Conditions for Positive and Negative Correlations Between Fitness and Heterozygosity in Equilibrium Populations

Hong-Wen Deng* and Yun-Xin Fu

*Osteoporosis Research Center, Creighton University, Omaha, Nebraska 68131 and 'Human Genetics Center, SPH, University of Texas, Houston, Texas 77225

Manuscript received August 12, 1997
Accepted for publication November 7, 1997

ABSTRACT

The past decades have witnessed extensive efforts to correlate fitness traits with genomic heterozygosity. While positive correlations are revealed in most of the organisms studied, results of no/ negative correlations are not uncommon. There has been little effort to reveal the genetic causes of these negative correlations. The positive correlations are regarded either as evidence for functional overdominance in large, randomly mating populations at equilibrium, or the results of populations at disequilibrium under dominance. More often, the positive correlations are viewed as a phenomenon of heterosis, so that it cannot possibly occur under within-locus additive allelic effects. Here we give exact genetic conditions that give rise to positive and negative correlations in populations at Hardy-Weinberg and linkage equilibria, thus offering a genetic explanation for the observed negative correlations. Our results demonstrate that the above interpretations concerning the positive correlations are not complete or even necessary. Such a positive correlation can result under dominance and potentially under additivity, even in populations where associated overdominance due to linked alleles at different loci is not significant. Additionally, negative correlations and heterosis can co-occur in a single population. Although our emphasis is on equilibrium populations and for biallelic genetic systems, the basic conclusions are generalized to non-equilibrium populations and for multi-allelic situations.

During the past three decades, numerous efforts have attempted to correlate fitness (or related characters) with genomic heterozygosity as reflected by molecular marker heterozygosity in natural populations (Mitton and Grant 1984; Allendorf and Leary 1986; Zouros and Foltz 1987; Lynch and Walsh 1997). While a number of studies have found no or negative correlations (e.g., Gains et al. 1978; Pierce and Mitton 1982), positive correlations have been revealed in most organisms studied. There has been little effort to search for satisfactory genetic explanations for the negative correlations, and it is generally regarded that with a parametric slope of nearly zero, negative correlations can result by chance. On the other hand, the more commonly observed positive correlations have stimulated a great deal of research interest and there have been some distinct genetic explanations for them.

One explanation is that (Mitton and Grant 1984; Smouse 1986; Zouros and Foltz 1987) overdominance is the cause, and the positive correlation is a piece of evidence for overdominance underlying the fitness loci. This is especially true in large randomly mating populations in which linkage disequilibrium is not significant. The basis for this view is that with overdominance, one would expect individual fitness to increase with the fraction of the genome that is heterozygous. This fraction should be correlated with the number of heterozygous molecular marker loci if they are tightly linked to polymorphic loci underlying fitness or themselves influence fitness. Under dominance, a correlation between multilocus heterozygosity and fitness cannot arise except when there is a correlation between multilocus homozygosity and the level of inbreeding. Such an association is unlikely to be pronounced in large, randomly mating populations. However, the overdominance argument encounters some difficulties. For example, there are several findings of heterozygote deficiency in populations showing positive correlations between heterozygosity and fitness-related traits (Gaffney et al. 1990; Lynch and Walsh 1997), which should not be expected in large randomly mating populations with the overdominance hypothesis.

Another explanation is that the populations studied may not be strictly panmictic but instead have local inbreeding. Nonrandom mating may cause correlations between homozygosity in the genome, even with unlinked loci (Haldane 1949; Ohta and Cockerham 1974; Houle 1994). The allozyme heterozygosity may thus be correlated with an individual level of inbreeding, and hence the associations between heterozygosity and fitness are largely the consequence of variation in the level of inbreeding among individuals (Ledig et al. 1983; Strauss 1986). However, this explanation was not
supported by a few studies (e.g., Leary et al. 1987), and large studies in marine bivalves normally exclude inbreeding as an explanation for the observed positive correlation (Gaffney et al. 1990). In randomly mating populations, if linkage disequilibrium is present, regardless of its causes, linked deleterious alleles with dominance at different loci will likely cause associated overdominance (Houle 1989). This also may explain the positive correlation. However, despite a large number of studies, evidence for linkage disequilibrium in large randomly mating populations remains controversial (e.g., Barker 1979; Lewontin 1985; Smit-McBride et al. 1988; Houle 1989; Zapatia and Alvarez 1992, 1993; Lynch and Deng 1994; Deng and Lynch 1996a).

