Some consideration on diversifying selection

Fumio TAJIMA¹ and Terumi MUKAI²

¹National Institute of Genetics, Mishima, Shizuoka-ken 411
²Department of Biology, Kyushu University 33, Fukuoka 812

(Received 7 May 1990)

ABSTRACT

The diversifying selection due to genotype-environment interaction can increase the genetic variation in natural populations. It is known, however, that the conditions for stable genetic polymorphism or marginal overdominance are quite restricted in this selection model. In this paper a simple model of diversifying selection was examined, and the following results were obtained: (1) Even when the conditions for marginal overdominance are not satisfied, if the diversifying selection is operating, the frequency of mutants can be higher than that in the case of simple mutation-selection balance. (2) This selection model causes a large amount of genetic load (environment load), even when the conditions for marginal overdominance are not satisfied, namely even when the equilibrium frequency of mutant is very low. From these results it can be concluded that the number of loci on which this type of diversifying selection is operating is very small, if any.

1. INTRODUCTION

In natural populations of Drosophila melanogaster, the additive genetic variance for viability shows a north-to-south cline (Mukai, 1985, 1990). The additive genetic variance for viability can be explained by mutation-selection balance in a northern population (Kusakabe and Mukai, 1984), whereas large amount of additive variance observed in southern populations may not be explained by mutation-selection balance (Mukai et al., 1974; Mukai and Nagano, 1983; Tachida et al., 1983). Furthermore, large genotype-environment interaction variance for viability was observed in southern populations (Tachida and Mukai, 1985; Takano et al., 1987), but not in a northern population (Takano et al., 1987). From these observations, Mukai concluded that the diversifying selection due to genotype-environment interaction may be operating on some of viability polygenes (see Mukai, 1990). That is, the southern population might have various environmental conditions as compared with the northern population, so that a large amount of genetic variation can be maintained in the southern population if fitness of genotype is different in different environments.

The effect of genotype-environment interaction on genetic variation was extensively studied as a potential mechanism for the maintenance of genetic variation in a population (e.g., Levene, 1953; Levins, 1962). As indicated by Maynard Smith
(Maynard Smith, 1966; Maynard Smith and Hoekstra, 1980), a stable genetic polymorphism can be maintained only when restricted conditions are satisfied. The amount of genetic variation which can be maintained in the population when these conditions are not satisfied, however, is not well known. In this paper we shall study this problem by using a simple model of diversifying selection. We shall also study the genetic load caused by diversifying selection.

2. MODEL

Assume a large random mating population, and consider a particular locus with two alleles, A and a. The population is assumed to have two environmental conditions, Environments 1 and 2. The proportions of Environments 1 and 2 are denoted by \( f_1 \) and \( f_2 (=1-f_1) \), respectively. The diversifying selection due to genotype-environment interaction is assumed to be operating in such a manner that in Environment 1 allele A is advantageous over allele a while in Environment 2 allele a is advantageous over allele A. This selection scheme is shown in Table 1, where the selection coefficients, \( s_1 \) and \( s_2 \), are assumed to be substantially smaller than unity. We denote the mutation rate from allele A to allele a by \( u \) per generation, and that from allele a to allele A by \( v \) per generation, and assume that \( u \) and \( v \) are substantially smaller than \( s_1 \) and \( s_2 \).

<table>
<thead>
<tr>
<th>Genotype Frequency</th>
<th>AA   ( p^2 )</th>
<th>Aa   ( 2pq )</th>
<th>aa   ( q^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment 1</td>
<td>1</td>
<td>1 - ( h_1 s_1 )</td>
<td>1 - ( s_1 )</td>
</tr>
<tr>
<td>Environment 2</td>
<td>1 - ( s_2 )</td>
<td>1 - ( h_2 s_2 )</td>
<td>1</td>
</tr>
<tr>
<td>Average</td>
<td>1 - ( f_2 s_2 )</td>
<td>1 - ( f_1 h_1 s_1 + f_2 h_2 s_2 )</td>
<td>1 - ( f_1 s_1 )</td>
</tr>
</tbody>
</table>

\( f_1 \) and \( f_2 \) are the proportions of Environments 1 and 2, respectively.

