

# INTERACTION OF GENETIC LOCI: THE EFFECT OF LINKAGE DISEQUILIBRIA ON HARDY-WEINBERG EXPECTATIONS

JEFFREY R. POWELL

*Department of Biology, Yale University, New Haven, Conn. 06520*

Received 6.iii.73

## SUMMARY

Interactions of a neutral polymorphic locus with a heterotic locus are investigated. In particular, deviations from Hardy-Weinberg expectations at the neutral locus are investigated. The amount of deviation is shown to be determined by the amount of linkage disequilibrium between alleles at the loci and the strength of heterosis at the selected locus. Under some conditions apparent selection coefficients can be determined for the genotypes at the neutral locus. Actually observed deviations from *H-W* for enzyme loci are discussed. From the results of the present investigation, it seems unlikely that the heterozygote excesses observed at these loci can be accounted for by associative overdominance alone.

## 1. INTRODUCTION

HETEROZYGOTE excess is an indication of overdominance at a locus. However, if the locus being observed is not overdominant but alleles at the locus are in linkage disequilibrium with an overdominant locus, then the observed locus will appear overdominant. Such a phenomenon has been termed pseudo-overdominance (Lerner, 1958) or associative overdominance (Frydenberg, 1963). The purpose of this report is to quantify the effect of associative overdominance on Hardy-Weinberg equilibrium at an unselected locus. We ask the question: How much deviation from Hardy-Weinberg can be expected at a "neutral" locus for a given amount of linkage disequilibrium with a heterotic locus? In certain cases, apparent selection coefficients can be calculated for genotypes at the neutral locus. The results of this indicate that, in general, associative overdominance can account for only slight deviations from *H-W* at the neutral locus. This treatment is then used to analyse some of the data on protein polymorphisms. The conclusion is reached that associative overdominance alone probably cannot account for the observed heterozygote excess at several protein loci.

## 2. MODEL

A random mating diploid population is assumed. The population is assumed to be at equilibrium with respect to gene frequencies. Furthermore, we assume that a certain amount of linkage disequilibrium is exhibited by the loci at the time of sampling. The actual cause of the disequilibrium does not directly concern us. Consider two loci with two alleles each: *A*, *a* and *B*, *b*.

The "A" locus is assumed to be "neutral", that is, all three genotypes have the same fitness. The "B" locus is heterotic. We define

$$\begin{aligned} p &= \text{the frequency with which the } A \text{ allele is found with } B \\ 1-p &= \text{the frequency with which the } a \text{ allele is found with } B \\ q &= \text{the frequency with which the } A \text{ allele is found with } b \\ 1-q &= \text{the frequency with which the } a \text{ allele is found with } b. \end{aligned}$$

Thus the difference between  $p$  and  $q$  is a measure of the disequilibrium between the two loci. If  $p$  and  $q$  are equal, then there is no linkage disequilibrium. The following fitnesses are assigned to "B" genotypes:  $B/B, 1-s$ ;  $B/b, 1$ ;  $b/b, 1-t$ . Thus the following equilibrium frequencies are observed:

$$\begin{aligned} f_B &= \frac{t}{s+t} \\ f_b &= \frac{s}{s+t} \\ f_A &= p \left( \frac{t}{s+t} \right) + q \left( \frac{s}{s+t} \right) \\ f_a &= (1-p) \left( \frac{t}{s+t} \right) + (1-q) \left( \frac{s}{s+t} \right). \end{aligned}$$

Fig. 1 summarises the model and also gives the expected frequency of each genotype before selection.

GENOTYPE		GENOTYPE		
		B / B	B / b	b / b
FREQUENCY BEFORE SELECTION		$\left( \frac{t}{s+t} \right)^2$	$2 \left( \frac{t}{s+t} \right) \left( \frac{s}{s+t} \right)$	$\left( \frac{s}{s+t} \right)^2$
FITNESS		$1-s$	$1$	$1-t$
A / A	$1$	$p^2$	$pq$	$q^2$
A / a	$1$	$2p(1-p)$	$(1-p)q + (1-q)p$ or $p+q-2pq$	$2q(1-q)$
a / a	$1$	$(1-p)^2$	$(1-p)(1-q)$	$(1-q)^2$

FIG. 1.—Summary of model. "A" is a selectively neutral locus, *i.e.* the fitnesses of the three genotypes are equivalent. "B" is a heterotic locus with selection coefficients  $s$  and  $t$ .  $p$  and  $q$  are measures of the linkage disequilibrium between the two loci. The cells show the expected frequency of each combination of genotypes at equilibrium before selection.