The correlation approach for fitness and genomic heterozygosity and the different explanations for the observed data are highly relevant to one fundamental and long-standing issue in population genetics: How is genetic variation maintained in natural populations? Overdominance essentially encompasses all forms of balancing selection at the allelic level and dominance is compatible with mutation-selection balance. The following two inferences are common to the above two genetic explanations. First, the positive correlation reflects the phenomenon of heterosis. Hence in large randomly mating populations, it cannot possibly exist under within-locus additive allelic effects. In addition, heterosis should be incompatible with the negative correlations and they cannot co-occur within individual populations. Second, in populations at genetic equilibria, the positive correlations cannot exist with dominance. These concepts have been widely held among the researchers in this field. However, are they always true? Throughout, unless otherwise specified, (genetic) equilibria refer to Hardy-Weinberg and/or linkage equilibria.

Employing a multilocus biallelic model, and by theoretical analyses supplemented by computer simulations, we show that these two concepts are not true. Moreover, we demonstrate that negative correlations between fitness and genomic heterozygosity are not unexpected, and we give explicit genetic conditions for both positive and negative correlations to occur. Our focus is on equilibrium populations. However, the conclusions derived for multiple loci under equilibrium are generalized to nonequilibrium populations and multi-allelic systems through one locus model.

There are extensive data existing for the discovered correlations. There are also some potential limitations of the correlation approach for fitness and genomic heterozygosity (see discussion). Therefore, our focus here is to show when correlations do indeed exist in equilibrium populations, and how to interpret them when they are detected. Hence, some practical issues are not dealt with here, such as what sample sizes are needed and how many loci need to be assayed in order to detect a correlation when it indeed exists. Some of these, or related practical problems, have been addressed before (e.g., Miltton and Pierce 1980; Chakraborty 1981).

**THEORY**

Consider a simplified situation, in which there are N polymorphic loci underlying fitness, each having two alleles A and a. The allelic effects across loci may vary so that the equilibrium frequencies for the ith locus is $p_i$, $q_i$, respectively. Let the three genotypic values be:

\[
\begin{align*}
AA &= 1 - h_i s_i \\
Aa &= 1 - s_i \\
aa &= 1 - q_i s_i
\end{align*}
\]

Then for the ith locus, $h_i < 0.0$ implies overdominance, $h_i = 0.5$ implies additivity, $0 < h_i < 1.0$ ($h_i > 0.5$) implies dominance and $h_i > 1.0$ implies underdominance. For the time being, we will assume $h_i = h$, $s_i = s$, and that mutation-selection balance has been established so that $q_i = q$ where $q$ is the frequency of the less fit allele $a$ at the ith locus. We will, later in the discussion section, consider the situation where $h_i$, $s_i$, and $q_i$ vary across loci. $q_i$ may vary across loci either due to variable $h_i$ and $s_i$ in populations at mutation-selection balance, or in populations experiencing recent expansion after a population bottleneck (even with constant $h$ and $s$), where genetic disequilibrium could be negligible but mutation-selection balance has not been reached.

The multiplicative fitness function is biologically plausible by direct and indirect evidence (Morten et al. 1956; Crow 1986; Fu and Ritland 1996) and will be assumed throughout. The fitness $W(n_1,n_2)$ of an individual is totally determined by the number of heterozygous ($n_1$) and homozygous ($n_2$) loci for allele in the genome, regardless of the specific genotypes at particular loci:

\[
W(n_1,n_2) = (1 - h s)^n_1 (1 - s)^n_2,
\]

where $0 \leq n_1 + n_2 \leq N$.