3. THEORY

Equilibrium gene frequency

Let the frequencies of alleles A and a be \( p \) and \( q (=1-p) \), respectively. Then, the change in the frequency of allele A is approximately given by

\[
\Delta p \approx pq [ | f_1 h_1 s_1 - f_2 (1-h_2) s_2 | p + | f_1 (1-h_1) s_1 - f_2 h_2 s_2 | q ] - pu + qv. \tag{1}
\]

From \( \Delta p = 0 \) we can obtain the equilibrium gene frequency. Here we examine several cases.

(a) \( f_1 s_1 < f_1 h_1 s_1 + f_2 h_2 s_2 \) and \( f_1 s_1 < f_2 s_2 \): In this case the average fitness of genotype aa is higher than those of genotypes Aa and AA, so that the equilibrium
frequency of allele A is approximately given by

\[ \hat{p} \approx \frac{v}{f_2 h s_2 - f_1 (1-h_1) s_1} . \]  \hspace{1cm} (2)

When \( h_1 = h_2 = h \) and \( s_1 = s_2 = s \), this formula becomes

\[ \hat{p} \approx \frac{v}{(h-f_1)s} , \]  \hspace{1cm} (3)

which indicates that the equilibrium frequency of allele A increases as the proportion of Environment 1 increases. This formula also indicates that \( \hat{p} \) is larger than \( v/hs \).

(b) \( f_1 s_1 > f_1 h_1 s_1 + f_2 h_2 s_2 \ll f_3 s_2 \): In this case allele A behaves like a recessive, deleterious allele, so that the equilibrium frequency of allele A is approximately given by

\[ \hat{p} \approx \frac{v}{f_2 s_2 - f_1 s_1} \]  \hspace{1cm} (4)

When \( s_1 = s_2 = s \), this formula becomes

\[ \hat{p} \approx \frac{v}{(1-2f_1)s} \]  \hspace{1cm} (5)

which is larger than \( \sqrt{v/s} \).

(c) \( f_1 s_1 > f_1 h_1 s_1 + f_2 h_2 s_2 \ll f_3 s_2 \): This case satisfies a condition for stable genetic polymorphism or marginal overdominance, so that, unlike in the other cases, the gene frequency can be intermediate. In this case the equilibrium frequency of allele A is approximately given by

\[ \hat{p} \approx \frac{f_1 (1-h_1) s_1 - f_2 h_2 s_2}{f_1 (1-2h_1) s_1 + f_2 (1-2h_2) s_2} . \]  \hspace{1cm} (6)

When \( h_1 = h_2 = h \) and \( s_1 = s_2 = s \), this formula becomes

\[ \hat{p} \approx \frac{f_1 - h}{1-2h} . \]  \hspace{1cm} (7)

(d) \( f_1 s_1 = f_1 h_1 s_1 + f_2 h_2 s_2 = f_3 s_2 \): In this case the average fitnesses are the same for all the genotypes, namely the alleles are selectively neutral [for the neutral theory, see Kimura (1968, 1983)]. Therefore, the equilibrium frequency of allele A is given by

\[ \hat{p} = \frac{v}{u+v} . \]  \hspace{1cm} (8)
Genetic load

The genetic load under the equilibrium condition is defined as

\[ L = \frac{W_{\text{max}} - \bar{W}}{W_{\text{max}}} \]  \hspace{1cm} (9)

(Crow, 1958), where \( W_{\text{max}} \) is the highest fitness and \( \bar{W} \) is the average fitness in the population. In the present selection scheme it becomes

\[ L = f_1(2\hat{p}\hat{q}h_1s_1 + \hat{q}^2s_1) + f_2(2\hat{p}\hat{q}h_2s_2 + \hat{p}^2s_2). \]  \hspace{1cm} (10)

Let us now examine several cases as before.