As a measure of deviation from Hardy-Weinberg expectations, we define:

$$H = \frac{\text{observed frequency of heterozygotes}}{\text{expected frequency of heterozygotes}}$$

In general, if the selection is acting at the locus being observed,  $H = \frac{1}{\bar{w}}$  which is what  $H$  is for locus "B" in the above example.

### 3. RESULTS

After selection at the "B" locus has occurred, genotype frequencies at locus "A" will be:

$$f(A/A) = \left[ p^2(1-s) \left( \frac{t}{s+t} \right)^2 + 2pq \left( \frac{t}{s+t} \right) \left( \frac{s}{s+t} \right) + q^2 \left( \frac{s}{s+t} \right)^2 \right] / \bar{w} \quad (1)$$

$$f(A/a) = \left[ 2p(1-p)(1-s) \left( \frac{t}{s+t} \right)^2 + 2(p+q-2pq) \left( \frac{t}{s+t} \right) \left( \frac{s}{s+t} \right) + 2q(1-q)(1-t) \left( \frac{s}{s+t} \right)^2 \right] / \bar{w} \quad (2)$$

$$f(a/a) = \left[ (1-p)^2(1-s) \left( \frac{t}{s+t} \right)^2 + 2(1-p)(1-q) \left( \frac{t}{s+t} \right) \left( \frac{s}{s+t} \right) + (1-q)(1-t) \left( \frac{s}{s+t} \right)^2 \right] / \bar{w} \quad (3)$$

where

$$\bar{w} = 1 - s \left( \frac{t}{s+t} \right)^2 - t \left( \frac{s}{s+t} \right)^2. \quad (4)$$

The expected frequency of  $A/a$  is:

$$2f_A f_a = 2 \left[ (1-p) \left( \frac{t}{s+t} \right) + (1-q) \left( \frac{s}{s+t} \right) \right] \left[ p \left( \frac{t}{s+t} \right) + q \left( \frac{s}{s+t} \right) \right] \quad (5)$$

From equations (2), (4) and (5) we may now determine the heterozygote excess for the neutral locus. For locus "A" with the parameters defined above:

$$H_A = \frac{2p(1-p)(1-s) \left( \frac{t}{s+t} \right)^2 + 2(p+q-2pq) \left( \frac{t}{s+t} \right) \left( \frac{s}{s+t} \right) + 2q(1-q)(1-t) \left( \frac{s}{s+t} \right)^2}{2 \left[ (1-p) \left( \frac{t}{s+t} \right) + (1-q) \left( \frac{s}{s+t} \right) \right] \left[ p \left( \frac{t}{s+t} \right) + q \left( \frac{s}{s+t} \right) \right]} \cdot \frac{\left[ 1 - s \left( \frac{t}{s+t} \right)^2 - t \left( \frac{s}{s+t} \right)^2 \right]}{\left[ 1 - \frac{st[tp(1-p) + sq(1-q)]}{[t(1-p) + s(1-q)][tp + sq]} \right]} \quad (6)$$

If we make two simplifying assumptions, heterosis is symmetrical (*i.e.*  $s = t$ ) and the disequilibrium is symmetrical (*i.e.*  $p = 1 - q$ ), then equation (6) takes on the relatively simple form

$$H_A = \frac{1 - 2pqs}{1 - (1/2)s} = (1 - 2pqs) \left( \frac{1}{\bar{w}} \right). \quad (7)$$

Since  $H$  for the selected locus ("B") is  $\frac{1}{\bar{w}}$ , then for this special case,  $H$  for the neutral locus is diminished by a factor  $2pqs$ .