Under random mating and linkage equilibrium, genomic genotypes for individuals, as determined by $n_1$ and $n_2$, follow a trinomial distribution:

\[
f(n_1,n_2) = \frac{N!}{n_1! n_2! (N - n_1 - n_2)!} (2pq)^{n_1} p^{2(n_1 - n_2 - q)x} q^{n_2}.
\]

Conditional on having $n_1$ heterozygous loci, the probability of having $n_2$ homozygous loci for the $a$ allele in the genome is a binomial distribution:

\[
f(n_2|n_1) = \frac{(N - n_1)!}{n_2!(N - n_1 - n_2)!} (pq)^{(N-n_1-n_2)/q} (1 - 2pq)^{n_2}.
\]

Thus, conditional on having $n_1$ heterozygous loci in the genome, the expected fitness of a genotype [E(W/n_1)] is:
\[
E(W/n_j) = \sum_n \frac{(N - n_j)^!}{n_j(N - n_j - n_j)!} \left( \frac{p^2}{1 - 2pq} \right)^{n_j} \left( \frac{q^2(1 - s)}{1 - 2pq} \right) (1 - hs)^n_j
\]

\[
= (1 - hs)^n_j \left( \frac{q^2(1 - s)}{1 - 2pq} + \frac{p^2}{1 - 2pq} \right) \left( \frac{1 - hs}{1 - 2pq} \right)^{n_j}
\]

\[
= (1 - hs)^n_j \left( 1 - \frac{q^2}{1 - 2q + 2q^2} \right)^{n_j}
\]

Equation 4 can be re-written as:

\[
E(W/n_j) = \left( \frac{q^2(1 - s)}{1 - 2pq} + \frac{p^2}{1 - 2pq} \right) \left( \frac{1 - hs}{1 - 2pq} \right)^{n_j}
\]

It is then easy to see that \( E(W/n_j) \) is a monotonically increasing function of \( n_j \), if

\[
1 - hs > \frac{q^2(1 - s)}{1 - 2pq} + \frac{p^2}{1 - 2pq},
\]

otherwise it is a monotonically decreasing function of \( n_j \). Let

\[
h_c = \frac{q^2}{1 - 2q + 2q^2};
\]

then Equation 6 is equivalent to

\[
h < h_c.
\]

In other words, if the heterozygote fitness is larger than the weighted mean fitness of the two homozygotes (Equation 6), or if \( h \) satisfies Equation 8, then the expected fitness of a genotype increases monotonically with the number of heterozygous loci \( n_j \) in the genome; otherwise it decreases monotonically with \( n_j \). This result may be easier to understand when we only consider, later in this section, the simplest case of a single locus with two alleles.

Therefore, both positive and negative relationships between fitness and genomic heterozygosity can exist. Which one exists critically depends on the above condition (Equation 6 or 8) under our assumptions of random mating, no significant linkage disequilibrium, and multiplicative fitness function. Figure 1 depicts the parameter space of \( h \) and \( q \) that gives rise to positive and negative correlations between fitness and genomic heterozygosity. Note that Figure 1 does not imply any true relationship of \( h \) and \( q \). It just graphically depicts the outcome regions of the relationship of fitness and heterozygosity given the true relationship of \( h \) and \( q \). It can be seen that negative correlations are possible under a large range of parameter space of \( h \) and \( q \).

Figure 1.—Parameter spaces of \( h \) (Y-axis) and \( q \) (X-axis) for positive and negative correlations to exist in population at genetic equilibria. The curve is based on Equation 7. Region A indicates the parameter space where both population inbreeding depression (heterosis) and a negative correlation can occur.
The essential question is then: Which has a higher fitness, a heterozygote or a homozygote? With the heterozygote, the fitness is $1 - hs$. With a homozygote, it can be either AA or aa with respective frequencies in the population under Hardy-Weinberg equilibrium being $p^2$ and $q^2$. Thus the expected fitness of a homozygote in the population is

