

Heterosis: Theory and Estimation

Kendall R. Lamkey and Jode W. Edwards

Heterosis is the foundation of modern maize breeding programs. Despite its importance in maize production, we know surprisingly little about the genetic basis of heterosis. Shull (1952) coined the term heterosis and defined it as: "... the interpretation of increased vigor, size, fruitfulness, speed of development, resistance to disease and to insect pests, or to climatic rigors of any kind, manifested by crossbred organisms as compared with corresponding inbreds, as the specific results of unlikeness in the constitutions of the uniting parental gametes." This definition is often interpreted as not implying a genetic basis for heterosis, because the definition basically describes the phenotype that results from crossing two different inbred lines.

For our purposes, we will define heterosis or hybrid vigor as the difference between the hybrid and the mean of the two parents (Falconer and Mackay, 1996). This difference is often called midparent heterosis. Midparent heterosis is often expressed as a percentage in the literature, but it is important to note that from a quantitative genetic point of view, percentage midparent heterosis is difficult to interpret. The concept of midparent heterosis is rather generic, because it does not refer to the genetic architecture of the parents crossed to produce the hybrids. In many situations this is irrelevant, but if heterosis is to be interpreted by using quantitative genetic theory, then the genetic architecture of the populations becomes extremely important.

Classic theories of heterosis do not distinguish between heterosis at the population level (we will call this panmictic-midparent heterosis) and heterosis resulting from the cross of inbred or partially inbred populations (inbred-midparent heterosis). The objectives of our paper are 1) to present the theory that unifies, conceptually at least, the concepts of panmictic- and inbred-midparent heterosis, and 2) to present some examples from published and unpublished literature of the estimation of heterosis and application of these concepts. The concepts presented in this paper are based on a paper presented at the International Symposium on the Genetics and Exploitation of Heterosis, Mexico City, Mexico (Lamkey and Edwards, 1998).

Population-Level Heterosis Theory

Consider two random-mating populations (P1 and P2) with two alleles per locus (A_1 and A_2). Let the frequency of the i^{th} allele be p_i in Pop 1 and p'_i in Pop 2 (Fig. 1). The F_1 or hybrid population between P1 and P2 is produced by making random crosses between individual plants in P1 and individual plants in P2. Also consider selfing P1 and P2 to complete homozygosity by a procedure such as single-seed descent and designating the inbred versions P1-I and P2-I. P1-I and P2-I can be thought of as single-seed bulks of many individual inbred lines. The F_1 population between P1-I and P2-I is produced in the same way as the F_1 between the noninbred populations. Note then that P1 and P2 and P1-I and P2-I will have different genotypic arrays but identical gametic arrays. The F_1 s produced by crossing P1 x P2 and P1-I x P2-I will, therefore, be identical. The F_2 generation is produced by random mating the F_1 generation. Then by extending the theory of Willham and Pollak (1985) the genetic expectations of the means of each generation are:

$$\bar{F}_{1(f)} = (1 - f)(\bar{F}_2 + 2\Delta^2 d) + fa(\bar{p}_1 - \bar{p}_2),$$

$$\bar{F}_{2(f)} = (1 - f)(\bar{F}_2) + fa(\bar{p}_1 - \bar{p}_2),$$

$$\bar{P}_{1(f)} = (1 - f)(\bar{F}_2 + 2\Delta\alpha - 2\Delta^2 d) + fa(\bar{p}_1 - \bar{p}_2 + 2\Delta), \text{ and}$$

$$\bar{P}_{2(f)} = (1 - f)(\bar{F}_2 - 2\Delta\alpha - 2\Delta^2 d) + fa(\bar{p}_1 - \bar{p}_2 - 2\Delta),$$

where

f = inbreeding coefficient of an generation;

p_i = frequency of the i^{th} allele in P1;

p'_i = frequency of the i^{th} allele in P2;

$\bar{p}_i = \frac{p_i + p'_i}{2}$ = average allele frequency in the cross P1 x P2 or P1-I x P2-I; and

$\delta_i = \frac{p_i - p'_i}{2}$ = one-half the difference in allele frequency between populations.