For this simplified case of symmetrical heterosis and disequilibrium we may now ask what will be the apparent selection coefficient at the neutral locus due to this associative overdominance. In this case,  $\bar{w} = 1 - (1/2)s$  and  $H = \frac{1}{1 - (1/2)s}$ . We now define  $s'$  as the apparent selection coefficient of the neutral locus. From (7) above, we then get

$$H_A = \frac{1 - 2pqs}{1 - (1/2)s} = \frac{1}{1 - (1/2)s'} \quad (8)$$

$$s' = 2 - \frac{2 - s}{1 - 2pqs}. \quad (9)$$

(It should be noted that unless  $s = t$ ,  $H$  alone does not allow one to calculate values of  $s$  and  $t$ .)

Table 1 gives various values of  $s$ ,  $t$ ,  $p$ ,  $q$ ,  $H_A$  and  $s'$ . Figs. 2 and 3 give graphical representations of these data. In fig. 2 it is apparent that for a given amount of disequilibrium, as the heterosis of the interacting locus

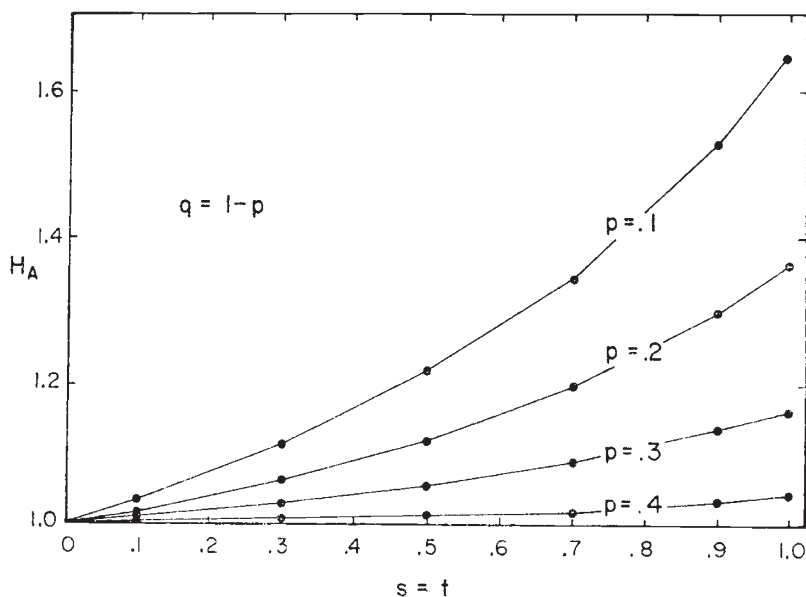


FIG. 2.—Graph showing relationship between amounts of linkage disequilibrium and heterosis, and the heterozygote excess observed at a selectively neutral locus. Symbols same as table 1. This graph represents the case of symmetrical heterosis and linkage disequilibrium.

increases, the  $H$  value of the "neutral" locus increases. Also in this graph it can be seen that for a given amount of heterosis, as the disequilibrium increases, so does  $H_A$ . Fig. 3 gives the relationship between  $s$  and  $s'$ . As

TABLE 1

*The heterozygote excess over H-W expectations at a selectively "neutral" locus ( $H_A$ ) when in linkage disequilibrium with a heterotic locus.  $p$  and  $q$  are measures of the disequilibrium and  $s$  and  $t$  determine the strength of the heterosis.  $s'$  is the apparent selection coefficient for the neutral locus in the cases of symmetrical heterosis and linkage disequilibrium. See text for details*

$p$	$q$	$s$	$t$	$H_A$	$s'$
0.1	0.9	0.1	0.1	1.034	0.065
		0.3	0.3	1.113	0.203
		0.5	0.5	1.213	0.352
		0.7	0.7	1.345	0.513
		0.9	0.9	1.524	0.687
		1.0	1.0	1.640	0.780
0.2	0.8	0.1	0.1	1.019	0.037
		0.3	0.3	1.064	0.120
		0.5	0.5	1.120	0.214
		0.7	0.7	1.194	0.325
		0.9	0.9	1.295	0.456
		1.0	1.0	1.360	0.529
0.3	0.7	0.1	0.1	1.008	0.016
		0.3	0.3	1.028	0.054
		0.5	0.5	1.053	0.101
		0.7	0.7	1.086	0.158
		0.9	0.9	1.131	0.232
		1.0	1.0	1.160	0.276
0.4	0.6	0.1	0.1	1.002	0.004
		0.3	0.3	1.007	0.014
		0.5	0.5	1.003	0.026
		0.7	0.7	1.022	0.043
		0.9	0.9	1.003	0.064
		1.0	1.0	1.040	0.077
0.2	0.8	0.2	0.1	1.024	—
		0.2	0.3	1.048	—
		0.2	0.5	1.052	—
		0.2	0.7	1.051	—
		0.2	0.9	1.049	—
		0.2	1.0	1.048	—
0.1	0.2	0.5	0.5	1.007	—
		0.4	0.5	1.027	—
		0.6	0.5	1.092	—
		0.8	0.5	1.165	—
		0.9	0.5	1.213	—