$$\frac{p^2}{p^2 + q^2} * 1 + \frac{q^2}{p^2 + q^2} *(1 - s),$$

which is

$$1 - \frac{q^2}{p^2 + q^2} * s.$$

Therefore, if

$$h < h_c = \frac{q^2}{p^2 + q^2} = \frac{q^2}{1 - 2q + 2q^2}, \quad (9)$$

a heterozygote has a higher expected fitness than a homozygote. Otherwise, if $h > h_c$, a heterozygote has lower expected fitness than a homozygote. If $q$ is such that $h > 0.5$, then the heterozygote has a higher expected fitness than a homozygote regardless of allelic effects being overdominant ($h < 0$), dominant ($0 < h < 1.0$, $h \neq 0.5$) or additive ($h = 0.5$). Note that Equation 9 is actually equivalent to Equations 7 and 8. This is exactly the condition we found earlier when we considered multiple loci under a multiplicative fitness function (Equations 6, 7 and 8), where we assumed constant $h$, $s$, $q$, linkage equilibrium and mutation-selection balance. However, in the one locus case, no assumption needs to be made about the linkage disequilibrium and mutation-selection balance. Therefore, regardless of linkage disequilibrium and mutation-selection balance, as long as Equation 9 is satisfied at each locus, a positive correlation will exist.

Now, let us relax our assumptions even further and assume a general population where even Hardy-Weinberg equilibrium may not hold. A population resulting from the mixing of different populations may represent such a scenario. The essential question still is: At each polymorphic locus, which has a higher fitness, a heterozygote or a homozygote? For a biallelic locus as above, let us denote the genotype frequencies as $P_{AA}$, $P_{AB}$ and $P_{BB}$ respectively. The expected fitness of a genotype conditional on that it is homozygous is:

$$\frac{P_{AA}}{P_{AA} + P_{AA}} * 1 + \frac{P_{AB}}{P_{AA} + P_{AB}} *(1 - s) = 1 - \frac{P_{AA}}{P_{AA} + P_{AB}} * s.$$

Again, this is the weighted fitness of the homozygotes, with the weight being their conditional genotype frequencies. Therefore, if

$$1 - hs > 1 - \frac{P_{AA}}{P_{AA} + P_{AB}} * s,$$

a positive correlation will exist. The above inequality is equivalent to

$$h < h_c = \frac{P_{AA}}{P_{AA} + P_{AB}}. \quad (10)$$

Note there is no assumption concerning the (Hardy-Weinberg/ linkage disequilibrium) equilibrium in the above derivation. It can be easily seen that if Hardy-Weinberg equilibrium is assumed, Equation 10 can be reduced to Equation 9. If $h > h_c$, a negative correlation will exist.

All the above analyses are for biallelic genetic systems, which are applicable for many allozyme loci and restriction fragment polymorphisms (RFLPs). However, some allozyme loci have more than two alleles and the increasingly employed micro-satellite marker loci are even more polymorphic. In the following, we are going to give the general genetic conditions of the correlation relationships for multi-allelic systems. We will use the tri-allelic genetic system as an example; extensions to genetic systems with more alleles are straightforward and can be obtained similarly.

Let the genotypic values and frequencies of a tri-allelic locus be:

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>BB</td>
<td>CC</td>
</tr>
<tr>
<td>1</td>
<td>1 - s_1</td>
<td>1 - s_2</td>
</tr>
<tr>
<td>1 - s_1 h_1</td>
<td>1 - s_1 h_2</td>
<td>1 - s_1 h_3</td>
</tr>
</tbody>
</table>

The essential question is again the same as before: Which has a higher fitness, a heterozygote or a homozygote? The expected fitness of a genotype given that it is homozygous is:

$$\frac{P_{AA} * 1 + P_{BB} * (1 - s_1) + P_{CC} * (1 - s_2)}{P_{AA} + P_{BB} + P_{CC}} = 1 - \frac{P_{BB} * s_2 + P_{CC} * s_3}{P_{AA} + P_{BB} + P_{CC}},$$