In the two allele case,

$$\delta_1 = -\delta_2 = \Delta;$$

d = the deviation of the heterozygote from the homozygote midparent;

a = half the difference between homozygotes;

$\alpha = a + d(\bar{p}_2 - \bar{p}_1)$ = average effect of an allele substitution; and

$\bar{F}_2 = a(\bar{p}_1 - \bar{p}_2) + 2\bar{p}_1\bar{p}_2d$ = mean of the F_2 generation.

Note that f , the inbreeding coefficient of a generation, refers to the genetic expectation resulting from selfing a particular generation. For example, if the F_1 generation is selfed one generation, then the genetic expectation is $\bar{F}_{1(0.5)} = 0.5(\bar{F}_2 + 2\Delta^2d) + 0.5a(\bar{p}_1 - \bar{p}_2)$. When $f = 0$, these equations are identical to those of Falconer and Mackay (1996), except that our equations use the F_2 as the reference population, whereas Falconer and Mackay deviated there results from the F_2 mean.

Panmictic-midparent heterosis is defined as the difference between the mean of the F_1 hybrid and mean of the two random-mating parental populations (panmictic-midparent value). F_2 -midparent heterosis is defined as the difference between the mean of the F_2 generation (derived by random mating the F_1) and the panmictic-midparent value. Heterosis of the F_1 -population selfed is defined as the difference between the mean of the F_1 -population selfed and panmictic-midparent value. Algebraically, these heterosis values are:

$$\text{Panmictic-midparent heterosis} = 4\Delta^2d ,$$

$$\text{Panmictic-midparent } F_2 \text{ heterosis} = 2\Delta^2d , \text{ and}$$

$$\text{Panmictic-midparent } F_1\text{-selfed heterosis} = 3\Delta^2d - \bar{p}_1\bar{p}_2d .$$

Four conclusions can be drawn from these expressions: 1) heterosis is dependent on directional dominance; 2) heterosis is a function of the square of the difference in allelic frequency between two populations and therefore, heterosis is specific to a particular cross; 3) randomly mating the F_1 reduces heterosis by 50%; and 4) selfing the F_1 reduces heterosis and is a function of both the difference in allelic frequency and the average allele frequencies between populations at loci with dominance. Although genetic divergence (difference in allelic frequency) and dominance are necessary for there to be heterosis, they are not sufficient in the case of multiple alleles. Cress (1966) showed that with multiple alleles segregating in a population the lack of heterosis cannot be used to infer a lack of genetic divergence between the parental populations. This result has important implications when breeders are screening populations to establish new heterotic groups.

At this point, it is important to understand that we are dealing with population level heterosis, which is much different from the usual case in maize of crosses between two inbred lines. We have started with two random-mating populations, selfed them at random (by a procedure such single seed descent), and then made the F₁ hybrid between the two random mating populations and the two inbred populations. The F₁ hybrid is then a population of hybrids that has a mean and associated variance. Selfing the F₁ and random mating the F₁ lead to different results because of the genetic structure of the populations.

The population concept of heterosis calculates heterosis relative to the mean of two random mating populations, whereas, in the inbred line-hybrid concept of heterosis, heterosis is calculated relative to the mean of the two inbred lines. An equivalent concept in the population model, to the inbred line-hybrid concept of heterosis, is to calculate heterosis relative to the mean of the two inbred ($f=1$) populations (the inbred-midparent value). Algebraically, this leads to the following expressions:

$$\text{Inbred-midparent heterosis} = 2\bar{p}_1\bar{p}_2d + 2\Delta^2d ,$$

$$\text{Inbred-midparent } F_2 \text{ heterosis} = 2\bar{p}_1\bar{p}_2d , \text{ and}$$

$$\text{Inbred-midparent } F_1\text{-selfed heterosis} = \bar{p}_1\bar{p}_2d + \Delta^2d .$$

Note that this leads to different results than when heterosis is calculated relative to the panmictic-midparent value. Inbred-midparent F₂ heterosis is equal to the inbreeding depression in the F₂ of the hybrid population when the F₂ is inbred to $f=1$. This is also equal to the average inbreeding depression in P1 and P2. Inbred-midparent heterosis is then equal to the average inbreeding depression plus one-half of the panmictic-midparent heterosis. Selfing the F₁ results in a 50% reduction in heterosis.

