HETEROSIS OF THE HYBRID RELATED TO GENE FREQUENCY DIFFERENCES BETWEEN TWO POPULATIONS[1]

C. E. CRESS[2]

Statistical Laboratory, Iowa State University, Ames

Received September 20, 1965

GENETIC diversity of the parents has long been recognized as a requirement for the expression of heterosis to be manifest in the cross of two lines or varieties. Experimental evidence in Drosophila (BRNCIC 1954; WALLACE 1955) has demonstrated that crosses of strains of diverse origin exhibit greater heterotic response than similar crosses of strains from the same origin. The foundation of hybrid corn production is based on the principle of crosses between unrelated inbred lines. In more recent years, interest in the introduction of exotic corn stocks to increase the genetic variability has developed. The initial studies attempted to relate diversity of origin to heterotic response obtained. WELLHAUSEN (1952) recorded the responses of F_1's for a large number of diverse Mexican races. The heterotic responses ranged from large positive effects down to performance below either parent variety. LONNQUIST and GARDNER (1961) found that intervarietal crosses among 12 parents representing a range of Cornbelt germ plasm produced mean F_1 yields above the midparent. Of the 66 individual F_1 yields, five were below the midparent value. PATERNIANI and LONNQUIST (1963) obtained similar results from 63 F_1 crosses among South American races of maize. MOLL, SALHUANA and ROBINSON (1962) examined the performance of 15 crosses of six varieties representing three widely dispersed geographical regions. In each of these studies there was a positive correlation between genetic diversity, geographical or plant type, and the degree of heterotic response. At the same time, each study revealed individual crosses that have not performed as well as expected when judged by the parental yields.

This study reports an examination of the role of gene frequency differences between two populations as they relate to the heterotic response of the hybrid.

Notation

Heterosis will be used here in the sense of the difference between population hybrid and average of the parents (midparent). Random mating within the parents, designated A and B, and random selection of parent plants for the hybrid are assumed. The expression developed will be for an arbitrary number of alleles at a locus.

[1] This research was supported by Grants 18093 and 19218 from the National Science Foundation.

[2] Present address: Statistics Center, Rutgers—The State University, New Brunswick, New Jersey.

The mean of population A is

$$\mu_A = \sum_{i,k=1}^{r} ({}_a p_i)({}_a p_k) g_{ik}$$

where ${}_a p_i$ is the gene frequency of the *ith* allele in population A, and g_{ik} is the genotypic value of an individual with allelic constitution $G_i G_k$. Similarly the mean of B is

$$\mu_B = \sum_{m,n=1}^{r} ({}_b p_m)({}_b p_n) g_{mn}$$

With a random choice of parents, the population hybrid has a mean

$$\mu_{AB} = \sum_{i,n=1}^{r} ({}_a p_i)({}_b p_n) g_{in}$$

Then, heterosis as measured from the midparent is given by

$$\begin{aligned} M_H &= \mu_{AB} - \tfrac{1}{2}(\mu_A + \mu_B) \\ &= \sum_{i,n} ({}_a p_i)({}_b p_n) g_{in} - \tfrac{1}{2} \sum_{i,k} ({}_a p_i)({}_a p_k) g_{ik} - \tfrac{1}{2} \sum_{m,n} ({}_b p_m)({}_b p_n) g_{mn} \end{aligned} \qquad (1)$$

If reciprocal crosses are assumed equal $g_{in} = g_{ni}$, then (1) simplifies to

$$M_H = \tfrac{1}{2} \sum_{i,n} [({}_a p_i - {}_b p_i)({}_b p_n - {}_a p_n)] g_{in}$$

This form of the expression shows that heterosis is a function of the differences in the frequencies of the alleles in the two populations and the genotypic values. The expression is written for a single locus, but may be summed over loci if no interlocus epistasis and independent segregation are assumed. In matrix notation

$$M_H = \tfrac{1}{2} P' G(-P) = \tfrac{1}{2} P'(-G) P,$$

where G is the $r \times r$ matrix of genotypic values for the r alleles at a single locus, and P is the $r \times 1$ vector of differences of gene frequencies between the two populations. For convenience, the homozygous genotype with the smallest genotypic value can be coded to have a value of zero.

Properties of M_H

In order to examine and categorize M_H, it was found necessary to place a restriction on the genotypic values. The restriction is such that any heterozygous genotypic value ($g_{in}, i \neq n$) will be greater than or equal to the average genotypic value of the homozygotes, that is

$$g_{in} = g_{ni} \geq (g_{ii} + g_{nn})/2, \qquad i \neq n$$

Because of this restriction, all g_{in} are non-negative. The restriction excludes what could be termed negative dominance, and seems to be in accord with what is generally observed for most economically important traits.