and the expected fitness of a genotype given that it is heterozygous is:
A more counterintuitive conclusion is that with the definition of \( P_{ABs} \), \( P_{CCs} \) and \( P_{BCs} \), which directly link the genetic effect \( h \) (and \( s \)) with the population property of gene (or genotype) frequencies. The conditions for these counterintuitive phenomena to exist do not seem to be prohibitive given some level of biological knowledge. For a positive correlation to exist, \( q \) does not have to be very common in the dominance case (Figure 1). In the additive case, in order for \( h_1 > 0.5 \) so that a positive correlation could possibly exist (Equations 6, 8 and 9), \( q \) has to be greater than 0.5 in populations at genetic equilibrium or \( P_{AA} < P_{B0} \) in populations at disequilibrium (Equation 10). Are these entirely impossible? Our knowledge is very limited on fitness effects at polymorphic loci such as those revealed by molecular markers (Kimura 1983; Nei 1987; Li 1997). For example, how is the extensive polymorphism in natural populations maintained? What is the difference in the fitness effects of different polymorphisms? How much genetic variation of fitness can polymorphism at a particular single locus explain? What is the mutation rate at a locus? Our knowledge about the population history, such as population dynamics and population admixture, is also very limited. Therefore, the possibility of \( q > 0.5 \) cannot be entirely ruled out in at least three conceivable situations. The first is in populations that have been large for long enough generations, so that a mutation-selection balance has been approximately reached. For instance, for slightly deleterious mutations (Ohta 1973, 1974), it is not unreasonable to assume \( s = 0.0001 \). Even in the most prohibitive case of additive effects, with mutation-selection balance, the locus mutation rate \( u \) is inferred to be on the order of 1.0E-5 (Crow and Kimura 1970) in order for \( q > 0.5 \). This is roughly on the order of the few mutation rates inferred for allozyme loci (Hartl and Clark 1989; Maynard-Smith 1989). For instance, Schléger and Dickie (1971) estimated that, for five loci tested in the mouse, the average mutation rate is 1.1E-5. Since this is only for visible mutants, it is likely to be a lower bound. It is noted that by inferring the order of \( u \), reversible mutations are ignored. This may be partially justified by the fact that mutations at many of its nucleotide sites within a locus may be slightly deleterious; however, in a mutant allele, only reversible mutations at those few mutated nucleotide sites can restore the original wild-type allele.

\[
P_{AB} \cdot (1 - s) + P_{AC} \cdot (1 - s) + P_{CC} \cdot (1 - s) =
\]

\[
\frac{P_{AB} \cdot P_{AC} + P_{BC}}{P_{AA} + P_{AB} + P_{AC}}
\]

Let \( s = \frac{P_{BB} \cdot s_1 + P_{CC} \cdot s_2}{P_{AA} + P_{BB} + P_{CC}}. \) That is, \( s \) is the weighted homozygous effect of the less fit alleles, with the weight being the conditional probabilities

\[
\begin{align*}
&\left(\frac{P_{BB}}{P_{AA} + P_{BB} + P_{CC}}\right) \text{ and } \left(\frac{P_{CC}}{P_{AA} + P_{BB} + P_{CC}}\right) \\
&\text{of being a specific less fit homozygous genotype, given that the genotype is homozygous.}
\end{align*}
\]

Let \( \overline{h} = \frac{P_{AB} \cdot s_1 + P_{AC} \cdot s_2 + P_{BC} \cdot s_3}{P_{AA} + P_{AC} + P_{BC}}. \) That is, \( \overline{h} \) is the weighted dominance coefficient, with the weight being the product of the corresponding selection coefficients \( (s_1, s_2 \text{ and } s_3) \) and the conditional probabilities

\[
\begin{align*}
&\left(\frac{P_{AB}}{P_{AA} + P_{AB} + P_{BC}}\right) \text{ and } \left(\frac{P_{AC}}{P_{AA} + P_{AC} + P_{BC}}\right) \text{ and } \left(\frac{P_{BC}}{P_{AA} + P_{AC} + P_{BC}}\right) \\
&\text{of being a specific heterozygous genotype, given that the genotype is heterozygous.}
\end{align*}
\]

Therefore, if

\[
1 - s < 1 - \overline{h}, \text{ or } s > \overline{h} \quad (11)
\]

da positive correlation is expected for fitness and heterozygosity for this tri-allelic locus. Otherwise, a negative correlation is expected. The results can be extended easily to loci with more than three alleles to show that the conclusions are the same. Note there is no assumption concerning the equilibrium in the above derivation. It can be easily seen that the result, derived from a single locus, is applicable to multiple loci under additive and multiplicative fitness functions. It can also be easily seen that with the definition of \( s \) and \( \overline{h} \) as above, Equation 11 can be reduced to Equation 10 for the biallelic situations by just setting \( P_{CC} = 0, P_{AC} = 0, P_{BC} = 0 \) (since only two alleles A and B exist at the locus).

