility of an individual (\( k \)) should be equal to or higher than \( 1 + sq \), neglecting the mortality due to environmental causes. If \( k \) is smaller than \( 1 + sq \), the rate of gene substitution is slowed down, but the population size does not decrease unless \( k \) is smaller than unity. This is a mathematical version of the statement by Kimura and Crow (1969) that “if the population does not have the required average fertility, it does not necessarily mean that it becomes smaller; it may only mean that it cannot make gene substitutions at the calculated rate.” Of course, in most organisms \( k \) is much larger than \( 1 + sq \), of which the maximum is close to 2 when \( s = 1 \) and \( q \) is close to 1.

Haldane (1957; 1960) and other authors have calculated the cost of natural selection or substitution load (Kimura 1960) as the total amount of reduction in mean fitness during gene substitution compared with the fitness of the advantageous genotype, which is assumed to be constant. In the present case the mean fitness is constant, while the fitness of the advantageous genotype varies with generation. Nevertheless, we can compute a similar quantity and it turns out to be the same as the cost of natural selection. Namely, approximating \( \Delta p \) by \( dp/dt \),

\[
E = \int_{p_0}^{\infty} (w_t - \bar{w}) \, dt = \int_{0}^{\infty} sq \, dt
\]

\[
= \int_{p_0}^{1} \left( 1/p \right) \, dp = -\log p_0
\]

(17)

where \( p_0 \) is the initial gene frequency. However, I am reluctant to call this quantity the cost of natural selection. Rather, I propose that \( E \) be called accumulated fertility excess necessary for a gene substitution or more briefly fertility excess required, since this is conceptually different from the cost of natural selection. One of the assumptions made in Haldane’s theory is that a change in environment occurs in the habitat of a population, so that the previously disadvantageous gene becomes advantageous. In the present theory no such assumption is necessary, although even in such a situation it works. Nevertheless, its implication on the limit to the number of gene substitutions per unit length of time remains the same.

Recently, Dr. J. F. Crow (1970) stated: “The cost is the excess in survival and fertility that the favored genotype must have in order to carry out the gene substitution at a specified rate, while the entire population stays roughly constant.”

He then computes the cost as \( C = \int_{p_0}^{\infty} \frac{(w_t - \bar{w})}{\bar{w}} \, dt \). (See also Crow 1968.)

This quantity is very similar to \( E \) defined above, if \( \bar{w} \) is close to unity. Therefore, the above concept of fertility excess required is not necessarily new, though he did not use the terminology. Strictly speaking, however, \( \bar{w} \) is not constant in Crow’s formulation, because he used essentially the same fitness model as Hal-
ANE's. For the population size to be constant, selection must be competitive
rather than noncompetitive. Under noncompetitive selection, the population size
either decreases or increases, as pointed out by Moran (1970). Furthermore, even if selection is noncompetitive, which would occur in some cases, the
fertility excess required may be measured more appropriately by $E$ than by $C$,
ince the fertility excess is to be measured in absolute terms. In this case $c$ is al-
ways 0, so that we have
\[ E = (r_1 - r_2) - r_2 \log_2 p_0 \]
approximately. Therefore, only when $r_1$ and $r_2$ are both close to unity, $E$ is nearly
qual to $- \log_2 p_0$. If $r_3 < 1$, $E$ would be reduced, while if $r_3 > 1$ it would increase.

The present formulation indicates that even with competitive selection the
amount of fertility excess required for a gene substitution to proceed is the same
as that predicted by the theory of the cost of natural selection. In this respect
there is not much difference between competitive and noncompetitive selection.
ved (1968) and Maynard Smith (1968) argued that competitive selection re-
ults in threshold selection when many gene substitutions occur at the same time.
ved (1968) states: "If selection occurs as the result of inherited differences in
ility to compete for limiting resources, there is no longer any basis for assum-
ing independent selection against a number of characters. As Milkan (1967)
phasized, it is the individual rather than the locus that is the object of selec-
ion."