A sufficient, but not necessary, condition for the quadratic form $P'(-G)$P to be positive is that $(-G)$ be positive definite. The nature of assigning genotypic values in the G matrix ensures that the elements on the principle diagonal of $(-G)$ will be negative or zero. Thus, none of the matrices $(-G)$ can be positive definite. For an arbitrary vector P there are no general conditions on the matrix $(-G)$ that will assure a positive value for M_H. There is additional information concern-

ing the vector P that will be useful in determining the properties of M_H. Let f_i be the elements of the vector P. Then

$$\text{(a)} \quad \sum f_i = 0$$
$$\text{(b)} \quad \text{Maximum} \sum |f_i| = 2.0$$

As an initial step, the quadratic form $P'(-G)P$ will be examined for a few alleles per locus and then generalized to an arbitrary number of alleles per locus.

With two alleles per locus

$$M_H = \tfrac{1}{2}P'(-G)P = \tfrac{1}{2}(r\ s)\begin{pmatrix} g & a \\ a & 0 \end{pmatrix}\begin{pmatrix} r \\ s \end{pmatrix}$$

where $r + s = 0$ and elements g and a are negative. Since $s = -r$,

$$M_H = \tfrac{1}{2}r^2(g - 2a).$$

With a positive degree of dominance $2|a| \geqq |g|$. Therefore, M_H will be non-negative. With two alleles, M_H is zero only when $2g = a$ (additive gene action) or $r = s = 0$ (no difference in gene frequency between populations A and B).

With three alleles per locus

$$M_H = \tfrac{1}{2}(r\ s\ t)\begin{pmatrix} g & a & b \\ a & h & c \\ b & c & 0 \end{pmatrix}\begin{pmatrix} r \\ s \\ t \end{pmatrix}$$

$$= \tfrac{1}{2}(gr^2 + 2ars + 2\,brt + hs^2 + 2,cst).$$

With four alleles per locus

$$M_H = \tfrac{1}{2}(r\ s\ t\ u)\begin{pmatrix} g & a & b & c \\ a & h & d & e \\ b & d & i & f \\ c & e & f & 0 \end{pmatrix}\begin{pmatrix} r \\ s \\ t \\ u \end{pmatrix}$$

$$= \tfrac{1}{2}(gr^2 + 2ars + 2brt + 2cru + hs^2 + 2dst + 2esu + it^2 + 2ftu). \qquad (2)$$

The pattern of expansion is readily apparent and extension to any number of alleles can easily be done. The four allele case will be used for discussion with the understanding that the same argument can be used for any number of alleles greater than two. The cross product terms in (2) can be of either sign, and little can be determined about the sign of M_H by inspection. Certain special cases will be informative for investigating the nature of the general equation 2.

With multiple alleles, the definition of additive, dominant, and overdominant gene effects will be such that for any *pair* of alleles the customary interpretation of gene effects will be used. For example, a locus with all pairs of alleles exhibiting completely dominant gene effects would have a $(-G)$ matrix of the form

$$\begin{pmatrix} g & g & g & g \\ g & h & h & h \\ g & h & i & i \\ g & h & i & 0 \end{pmatrix}$$

where $|g|>|h|>|i|>0$. In equation 2, this means that $g=a=b=c$, $h=d=e$ and $i=f$. Making this substitution one gets

$$M_H = \tfrac{1}{2}(gr^2 + 2grs + 2grt + 2gru + hs^2 + 2hst + 2hsu + it^2 + 2itu).$$

Since $r+s+t+u=0$ then,

$$M_H = \tfrac{1}{2}[(-g+h)(s+t+u)^2 + (-h+i)(t+u)^2 + (-i)u^2].$$

But $(-g+h)>0$, $(-h+i)>0$, $(-i)>0$; therefore, $M_H>0$ except for the trivial case $g=h=i=0$. Under the conditions set up for the completely dominant model, M_H will always be positive. However, this is a very special case.

The case of pure overdominance can be illustrated by setting $g=h=i=0$. Then, $(-G)$ is of the form

$$\begin{pmatrix} 0 & a & b & c \\ a & 0 & d & e \\ b & d & 0 & f \\ c & e & f & 0 \end{pmatrix}$$

Equation 2 becomes

$$M_H = ars + brt + cru + dst + esu + ftu. \qquad (3)$$

By inspection of (3), one can see that any term may predominate and be of either sign, thus M_H can be positive or negative. Only a few numerical examples are required in order to be convinced that a negative M_H, with an underlying scale of pure overdominance, is neither unlikely nor aberrant. Let the negative of the matrix of genotypic values be arbitrarily chosen to be

$$(-G) = \begin{pmatrix} 0 & -3 & -10 & -1 \\ -3 & 0 & -2 & -3 \\ -10 & -2 & 0 & -2 \\ -1 & -3 & -2 & 0 \end{pmatrix}$$

P vectors which give negative measures of heterosis are quite easy to construct. The three examples in Table 1 were chosen to represent realistic crosses where multiple alleles probably are found.