**DISCUSSION**

Under the simplified model of two alleles at each locus, in randomly mating populations with genetic equilibria, a positive correlation between genomic heterozygosity and fitness with dominance seems to be counterintuitive and has not been revealed before (Figures 1 and 2). A more counterintuitive conclusion is that a positive correlation could potentially exist even under within-locus additive allelic effects, without any dominance or overdominance in equilibrium populations (Figure 1, Equations 6–9). However, our analyses clearly demonstrate that these are entirely possible for a range of parameter space under Hardy-Weinberg and linkage equilibria. Importantly, negative correlations between fitness and genomic heterozygosity are actually not unexpected under a wide range of plausible parameter space of \( h \) and \( q \) in equilibrium populations (Figure 1). Additionally, in nonequilibrium populations, a positive correlation can also exist with dominance, overdominance, and additive allelic effects if Equation 10 holds at polymorphic loci; otherwise a negative correlation may exist. Furthermore, with no assumption about the genetic equilibrium, the exact genetic conditions for the positive and negative correlations for general multiallelic systems are also given. All of these results are new in that they give genetic conditions for the correlations, which directly link the genetic effect \( h \) (and \( s \)) with the population property of gene (or genotype) frequencies.

The conditions for these counterintuitive phenomena to exist do not seem to be prohibitive given some level of biological knowledge. For a positive correlation to exist, \( q \) does not have to be very common in the dominance case (Figure 1). In the additive case, in order for \( h_1 > 0.5 \) so that a positive correlation could possibly exist (Equations 6, 8 and 9), \( q \) has to be greater than 0.5 in populations at genetic equilibrium or \( P_{AA} < P_{B0} \) in populations at disequilibrium (Equation 10). Are these entirely impossible? Our knowledge is very limited on fitness effects at polymorphic loci such as those revealed by molecular markers (Kimura 1983; Nei 1987; Li 1997). For example, how is the extensive polymorphism in natural populations maintained? What is the difference in the fitness effects of different polymorphisms? How much genetic variation of fitness can polymorphism at a particular single locus explain? What is the mutation rate at a locus? Our knowledge about the population history, such as population dynamics and population admixture, is also very limited. Therefore, the possibility of \( q > 0.5 \) cannot be entirely ruled out in at least three conceivable situations. The first is in populations that have been large for long enough generations, so that a mutation-selection balance has been approximately reached. For instance, for slightly deleterious mutations (Ohta 1973, 1974), it is not unreasonable to assume \( s = 0.0001 \). Even in the most prohibitive case of additive effects, with mutation-selection balance, the locus mutation rate \( u \) is inferred to be on the order of 1.0E-5 (Crow and Kimura 1970) in order for \( q > 0.5 \). This is roughly on the order of the few mutation rates inferred for allozyme loci (Hartl and Clark 1989; Maynard-Smith 1989). For instance, Schléger and Dickie (1971) estimated that, for five loci tested in the mouse, the average mutation rate is 1.1E-5. Since this is only for visible mutants, it is likely to be a lower bound. It is noted that by inferring the order of \( u \), reversible mutations are ignored. This may be partially justified by the fact that mutations at many of its nucleotide sites within a locus may be slightly deleterious; however, in a mutant allele, only reversible mutations at those few mutated nucleotide sites can restore the original wild-type allele.
The second situation where $q$ may exceed 0.5 may be in populations experiencing recent expansion, where mutation-selection balance is not established yet but Hardy-Weinberg and linkage disequilibria are not significant. We focus on considering the conditions for different correlations under Hardy-Weinberg and linkage equilibria. In order to assume constant $q$ under constant $h$ and $s$ across loci, mutation-selection balance is assumed when we derive Equation 8 for the case of multiple loci. However, mutation-selection may not be an essential assumption for our conclusions. This can be easily seen, since the same basic conclusion (Equation 8) is derived for the single locus case (Equation 9) without assuming mutation-selection balance. It is known that Hardy-Weinberg equilibrium can be established by just one generation of random mating; linkage disequilibrium decays at a rate of $r$, which is the recombination rate between two loci at disequilibrium. However, to reach a mutation-selection balance, roughly, for those mutants with $s$ greater than $10/N$, where $N$ is the effective population size, the population has to have an annual population size in excess of $N$ for a time span (in generations) of at least a few $N$ (Kimura et al. 1963; Lynch et al. 1995; Deng and Lynch 1996b). Therefore, it takes a much longer time for a population to reach mutation-selection balance than to reach approximate genetic equilibria. So, it is possible for a population to be approximately in genetic equilibria without reaching mutation-selection balance. The third scenario is that for non-equilibrium populations due to recent population admixture, we can have $P_{aa} < P_{ab}$ fairly easily if two populations mix and the larger proportion is from the population homozygous for $a$.