Calculating heterosis relative to the inbred midparent reveals that inbred-midparent heterosis is more than just a function of divergence in allelic frequencies between two populations. The cross between two inbred populations must recover the performance that was lost during inbreeding, plus have enough panmictic heterosis to produce a good hybrid. This gives rise to the concept of *baseline heterosis*. We define *baseline heterosis* as the difference between the panmictic-midparent value and the inbred-midparent value. It is the average amount of heterosis that will be present in crosses among inbred lines derived from the two populations.

Inbred-midparent heterosis is then equal to baseline heterosis plus panmictic heterosis.

Algebraically, the relationships can be presented as follows:

$$\begin{aligned}
 \text{Baseline heterosis} &= (\bar{P}_{1(0)} + \bar{P}_{2(0)})/2 - (\bar{P}_{1(1)} + \bar{P}_{2(1)})/2 \\
 &= \text{Panmictic midparent value} - \text{Inbred midparent value} \\
 &= \text{Inbred midparent heterosis} - \text{panmictic midparent heterosis} \\
 &= 2\bar{p}_1\bar{p}_2d - 2\Delta^2d .
 \end{aligned}$$

$$\begin{aligned}
 \text{Inbred-midparent heterosis} &= \text{baseline heterosis} + \text{panmictic heterosis} \\
 &= 2\bar{p}_1\bar{p}_2d + 2\Delta^2d .
 \end{aligned}$$

Baseline heterosis is adjusted for genetic divergence, because some of the inbreeding depression that occurs in the F_2 of the F_1 hybrid is due to genetic divergence. Note that inbred-midparent heterosis is a function of inbreeding depression, genetic divergence, and dominance whereas panmictic-midparent heterosis is a function only of genetic divergence and dominance. The relationships among these concepts is presented diagrammatically in Fig. 2.

Baseline and panmictic heterosis for dominance and overdominance are plotted in Fig. 3. The graphs are very similar for the two types of gene action with the major difference being the magnitude; more heterosis is observed with overdominance than with dominance. The second and most important point is that panmictic-midparent heterosis only exceeds baseline heterosis when allelic frequencies are at the extremes. This is an important point to keep in mind when studying heterosis among inbred lines. A significant portion of the heterosis among inbred lines is due simply to recovery of what was lost during inbreeding and in some instances little of the observed heterosis may actually be due to genetic divergence.

Inbred Line-Hybrid Heterosis Theory

The same theory can be used to derive the heterosis obtained from crossing two inbred lines if a few simple relationships are observed. At a two allele locus, $\bar{p}_1 = p_1 + \Delta$ and $\bar{p}_2 = p_2 - \Delta$. Because allele frequency can be only 0 or 1 in an inbred line, this implies that $\bar{p}_1 = \bar{p}_2 = 1/2$ and that $\Delta = 1/2$. Using this result, our model reduces to a standard generation

means model with the F_2 as the reference population. Algebraically, inbred line-hybrid heterosis can be expressed as:

$$\text{Inbred-midparent heterosis} = d,$$

$$\text{Inbred-midparent } F_2 \text{ heterosis} = 1/2d, \text{ and}$$

$$\text{Inbred-midparent } F_1\text{-selfed heterosis} = 1/2d.$$

As expected random mating the F_1 and selfing the F_1 give the same result and lead to a 50% reduction in heterosis. These equations also illustrate that there are two ways to calculate heterosis between two inbred lines. The first is the conventional way of subtracting the mean of the inbreds from the F_1 mean. The second is by selfing the F_1 hybrid and evaluating both the F_1 and F_2 generations in the field. If this is done, heterosis can be calculated as two times the difference between the F_1 and F_2 generations. This method has the advantage of not requiring estimates of inbred performance, which are often difficult to obtain and may lead to biased estimates of heterosis. Also note that under the inbred-line model, that baseline heterosis is zero. It is not possible to obtain information on how much of the heterosis between two inbred lines is due to genetic diversity and how much is due to inbreeding depression, because the two are confounded together in inbred lines.