Although selection certainly occurs among individuals, it does not necessarily
ply threshold selection. For threshold selection to occur, selection must take
dace through the end product of genotypic expression, as in the case of artificial
election for a quantitative character, since the individuals to be eliminated from
population are those whose total number of advantageous genes is less than a
certain threshold level with some environmental effects. In other words, the genes
olved in competitive selection must have expressed their effects on some
'phenotypic character' before competition occurs and selection must operate
ccording to the scale of this "phenotypic character." Here apparently the same
uation as warned by Robertson (1968) with respect to optimum model selec-
ion is necessary. It is a well-known fact that natural selection takes place at
arious stages of development. If selection or competition at each stage of de-
opment is controlled by a different gene, selection will be "independent" rather
an "threshold." Independent selection may occur even if many genes are con-
cerned with competitive selection at a single developmental stage, if they control
derrent characters concerned with competition. It is instructive to note that the
ovement of a single quantitative character by artificial selection is rather
y but the simultaneous improvement of many characters is a difficult task,
ply because there is limitation in reproductive capacity of an organism (cf.
aldane 1957; Wright 1969a).

While I do not completely deny the existence of threshold selection (for a pos-
able example, the winter survival of a bird determined by its position in a peck
der; Maynard Smith 1968), I believe that gene action in competitive selection
never exceptional and threshold selection is rather rare in nature. Crow (1970)
appears to be of the same opinion. Of course, the relative importance of “independent” and “threshold” selections should eventually be determined by experiments. There are some data on the interaction of viability (deleterious) genes in Drosophila, though these genes may not be important for evolution. They show that there is synergistic interaction between mildly deleterious genes (Dobzhansky, Spassky and Tidwell 1963; Spassky, Dobzhansky and Anderson 1965; Temin et al. 1969; Mukai 1969). Synergistic interaction is expected to reduce the fertility excess required to a certain extent (Kimura and Crow 1969; Crow 1970), but the amount of interaction experimentally observed is not so strong as to reduce it drastically.

Let us now consider the number of possible gene substitutions per generation for a given fertility (k) on the assumption of independent competitive selection and constant population size. Replacing summation by integral, the average fitness of genotype A during the process of gene substitution is given by

$$\bar{f} = \int_{t_0}^{t_1} (1 + sq) dt$$

where $t_0$ and $t_1$ refer to the times at which the gene substitution starts and finishes, respectively. We assume that during this time interval the gene frequency changes from $p_0$ to $p_1$, where $p_0$ and 1- $p_1$ are both close to but not equal to 0. This assumption seems to be reasonable, if we note that the deterministic theory usually fails at the region of gene frequency close to 0 or 1, and that the complete fixation of a gene rarely occurs in a large population owing to spontaneous mutations. Again using the equation $dp/dt = spq$, we have

$$\bar{f} = 1 + s/2$$

(18)

where $p_0$ and 1 - $p_1$ are assumed to be equal to each other. Therefore, the number of gene substitutions that can occur simultaneously (n) is obtained from $k \geq (1 + s/2)^n \approx e^{n/2}$, where $s$ is assumed to be small compared with unity. Namely, $n \leq (2\log_2 k)/s$. In this case, however, $- (2\log_2 p_0)/s$ generations are required for a gene substitution to be completed. Therefore, the number of possible gene substitutions per generation is

$$\nu = \log_2 k/(- \log_2 p_0)$$

(19)

This formula is remarkably simple and independent of selection intensity. It is in accord with the property of the accumulated fertility excess required or Haldane’s cost of natural selection. In fact, the above formula can be derived from the relation $k \geq \exp(- \nu \log_2 p_0)$, which was the basis of Haldane’s argument for the limit to the number of gene substitutions per unit length of time, though he considered noncompetitive selection.

There is another approach to the same question. The fertility excess approach gives only an upper limit to the number of gene substitutions. A better estimate of this number may be obtained by considering the genetic variance of fitness contributed by genes under substitution. In this method fertility differences can also be taken into account (cf. Crow 1968). Crow (1968) used the variance approach to a slightly different problem, i.e., the relationship between the amount of increase in population fitness due to gene substitution and the number of gen-
erations required. More recently, Crow (1970) studied the coefficient of variation ($\sqrt{V_w/\bar{w}}$) and his index of selection ($V_w/\bar{w^2}$) as contributed by genes under substitution, where $V_w$ is the variance of fitness. It seems to me, however, that the genetic variance itself is more meaningful for the present purpose.