The observed negative correlations between fitness traits and heterozygosity in a number of studies (Gains et al. 1978; Pierce and Mitton 1982; Mitton and Grant 1984; Allendorf and Leary 1986; Zouros and Foltze 1987; Lynch and Walsh 1997) have not attracted much attention. Hence, there has been hardly any satisfactory genetic explanation for them. Researchers have generally attributed these negative correlations to statistical artifacts and commented that with a parametric slope of nearly zero, negative correlations can result by chance. However, we show here that the negative correlations may indeed result if, at each locus, heterozygote fitness is smaller than the weighted mean fitness of homozygotes, or $h > h_c$. As noted in Figure 1 and Equations 6-10, for a di-allelic system, the negative correlations can actually occur under a wide range of parameter space of $h$ and $q$ in equilibrium and nonequilibrium populations. This conclusion contradicts that of Turelli and Ginzburg (1983) for multi-allelic systems. By numerical analyses via computer simulations, they concluded that in populations with little linkage disequilibrium, average fitness always increases with genomic heterozygosity. The different conclusions of ours and that of Turelli and Ginzburg (1983) are not due to the genetic systems under study (bi-allelic vs. multi-allelic), since we clearly show that the negative correlations can result under multi-allelic systems (Equation 11). Turelli and Ginzburg (1983) did not study the influence of the detailed genetic effect $h$ and $s$ in association with gene (or genotype) frequencies on the correlations.

The potentially common negative correlations are not inconsistent with the widely observed inbreeding depression. For a single locus with two alleles, the necessary and sufficient condition for population inbreeding depression to occur is that the heterozygote has fitness greater than the arithmetic mean fitness of the two corresponding homozygotes, not weighted by their population frequencies as in Equations 6 and 10 (Crow and Kimura 1970; Falconer and Mackay 1996). Therefore, the widely observed inbreeding depression does necessarily imply that the negative correlation should be rare. In fact, when $q$ is small, the negative correlation should be more likely unless $h$ is also very small (Figure 1). Please note, underdominance (where $h > 1$) may be an alternative and sufficient explanation for the negative correlations. However, it is not a necessary one, since $h$ does not have to be greater than 1.0 if $h_c < 1$. As long as $h > h_c$, a negative correlation will exist. Figure 1 depicts the parameter space where inbreeding depression and negative correlations can co-occur in a single population. Because of this potential co-occurrence of inbreeding depression (heterosis) and negative correlations, the positive correlations may not be interpreted as equivalent to heterosis.

It should be pointed out that although the derivation for multiple-loci in the theory section is based on the assumptions of constant effects ($h$ and $s$) and multiplicative fitness function, these assumptions are not essential for our main conclusions. In the theory section, for the one locus model, we also showed the same basic results as the multilocus model. If $h$, and $s$, are variable across loci, each locus will have its own peculiar $q$, and thus a peculiar $h_c$, at mutation-selection balance. At the $i$th locus, $h_c$ is determined by the equilibrium $q_i$ (Equation 7), which in turn depends on the specific allelic effects $h_i$ and $s_i$ for equilibrium populations at mutation-selection balance (Crow and Kimura 1970). Under either multiplicative or additive fitness functions across loci, as long as at each locus, $h_i < h_c$, being a heterozygote has higher expected fitness than being a homozygote; thus, a positive correlation between fitness and genomic heterozygosity will result. Similar conclusions hold for the situations of negative correlations. In cases where $h_i < h_c$ at some loci and $h_i > h_c$ at other loci, no correlation or either correlation could exist, which depends on the fitness effects and the number of loci with $h_i < h_c$ relative to those with $h_i > h_c$.