Estimation of Heterosis

There are many different ways of expressing heterosis. When heterosis is reported in the literature, we need to be sure that authors are clear on how they have calculated and interpreted their heterosis estimates. We have dealt entirely with absolute estimates of heterosis. Heterosis estimates expressed as a percentage do not have a quantitative genetic interpretation and should only be used for making comparisons among hybrids as to the relative magnitudes of heterosis. Interpreting changes in percentage heterosis in light of quantitative genetics has no basis in theory.

Example #1

To illustrate the complexity of interpreting heterosis estimates, we give an example from a reciprocal recurrent selection (RRS) program (Keeratinijakal and Lamkey, 1993a,b). Keeratinijakal and Lamkey (1993a,b) conducted a detailed evaluation of 11 cycles of RRS in

BSSS and BSCB1. Many different types of populations were evaluated in this study, but the ones of importance for heterosis are the Cn x Cn interpopulation crosses, the Cn x Cn interpopulation crosses selfed, and the Cn populations per se from BSSS and BSCB1. In Fig. 4a, we have plotted panmictic, inbred, and baseline heterosis for each cycle of selection. The three types of heterosis were calculated for each cycle as follows:

Panmictic-midparent heterosis = F_1 – panmictic midparent value ,

Inbred-midparent heterosis = $2(F_1 - F_1 \otimes)$

Baseline heterosis = Inbred-midparent heterosis - Panmictic-midparent heterosis.

Fig. 2a reveals two noteworthy features. 1) Panmictic heterosis increased substantially over cycles of selection; and 2) Inbred and panmictic heterosis are very similar, which results in negative estimates of baseline heterosis. Negative estimates of baseline heterosis are not expected for grain yield in maize. Clearly, something is amiss with these data. We believe that our estimate of inbred heterosis is very good. The only thing that would make the estimate better is additional generations of selfing of the F_1 hybrid. The estimate of baseline heterosis is obtained by subtraction, so it is not the problem. The problem clearly lies with the estimate of panmictic heterosis, although this was the estimate reported by Keeratinijakal and Lamkey (1993a). The pattern of selection response is such that the F_1 increased at a rate of about 7% per cycle, but the populations per se did not respond to selection. Keeratinijakal and Lamkey (1993a) reported that the increase in heterosis is primarily due to a lack of response in the populations per se.

The question about how panmictic heterosis has changed with cycles of selection still remains. Answering this question, requires an understanding of why the populations per se did not respond to selection. Keeratinijakal and Lamkey (1993b) in a more detailed genetic analysis reported that the lack of response in the populations was due to inbreeding depression from random genetic drift. By using the model that they fit to the data, it is possible to predict the means of the populations per se as $AOI + 2DOI + 2n(ALI + DLI)$, where n is the cycle of selection and the other terms are described in Keeratinijakal and Lamkey (1993b). This model predicts the performance of the populations per se in the absence of inbreeding depression due to genetic drift. The new estimates of the population means were then used to recalculate the panmictic-midparent value, panmictic heterosis, and baseline heterosis. Inbred heterosis was

unaffected by this calculation. The results are shown in Fig. 4b and are quite different from those shown in Fig. 4a. The results in Fig 4b show that most of the increase in inbred heterosis is due to an increase in baseline heterosis. Furthermore selection has done little to change panmictic heterosis and it appears to be decreasing rapidly during the last two cycles of selection. These two analyses of the same data lead to strikingly dissimilar conclusions.

Example #2

The second example of estimating heterosis is taken from Eyherabide and Hallauer (1991). This study was similar to the one of Keeratinijakal and Lamkey (1993a,b) only Eyherabide and Hallauer studied the response to eight cycles of full-sib reciprocal recurrent selection. Panmictic, inbred, and baseline heterosis were estimated in the same manner as for the Keeratinijakal and Lamkey study for both the observed and adjusted heterosis. If we only calculated panmictic heterosis, we would be lead to the conclusion that selection has substantially increased heterosis and that this is due to an increase in genetic divergence between the populations (Fig. 5a). When we calculate baseline heterosis, we find that it has been relatively flat over cycles of selection, whereas inbred heterosis has shown a rather steady increase over cycles (Fig. 5a.).