expected, as  $s$  increases, so does  $s'$ ; for a given  $s$ , as the disequilibrium increases,  $s'$  increases.

It is appropriate here to point out that the parameters  $p$  and  $q$  defined here are somewhat different from the usual parameters used in disequilibrium studies (*e.g.* Lewontin and Kojima, 1960). However, it is easy to translate between the two sets of terms. In fact, it can be shown that the difference

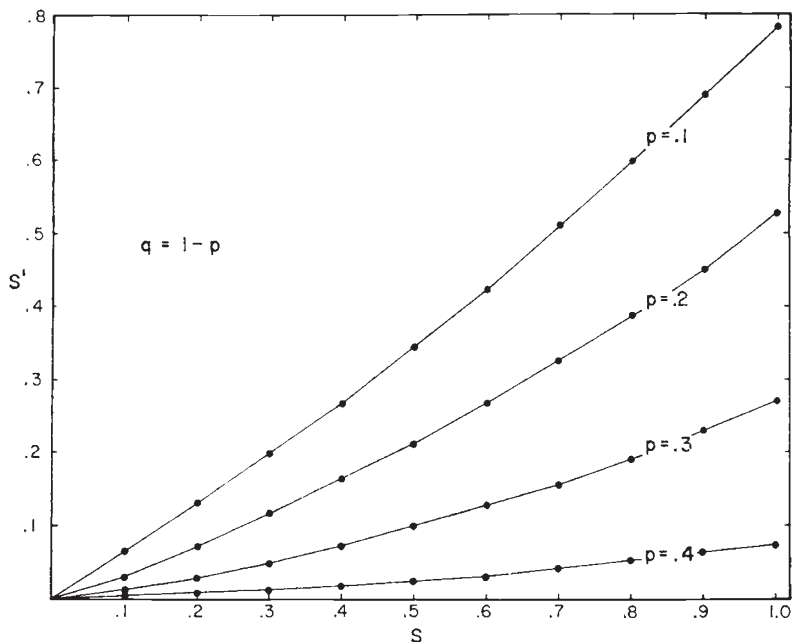


FIG. 3.—Graph showing relationship between the selection coefficients ( $s$ ) of the heterotic locus and the apparent coefficients ( $s'$ ) at the neutral locus.  $p$  and  $q$  indicate the amount of linkage disequilibrium between the two loci.

between  $p$  and  $q$  is a measure of linkage disequilibrium as defined by  $D$ . Using the terms defined above and  $D$  defined by Lewontin and Kojima:

$$D = \frac{st(p-q)}{(s+t)^2}; \quad (10)$$

for  $s = t$

$$D = \frac{p-q}{4}. \quad (11)$$

#### 4. DISCUSSION

It seems safe to conclude that to observe an excess of heterozygotes at a neutral locus it must be in considerable linkage disequilibrium with a strongly heterotic locus. For values of  $p$ ,  $q$ ,  $s$  and  $t$  which one might consider reasonable,  $H_A$  is usually less than 1.1. For example, with  $p$  and  $q$  equal to 0.2 and 0.8 respectively and  $s = t = 0.3$ ,  $H_A$  is 1.064. This would be a case of considerable linkage disequilibrium and moderately strong heterosis, yet the heterozygote excess at the neutral locus is relatively slight. A sample size around 3000 would be necessary to detect a significant (90 per cent sure) excess of heterozygotes at the neutral locus (Lewontin and Cockerham, 1959).