The explanation for the commonly observed positive correlations between fitness (or its related traits such as developmental stability) and molecular marker loci may be complex. In populations that are not strictly
panmictic with local inbreeding present, the genomic heterozygosity may be correlated with individual levels of inbreeding. Therefore, the associations between heterozygosity and fitness are largely the consequence of variation in the level of inbreeding among individuals (Ledig et al. 1983; Strauss 1986). In randomly mating populations, if linkage disequilibrium is present (whether due to random genetic drift, selection or other causes), linked deleterious alleles under dominance at different loci will likely cause associated overdominance (Houle 1989, 1994). This also may explain the positive correlations. However, we have particularly shown, via analytical approaches supplemented by simulations, two novel results for populations at genetic equilibria. The first is that the positive correlation can result with dominance. The second is that the positive correlation may not necessarily always reflect the phenomenon of heterosis, for the two reasons argued earlier: (1) the positive correlation may potentially exist even under within-locus additive allelic effects, and (2) the possible co-occurrence of inbreeding depression (heterosis) and negative correlations implies that the positive correlations may not be interpreted as equivalent to heterosis. In populations at genetic equilibria, functional overdominance may be an alternative and sufficient explanation (Smouse 1986); however, it is not a necessary one. Therefore, even for populations at equilibria, the positive correlation observed cannot by itself be evidence for overdominance; it may not even be evidence for heterosis expected with dominance/over-dominance either, since it could potentially exist under pure additive allelic effects.

An implication of our results is that there is probably a limitation of the correlation approach for distinguishing the genetic mechanisms responsible for the maintenance of genetic variability. This is because of the following reasons. First, in equilibrium populations, a positive correlation can be explained by both dominance ($0 < h < 1.0$) and overdominance ($h < 0$) as long as $h < h_c$ (Equations 6–10). Second, the negative correlations and heterosis can co-occur under the same genetic conditions (Figure 1). Third, even for the same genetic effect (i.e., the same $h$) in one species, both correlations could be revealed in different populations if these populations have different genotype frequencies due to different population origins and histories (Equation 10). The limitation of the correlation approach in inferring the mechanisms responsible for the maintenance of genetic variability was also pointed out before on different grounds (e.g., Houle 1994; Fu and Ritland 1996).

A potential application of the theoretical result here is that, for diallelic makers (such as those from RFLP), inference of the upper/ lower bounds of $h$ may be made given significant positive/negative correlations being found (Equations 6–10). $h$ is an important genetic parameter in population and evolutionary genetics and has been difficult to estimate (even for its bounds, Deng 1997; Deng et al. 1997), especially for those organisms for which controlled breeding is difficult. Whereas, using Equations 6–10, the bounds of $h$ may be estimated without controlled breeding with the application of the traditional correlation approach for fitness and genomic heterozygosity in natural populations.

We thank Drs. D. Charlesworth, M. Lynch, R. Chakraborty, and M. Johnson for discussions and comments. We are also grateful to Dr. Z.-B. Zeng and three anonymous reviewers for their helpful comments that improved the paper. H.-W. Deng would like to thank Dr. M. Lynch for years of advice, and Dr. D. Hedgecock for providing support to attend the conference "Genetic and Physiological Basis of Heterosis," which stimulated the development of this work. The work was in part supported by a FIRST AWARD from National Institutes of Health to Dr. Y.-X. Fu. H.-W. Deng was supported by a grant from Health Future Foundation to Drs. R. Recker and D. Kimmel while finalizing this paper.

**Literature Cited**


Communicating editor: Z-B. Zeng