If the base populations are adjusted for the amount of inbreeding depression that has occurred due to genetic drift, a different picture emerges as in the study of Keeratinijakal and Lamkey. Inbred heterosis is the same before and after adjustment for genetic drift (Fig. 5b). Baseline heterosis now increases at a rate that roughly parallels the increase in inbred heterosis. Panmictic heterosis is now relatively flat and fluctuates around a value of roughly 0.5 Mg ha^{-1} .

Interpretation

Proper interpretation of the results from these two studies will require more detailed analysis of the data and results from other aspects of the studies. The point to be made here is that rote estimation of heterosis can be deceiving. These examples clearly point out that heterosis at the population level is extremely sensitive to the inbreeding level of the populations. If all populations have the same level of inbreeding, then hybrid to hybrid comparisons can be made, but absolute estimates of panmictic heterosis may be substantially overestimated.

Conclusions

The data clearly indicate a need for further empirical and theoretical research into heterosis; however, we need to be very careful about how future experiments are designed and analyzed. Enfield (1977) was critical of empirical quantitative genetic experiments indicating that the literature was either cluttered with experiments that were meaningless because of standard errors that were too large or because standard errors were absent altogether. Despite the tremendous developments that have recently occurred in molecular biology, quantitative genetics is still the only theory linking genotype to phenotype. It is imperative that we design better experiments to test the adequacy and validity of quantitative genetic models.

We would like to propose several areas of research that are needed to better understand heterosis. We will not present experimental approaches, because we do not have all of the answers.

- 1) Gene action and effects are key to understanding the inheritance of quantitative traits. For maize at least, it seems that from average population estimates, there is no evidence for overdominance. Although this is useful information, what would be even better is to gain insight into the distribution of gene effects and gene action for individual traits. Are there lots of loci with equal and small effects or is there some type of distribution? The conventional approach to gene action studies will not answer this question and new approaches will be needed.
- 2) Selection experiments in plants need to be better designed. Most of our current recurrent selection experiments are not adequately designed to separate the effects of selection from drift, so it becomes nearly impossible to reliably interpret the results of these experiments. It is clear that recurrent selection works, and future experiments to demonstrate the effectiveness of recurrent selection are probably not needed. But we do need well-designed experiments with adequate controls and replication. New selection experiments should either be replicated or include an unselected, replicated control population of the same effective size.
- 3) The view of random genetic drift in agricultural is the one that drift leads to a loss of

heterozygosity and eventual erosion of genetic variance. Although this is true in the case of additive gene action, with dominance and epistasis, drift may reduce heterozygosity with a corresponding increase in additive genetic variance. We need to incorporate this new information from evolutionary biology into the design of our breeding programs.

- 4) Epistasis has long been ignored in breeding programs and is generally assumed to be absent or unimportant. Evidence from molecular biology clearly shows that genes interact. Recent theory from evolutionary biology that distinguishes physiological epistasis from population epistasis may indicate why population level epistasis may be undetectable. More theoretical work is needed to optimally design breeding programs to select for epistatic effects.
- 5) Despite several classic studies, we know very little about inbreeding depression in plants. Because much of the observed heterosis among inbred lines may be due to the recovery of inbreeding depression, the genetics of heterosis may be best elucidated by studying the genetics of inbreeding depression. Our preoccupation with heterosis has caused us to overlook the importance of inbreeding depression.

There are of course several problems in designing and conducting good quantitative genetics experiments. First, they are often large and consume considerable physical resources, even for lab species. Second, there seems to be little funding available in agricultural species to do quantitative genetics and plant breeding related research. Third, there are few scientists being trained to do this type of research. Lack of funding and support in quantitative genetics may in the long term severely limit future genetic gains.