In a saturated population the fitness of $A$ and $a$ are $1 + sq$ and $1 - sp$, respectively. Therefore, the variance of fitness is $\nu_w = pqs^2$. The accumulated genetic variance is then $s$, while the average variance is

$$\bar{\nu} = s^2/(-2\log p_0)$$

(20)

Hence if the total genetic variance is $V_g(= n\bar{\nu})$, we have

$$\nu = V_g/s$$

(21)

This formula is again simple but is expected to give a better estimate than formula (19) if $s$ is known, since it does not include the effect of mortality due to environmental factors. $V_g$ can be estimated from the variance and heritability of fertility. The recent study on the variability and heritability of human fertility conducted by Imaizumi, Nei and Furusho (1970) has indicated that the genetic variance of fitness in man is quite small. Of course, human populations are not necessarily stable, so that formula (21) should be slightly modified when applied to human data.

DISCUSSION

The ways in which natural selection operates in a population are generally very complicated. The models developed here are simplifications of such complicated systems of natural selection. There are several other factors to be considered in order to develop a more realistic model. For instance, facilitation (Lewontin 1955) or positive density-factor (Haldane 1956) is sometimes important when population density is low. The population-regulating factor, $c$, may be more complicated than $N/K$ and vary with genotype. It is known that fertility, $k$, is not necessarily constant but often depends on population density. The requirement of different niches by different genotypes may also be important. The effects of these factors are now under investigation. However, our conclusion about the fertility excess necessary for gene substitution does not appear to be much affected by such factors.

As already indicated, competition always results in frequency-dependent selection, i.e., the dependency of genotype fitness upon gene or genotype frequency. Experimental data have shown that natural selection is often frequency dependent (Gustafsson 1951; Levene, Pavlovsky and Dobzhansky 1954; Lewontin 1955; Sokal and Karten 1964; Harding, Allard and Smeltzer 1966; Kojima and Yarbrough 1967; and others). As already indicated by some authors (e.g., Schutz and Uzanis 1969), some, if not a large, part of this frequency-dependency appears to be due to competitive selection.

As we have seen, Haldane's theory of the cost of natural selection appears to be essentially correct even under competitive selection, if it is looked at from the fertility excess required. This has been anticipated by Kimura and Crow (1969) and Crow (1970), as is clear from the earlier quotations of their statements. The
present mathematical formulation of competitive selection, however, would provide a firmer theoretical basis for their statements. Further, the new definition and terminology of *fertility excess required* would help solve various problems posed by Li (1963), Van Valen (1963), Brues (1964; 1969), Mather (1969), and Moran (1970). The term "the cost of natural selection" or "substitutional load" appears to have led a number of authors to misunderstand Haldane's important contribution to evolutionary theory.

In this paper only the deterministic treatment was presented. In reality, however, natural populations are often quite small, so that stochastic elements become important. An elegant theory of the substitutional load in finite populations has recently been developed by Kimura and Maruyama (1969) and Kimura (1969). Most of the results obtained by these authors appear to apply to the present theory of fertility excess required.

Finally, it should be remarked that a certain degree of fertility excess is also necessary for maintaining genetic polymorphisms in populations, whether selection is competitive or noncompetitive. King (1967), Sved, Reed and Bodmer (1967), and Milkman (1967) claimed that a large number of polymorphic loci can be maintained by threshold selection through competition. As mentioned earlier, however, competition itself has nothing to do with threshold selection. Crow (1968) has criticized threshold selection from a different point of view. It seems that the theory of genetic loads as developed by Haldane (1937), Muller (1950), and Crow (1958) is essentially correct even in regulated populations, as far as concerned with random mating populations. Some kinds of genetic loads, however, would be better understood if they were expressed in terms of the fertility excess required. A study on this problem is now under way, taking into account various types of competitive selection and finiteness of population size.

I would like to thank Drs. J. F. Crow and B. R. Levin for their comments on the earlier version of this paper.

*Note added in proof:* After this paper was submitted for publication, Crow and Kimura (1970 op. cit. p. 252) and Ewens (1970) derived the same formula as (21) in the present paper. Using this formula, Ewens has discussed the cost of natural selection, but his conclusion is different from the present one.