Example 1 is intended to simulate a typical locus in the cross of inbred line B to an unrelated source. Example 2 could be the F_2 of a single cross mated to an unrelated source. The gene frequencies shown in Example 3 would be equivalent

TABLE 1

Gene frequencies and heterosis values for three representative crosses involving four alleles with an underlying scale of pure overdominance

Example 1		Example 2		Example 3	
a^pi	b^pi	a^pi	b^pi	a^pi	b^pi
.4	0	.5	0	.5	0
.1	0	0	.2	0	.5
.5	0	.5	0	.5	0
0	1.0	0	.8	0	.5
$M_H=-0.52$		$M_H=-1.28$		$M_H=-1.25$	

to a double cross of two unrelated single cross F_2's. Other permutations of the gene frequencies in these examples will give both positive and negative values for M_H. The significant observation is that negative values exist at all. Breeding theory and practice have tacitly assumed that overdominant gene effects combined with a cross of unrelated sources produce the greatest heterotic effect. It has been shown here that quite the opposite is possible, and likely at certain loci.

Thus far, we have considered only the type of overdominance where all the homozygotes are equal, and they in turn are less than any of the heterozygotes. Suppose the restriction on the homozygotes is relaxed to allow them to be nonzero. Upon examination of the general equation 2 it is evident that the only terms that were not considered in the previous overdominance model are $\frac{1}{2}(gr^2 + hs^2 + it^2)$. Since g, h, and i are all negative, we find that all homozygotes contribute *negatively* to M_H. Thus, M_H is more likely to be negative than the purely overdominant model with the same off-diagonal elements of the G matrix. When the condition prohibiting negative dominance is relaxed, then additional negative impetus is given to M_H.

Of the models considered, only the very special conditions for the completely dominant model gives assurance that M_H will be positive for all vectors P. In general, negative M_H are possible. This is true for partial dominance, over dominance, or combinations of gene effects. When multiple alleles are considered, the largest class of gene effects would fall into the category of a combination of intralocus gene effects. A relevant point to be concluded from this development is that when more than two alleles per locus are considered, one cannot expect each locus to contribute positively to the heterotic effect, even when the alleles in pairs have partially dominant, completely dominant, or over dominant effects. When all loci in a genotype are considered, the positive contributions will generally be larger in number and in magnitude than the negative contributions. This has been verified in many studies. Some were cited. However, the breeder should not be surprised to find certain crosses that yield less than the midparent value, even with a positive degree of dominance for all pairs of alleles.

APPLICATION

The amount of heterosis exhibited by a cross is customarily used as a measure of the genetic divergence of the parent stock. Without question, genetic divergence (difference in gene frequency) in the parents is required for heterosis to be manifest in the cross. As has been shown, the converse of this statement is not true; that is, the lack of heterotic response cannot be used to infer a lack of genetic divergence. The negative heterotic contributions at certain loci cancel positive responses at other loci. The net response in the hybrid may be little or no deviation from the midparent. Thus, the validity of evaluating the degree of genetic divergence based on the amount of heterotic response is subject to considerable question.

The choice of a tester for a heterogeneous population is often conditional on the average performance of the test crosses; that is, the tester with the highest

average cross performance is chosen. Unless the selected individuals are to be used immediately in hybrid combination with the tester this emphasis on heterotic response is misplaced. The heterotic response reveals little concerning the genetic potential and nothing concerning the expected rate of progress from selection.

In evaluating hybrids, breeders can point to numerous examples of unexpectedly poor hybrid performance in spite of superior parental stock and genetic diversity. Linkage and epistasis have been proposed as the cause of this poor performance. However, the effects of multiple alleles could account for this variable performance even in the absence of interlocus epistasis and linkage. If the percentage of loci contributing negatively to M_H is small, then a high performing hybrid results. Thus, it can be seen that genetic diversity is necessary for significant heterosis but not sufficient to guarantee it.

SUMMARY

Genetic diversity of two populations as measured by gene frequency differences, was related to the heterotic performance of the hybrid, where the measure of heterosis was deviation from the midparent. Heterosis (M_H) was found to be non-negative in the case of two alleles, if negative dominance was excluded. With more than two alleles per locus, negative contributions to M_H are to be expected at certain loci, and the net effect may result in a hybrid genotypic value equal to or below the midparent. This can occur even when the dominance relationships of all pairs of alleles are positive. Some implications of these results are given.

LITERATURE CITED

BRNCIC, D., 1954 Heterosis and the integration of the genotype in geographic populations of *Drosophila pseudoobscura*. Genetics **39**: 77–88.

LONNQUIST, J. H., and C. O. GARDNER, 1961 Heterosis in intervarietal crosses in maize and its implication in breeding procedures. Crop Science **1**: 179–183.

MOLL, R. H., W. S. SALHUANA, and H. F. ROBINSON, 1962 Heterosis and genetic diversity in variety crosses of maize. Crop Science **2**: 197–198.

PATERNIANI, E., and J. H. LONNQUIST, 1963 Heterosis in interracial crosses of corn (*Zea mays* L). Crop Science **3**: 504–507.

WALLACE, B., 1955 Inter-population hybrids in *Drosophila melanogaster*. Evolution **9**: 302–316.

WELLHAUSEN, E. J., 1952 Heterosis in a new population. pp. 418–450. *Heterosis*. Edited by J. W. GOWEN. Iowa State College Press, Ames.