ACKNOWLEDGEMENTS

Joint contribution from the Corn Insects and Crop Genetics Research Unit, Agricultural Research Service, USDA, Dept. of Agronomy, Iowa State Univ. and Journal Paper No. J-17770 of the Iowa Agric. and Home Economics Exp. Stn. Project No. 3495, and supported by Hatch Act and State of Iowa.

References

- Cress, C. E. 1966. Heterosis of the hybrid related to gene frequency differences between two populations. *Genetics* 53:269-74.
- Enfield, F. D. 1977. Selection experiments in *Tribolium* designed to look at gene action issues. p. 177-190. *In*: E. Pollak, O. Kempthorne, and T. B. Bailey, Jr. (eds.) Proceedings of the International Conference on Quantitative Genetics, Ames, IA. 16-21 Aug. 1976. Iowa State University Press, Ames.
- Eyherabide, G. H. and A. R. Hallauer. 1991. Reciprocal full-sib recurrent selection: II. Contributions of additive, dominance, and genetic drift effects. *Crop Sci.* 31:1442-1448.
- Falconer, D. S. and T. F. C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, Essex, England
- Keeratinijakal, V. and K. R. Lamkey. 1993a. Responses to reciprocal recurrent selection in BSSS and BSCB1 maize populations. *Crop Sci.* 33:73-77.
- Keeratinijakal, V. and K. R. Lamkey. 1993b. Genetic effects associated with reciprocal recurrent selection in BSSS and BSCB1 maize populations. *Crop Sci.* 33:78-82.
- Lamkey, K. R. and J. W. Edwards. 1998. The quantitative genetics of heterosis. p. xx-xx. *In*: Coors et. al. The genetics and exploitation of heterosis in crop plants. Crop Science Society of America, Madison, WI
- Shull, G. H. 1952. Beginnings of the heterosis concept. p. 14-48. *In*: Heterosis J. W. Gowen (ed.). Iowa State College Press, Ames.
- Willham, R. L. and E. Pollak. 1985. Theory of heterosis. *J. Dairy Sci.* 68:2411-2417.