**SUMMARY**

Mathematical models of natural selection in regulated populations are developed, taking into account competition. Two types of competition, i.e., competition at the individual level and at the genotypic level, are introduced. It is shown that competitive selection always results in frequency-dependent selection. In a saturated population where the population size is constant, the classical formula for the change in gene frequency holds true in haploids but not necessarily in diploids. In an unsaturated population, the frequency of genes under selection does not necessarily increase or decrease monotonically. Competitive selection itself has nothing to do with threshold selection as claimed by some authors. On the contrary, competitive selection necessitates the same degree of fertility
cess as that required by noncompetitive selection for a given rate of gene substitution to proceed. Haldane’s theory of the cost of natural selection appears to be essentially correct even in regulated populations, if it is expressed in terms of the newly defined fertility excess required for gene substitution. In addition to this parameter, the genetic variance necessary for gene substitution is formulated. These two parameters can be used also in the study of genetic polymorphisms.

LITERATURE CITED


———, 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61: 893–903.


LEVENE, H., 1953 Genetic equilibrium when more than one ecological niche is available. Am. Naturalist 87: 331–333.


GENE SUBSTITUTION BY COMPETITION

APPENDIX I

EXPECTED FITNESSES OF GENOTYPES WHEN COMPETITION OCCURS AMONG MORE THAN TWO INDIVIDUALS

Let us consider a case where three individuals are competing for a unit of resource. In this case there are four possible genotype combinations, of which the expected frequencies are given in Table 3. In the case of competition among three individuals of genotype $A$ or $a$, one of them wins with probability 1/3, so that the probability of success of $A$ or $a$ is 1, as is obvious. When competition occurs among two $A$'s and one $a$, it can be decomposed into the following three pairwise competitive events, if each member behaves individually.

Probability of success

$$
\begin{array}{ccc}
A & a \\
A:A & 1 & \\
A:a & (1 + s)/2 & (1 - s)/2 \\
A:a & (1 + s)/2 & (1 - s)/2 \\
\end{array}
$$

The above three events occur with equal frequencies. Therefore, the expected probabilities of success of $A$ and $a$ are $(2 + s)/3$ and $(1 - s)/3$, respectively. The probabilities of success of $A$ and $a$ in the case of competition among one $A$ and two $a$'s can be obtained in the same way. They are as given in Table 3. The expected fitnesses of genotypes $A$ and $a$ when competition occurs among three individuals then become $(1 + sq)$ and $(1 - ps)$, respectively. Similar arguments show that the genotype fitnesses are the same even if competition occurs among more than three individuals.

TABLE 3

Probabilities of success of the two genotypes in the haploid model when competition occurs among three individuals

<table>
<thead>
<tr>
<th>Competition between</th>
<th>Frequency</th>
<th>Probability of success $A$</th>
<th>Probability of success $a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A : A : A$</td>
<td>$p^3$</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>$A : A : a$</td>
<td>$3p^2q$</td>
<td>$(2 + s)/3$</td>
<td>$(1 - s)/3$</td>
</tr>
<tr>
<td>$A : a : a$</td>
<td>$3pq^2$</td>
<td>$(1 + s)/3$</td>
<td>$(2 - s)/3$</td>
</tr>
<tr>
<td>$a : a : a$</td>
<td>$q^3$</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

APPENDIX II

EXPECTED FITNESSES OF GENOTYPES WHEN MORE THAN ONE ECOCLOGICAL NICHE IS PRESENT

Suppose that there are $n$ different ecological niches or local environments. Let $e_i$ be the proportion of the total adult individuals found in the $i$th niche ($\sum e_i = 1$). Following LEVENE (1953), we assume that at the time of reproduction they leave the niches and mate at random. We further assume that the average viability or intensity of competition is the same for all niches but the competition
coefficient $s$ varies with niche. Let $s_i$ be the competition coefficient in the $i$th niche. Then the expected values of genotype fitnesses for $A$ and $a$ are given by $\Sigma e_i(1 + s_iq) = 1 + \bar{s}q$ and $\Sigma e_i(1 - s_ip) = 1 - \bar{s}p$, respectively, where $\bar{s} = \Sigma e_is_i$. Therefore, formulae (2a) and (2b) hold even in heterogeneous environments if $s$ is replaced by $\bar{s}$. 