(a) \( f_1s_1 < f_1h_1s_1 + f_2h_2s_2 \) and \( f_1s_1 < f_2s_2 \): In this case the equilibrium frequency of allele A is given by (2). Substitute (2) into (10), we have

\[ L \approx 2v + f_1s_1, \]  \hspace{1cm} (11)

which indicates that the genetic load is larger than that of deleterious mutant with partial dominance by \( f_1s_1 \). If we define environment load as \( f_1s_1 \), then the genetic load can be expressed as

Genetic Load = Mutation Load + Environment Load  \hspace{1cm} (12)

since the mutation load is \( 2v \) in this case.

(b) \( f_1s_1 = f_1h_1s_1 + f_2h_2s_2 \leq f_2s_2 \): In this case the substitution of (4) into (10) gives

\[ L \approx v + f_1s_1, \]  \hspace{1cm} (13)

which indicates that the genetic load is larger than that of recessive, deleterious mutant by \( f_1s_1 \). Since \( v \) is the mutation load in this case, equation (12) still holds.

(c) \( f_1s_1 > f_1h_1s_1 + f_2h_2s_2 \leq f_2s_2 \): As mentioned above, this case satisfies a condition for stable genetic polymorphism or marginal overdominance. In this case, substituting (6) into (10), we obtain

\[ L \approx \frac{f_1s_1f_2s_2 - (f_1h_1s_1 + f_2h_2s_2)^2}{f_1(1-2h_1)s_1 + f_2(1-2h_2)s_2}. \]  \hspace{1cm} (14)

When the fitnesses of genotypes AA, Aa, and aa are \( 1-S, 1, \) and \( 1-T \), respectively, the segregation load is given by \( ST/(S+T) \) [see Crow and Kimura (1970)]. In the present case \( S \) and \( T \) can be regarded as

\[ S \approx f_2s_2 - (f_1h_1s_1 + f_2h_2s_2), \]  \hspace{1cm} (15a)

\[ T \approx f_1s_1 - (f_1h_1s_1 + f_2h_2s_2), \]  \hspace{1cm} (15b)

approximately (see Table 1). Then the segregation load is approximately given by

\[ L_s \approx \frac{|f_2s_2 - (f_1h_1s_1 + f_2h_2s_2)| |f_1s_1 - (f_1h_1s_1 + f_2h_2s_2)|}{f_1(1-2h_1)s_1 + f_2(1-2h_2)s_2}. \]  \hspace{1cm} (16)
If we define environment load as \( f_1 h_1 s_1 + f_2 h_2 s_2 \) in this case, then the genetic load can be expressed as

\[
\text{Genetic Load} = \text{Segregation Load} + \text{Environment Load.} \quad (17)
\]

The definition of environment load in this case is different from that of (12). However, if we redefine the environment load as the smallest value among \( f_1 h_1 s_1 + f_2 h_2 s_2 \), and \( f_2 s_2 \), then this difference can be explained.

(d) \( f_1 s_1 = f_1 h_1 s_1 + f_2 h_2 s_2 = f_2 s_2 \): In this case, from (10) with (8) we have

\[
L = f_1 s_1, \quad (18)
\]

which is the environment load. This indicates that the genetic load is not zero although the equilibrium gene frequency is determined solely by the mutation rates.

4. CONCLUSION AND DISCUSSION

Relationship between the equilibrium gene frequency and the proportion of environmental condition

In the present model of diversifying selection the condition for stable genetic polymorphism or marginal overdominance is

\[
|f_1 s_1 > f_1 h_1 s_1 + f_2 h_2 s_2 < f_2 s_2|. \quad (19)
\]

For simplicity, consider only the case where the degree of dominance and selection coefficient in Environment 1 are the same as those of Environment 2, namely \( h_1 = h_2 = h \) and \( s_1 = s_2 = s \). Then, this condition becomes

\[
|f_1 > h < f_2| \text{ or } |h < f_1 < 1-h|, \quad (20)
\]

where \( h \) is smaller than 1/2. Under this condition the equilibrium frequency of allele A is approximately given by (7). When \( h \) is close to 1/2, however, the range where this condition is satisfied is very limited. For example, if \( h = 0.4 \) (Mukai and Yamazaki, 1968; Mukai 1969), this condition is satisfied only when \( 0.4 < f_1 < 0.6 \). This conclusion is consistent with Maynard Smith and Hoekstra (1980).