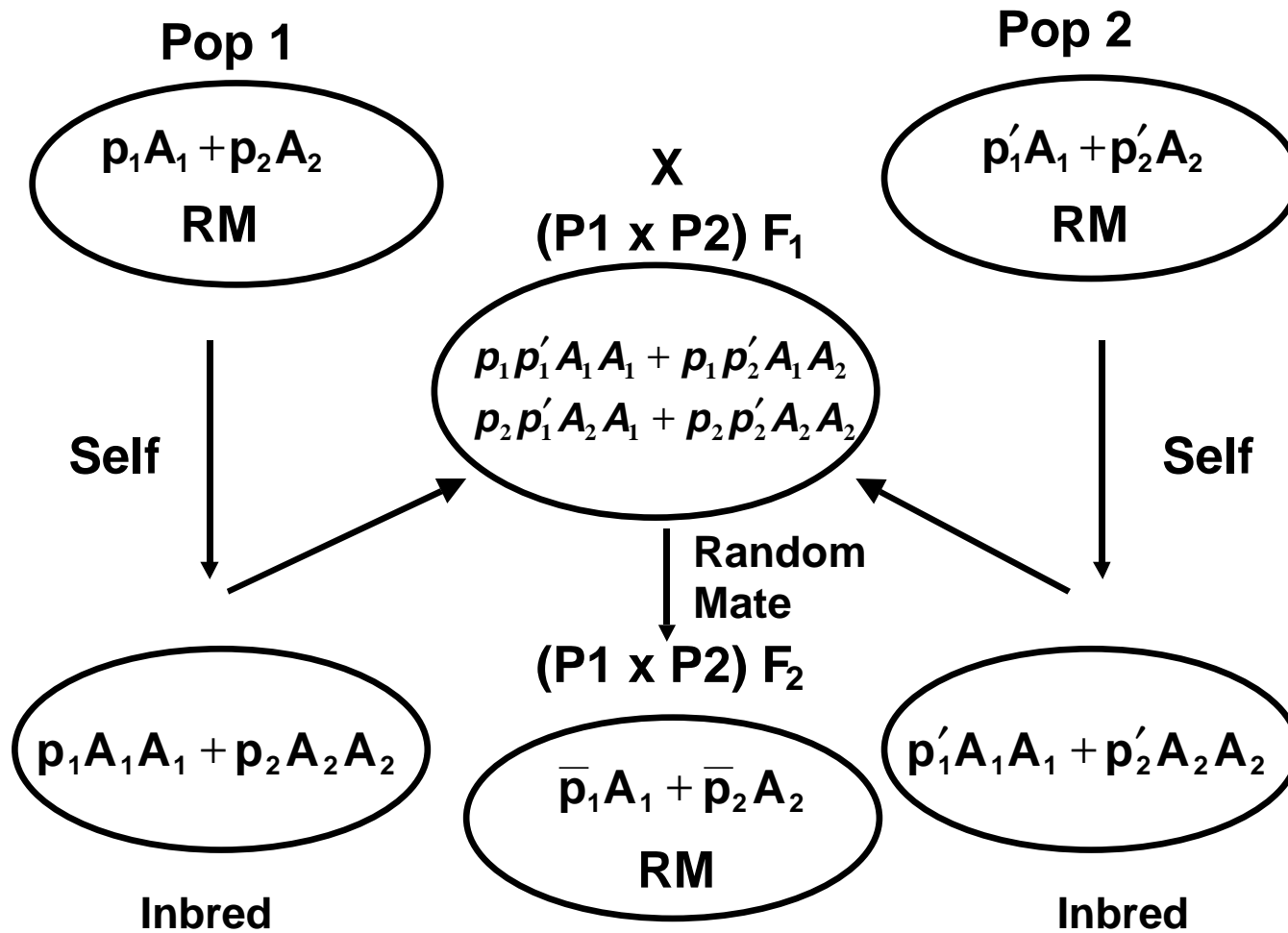


Fig. 1. Mating scheme for populations described in text. Populations 1 and 2 start with the gametic arrays shown and are in random mating equilibrium (panmixia). Crossing the panmictic populations together forms the F₁ shown. Random mating the F₁ gives rise to the F₂ generation, the population in which genetic effects are described in Willham and Pollak (1985). Inbreeding populations 1 and 2 to complete homozygosity generates the populations with the genotypic arrays shown. Crossing these two inbred populations produces the same F₁ as crossing the two panmictic populations.