Cases of heterozygote excesses have been reported for enzyme loci. For *Drosophila pseudoobscura*, Wills and Nichols (1971) obtained an  $H$  value of 1.25 at the *Odh* locus in the males which had been inbred and stressed by adding octanol to the medium. Richmond and Powell (1970) report an  $H$  value of

1.30 for the *To* locus of *D. paulistorum* from a natural population and Powell (1974) has found that the same locus gives an *H* of 1.12 in laboratory populations. From table 1 it can be seen that to obtain such values of *H* due to associative overdominance, the disequilibrium must be great and the heterosis very strong. It seems more likely that in these cases the observed heterozygote excess was at least partly due to the locus being studied.

A special case of associative overdominance is the inclusion of a locus in a heterotic chromosomal polymorphism. The present treatment can easily be extended to handle this situation, one need only substitute gene arrangements for the alleles at the selected "B" locus described above. Because of the great numbers of inversion polymorphisms in *D. paulistorum*, Powell (1974) tested for linkage disequilibrium between alleles at an apparently heterotic protein locus and gene arrangements in the chromosome on to which it maps. He found that for gene arrangements in the right arm of the X chromosome and alleles at the *To* locus the following values were obtained:

$$p = 0.52 \pm 0.05 = \text{frequency of } F \text{ allele with ST(XR),}$$

$$q = 0.47 \pm 0.05 = \text{frequency of } F \text{ allele with INV(XR).}$$

For the left arm of the X chromosome he obtained:

$$p = 0.48 \pm 0.04 = \text{frequency of } F \text{ allele with ST(XL),}$$

$$q = 0.56 \pm 0.075 = \text{frequency of } F \text{ allele with INV(XL).}$$

Even assuming errors of 2 S.E. for the *p* and *q* values and the greatest possible values of *s* and *t* (1.0), an *H* value of less than 1.10 is obtained. Since *H* values of 1.30 and 1.12 were observed at this locus in natural and experimental populations respectively, it is very unlikely that association of the alleles with gene arrangements can account for the observations.

Finally it should be stressed that the treatment here does not attempt to explain the cause of the assumed linkage disequilibrium. Only a few of the effects of certain kinds of disequilibria are investigated. Ohta and Kimura (1970) have shown that genetic drift in finite populations can be a cause of linkage disequilibrium leading to associative overdominance. (In their paper they derived an approximation—expression (4)—analogous to the deterministic equation (9) above.) However, Lewontin and Kojima (1960) have shown that in order to obtain permanent linkage disequilibria, there must be strong fitness interactions among the loci.

*Acknowledgments.*—I thank Professors Th. Dobzhansky and Wyatt W. Anderson for advice on this project and Dr Martin L. Tracey for fruitful discussion of the manuscript. I also thank the anonymous reviewer who found the simplified form of equation (6).

## 5. REFERENCES

- FRYDENBERG, O. 1963. Population studies of a lethal mutant in *Drosophila melanogaster*. I. Behavior in populations with discrete generations. *Hereditas*, 50, 89-119.
- LENER, I. M. 1958. *The Genetic Basis of Selection*. Wiley, New York.
- LEWONTIN, R. C., AND COCKERHAM, C. C. 1959. The goodness-of-fit test for detecting natural selection in random mating populations. *Evolution*, 13, 561-564.
- LEWONTIN, R. C., AND KOJIMA, K. 1960. The evolutionary dynamics of complex polymorphisms. *Evolution*, 14, 458-472.

- OHTA, T., AND KIMURA, M. 1970. Development of associative overdominance through linkage disequilibrium in finite populations. *Genet. Res. Camb.*, 16, 165-177.
- POWELL, J. R. 1974. Heterosis at an enzyme locus of *Drosophila*: Evidence from experimental populations. *Heredity*, 32, 105-108.
- RICHMOND, R. C., AND POWELL, J. R. 1970. Evidence of heterosis associated with an enzyme locus in a natural population of *Drosophila*. *Proc. Nat. Acad. Sci. USA*, 67, 1264-1267.
- WILLS, C., AND NICHOLS, L. 1971. Single-gene heterosis in *Drosophila* revealed by inbreeding. *Nature*, 233, 123-125.