Even when the condition for stable genetic polymorphism or marginal overdominance is not satisfied, however, if the genotype-environmental interaction is operating, the equilibrium gene frequency can be higher than that in the case of simple mutation-selection balance. For example, when \( f_1 < h \) and \( f_1 < 1/2 \), the equilibrium frequency of allele A is approximately given by (3), which is larger than \( \nu / (hs) \). Table 2 shows one example where \( h = 0.4 \), \( s = 0.03 \) (Mukai, 1964; Mukai et al., 1972) and \( u = v = 0.0001 \) are assumed. In this table the equilibrium frequency of allele A was obtained by using (1) directly. From this table we can see that in the present case the equilibrium gene frequency can be larger than
Table 2. Effect of the proportion of environmental condition on the equilibrium gene frequency and the genetic load

<table>
<thead>
<tr>
<th>Proportion of Environment 1</th>
<th>Equilibrium frequency of allele A</th>
<th>Genetic load (Environment load)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0082</td>
<td>0.0002 (0)</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0110</td>
<td>0.0032 (0.003)</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0164</td>
<td>0.0062 (0.006)</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0318</td>
<td>0.0092 (0.009)</td>
</tr>
<tr>
<td>0.4</td>
<td>0.1323</td>
<td>0.0120 (0.012)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0.0135 (0.012)</td>
</tr>
<tr>
<td>0.6</td>
<td>0.8677</td>
<td>0.0120 (0.012)</td>
</tr>
<tr>
<td>0.7</td>
<td>0.9682</td>
<td>0.0092 (0.009)</td>
</tr>
<tr>
<td>0.8</td>
<td>0.9836</td>
<td>0.0062 (0.006)</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9890</td>
<td>0.0032 (0.003)</td>
</tr>
<tr>
<td>1</td>
<td>0.9918</td>
<td>0.0002 (0)</td>
</tr>
</tbody>
</table>

$h_1=h_2=h=0.4$, $s_1=s_2=s=0.08$, and $u=v=0.0001$ are assumed. The equilibrium frequency of allele A was obtained by using (1), the genetic load was obtained from (10), and the environment load was obtained from the definition.

that of simple mutation-selection balance. For example, $\hat{p}=0.0318$ when $f_1=0.3$. Note that $\hat{p}=0.0333$ if (3) is used. The fact that the two values are very close to each other indicates that (3) is a good approximation even when $h$ is not substantially larger than $f_1$. In any case this value of $\hat{p}$ is about four times larger than 0.008 which is the equilibrium gene frequency in the case of simple mutation-selection balance.

Relationship between the equilibrium gene frequency and genetic load

As mentioned above, this diversifying selection model causes a large amount of genetic load and a majority of this large genetic load comes from the environment load. From (11), (13) and (18) we can see that the genetic load increases almost linearly with $f_1$, and that the genetic load is very large even when the condition for stable genetic polymorphism or marginal overdominance is not satisfied. This means that the diversifying selection model examined here creates a large amount of genetic load (environment load) even when the equilibrium gene frequency is very low. Table 2 shows the relationship between the genetic load (environment load) and the proportion of environmental condition, and we can see from this table that the genetic load increases almost linearly with $f_1$ until it reaches the maximum value and then decreases almost linearly with $f_1$.

From the above consideration we can conclude that the number of loci on which this type of diversifying selection is operating is very small, if any, because of large genetic load (environment load). Some model of diversifying selection such as habitat selection may not cause a large amount of genetic load. In this case,
however, strong linkage disequilibrium might be observed in the population, but
this is not the case at least in natural populations of D. melanogaster (Mukai and
Yamaguchi, 1974; Mukai and Voelker, 1977; Yamaguchi et al., 1980).

I thank Dr. H. Tachida and an anonymous reviewer for their useful suggestions and comments (F. T.). This is contribution no. 1888 from the National Institute of Genetics, Mishima, 411, Japan, and paper no. 59 from the Laboratory of Population Genetics, Department of Biology, Kyushu University, Fukuoka, 812, Japan.

REFERENCES

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