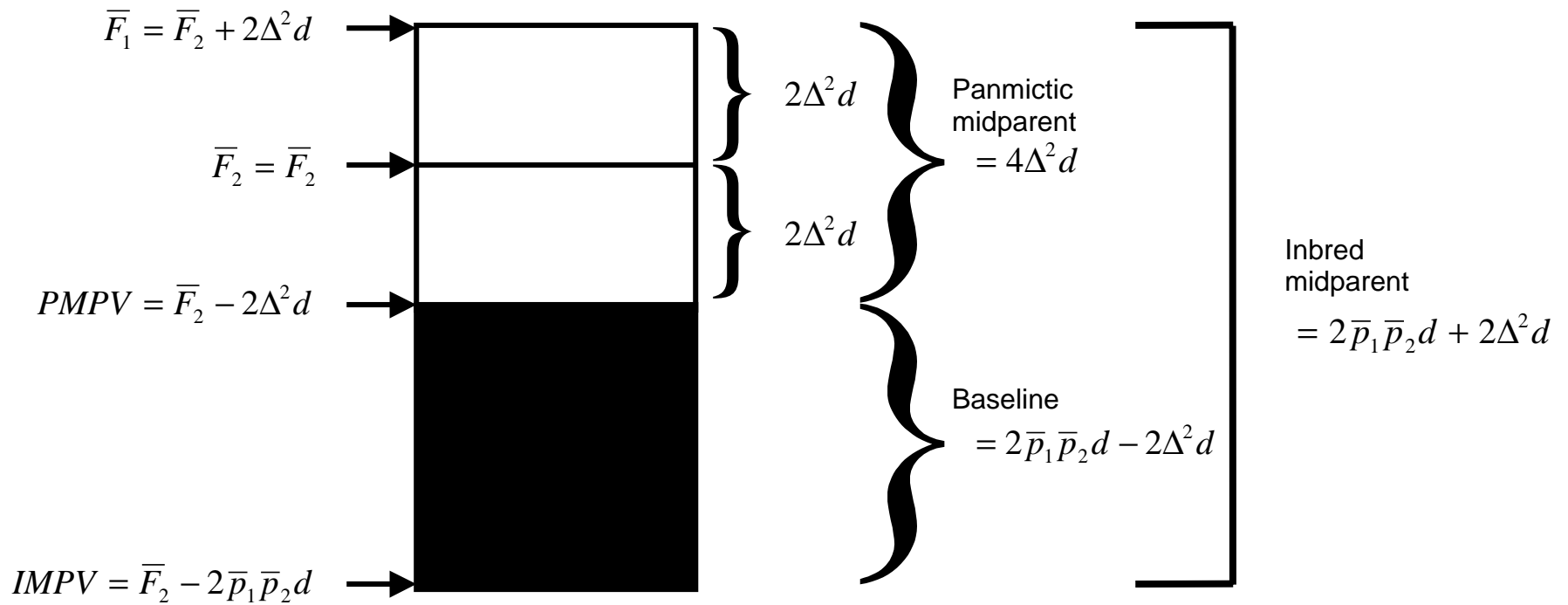


Fig. 2. A diagrammatic representation of population heterosis using the F2 population as the reference point. All symbols are defined in the text, except PMPV = panmictic midparent value and IMPV = inbred midparent value. Note that there was no intention to draw the diagram to scale.

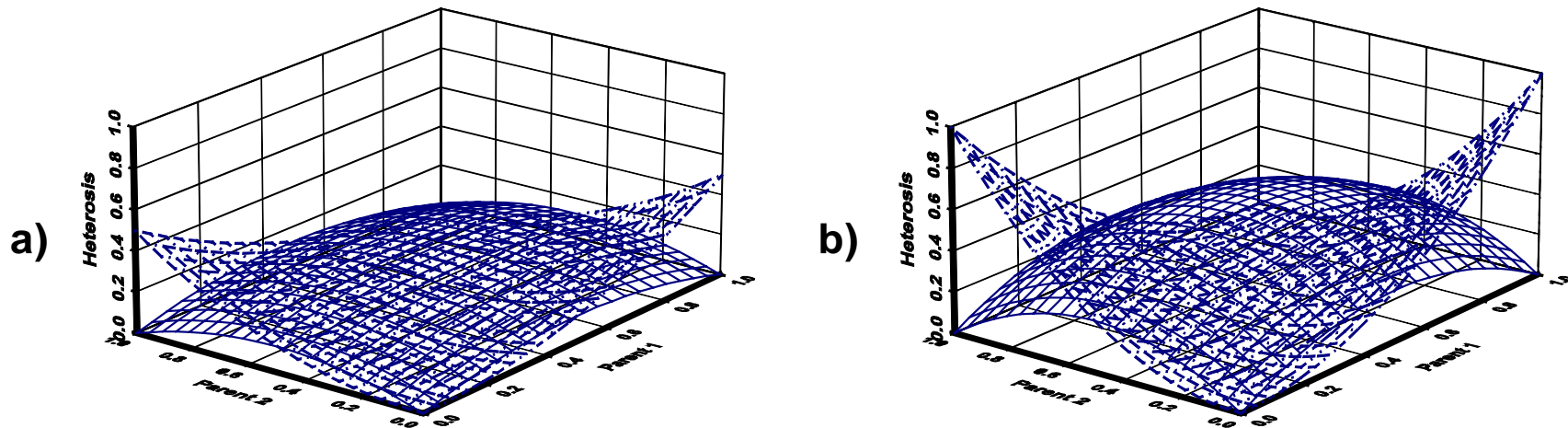


Fig. 3. Plots of baseline (solid lines) and functional heterosis (dashed lines) versus allelic frequencies in the parents. Inbred-midparent heterosis is the sum of panmictic-midparent heterosis and baseline heterosis. Fig. a) is for the case of complete dominance and b) is for the case of pure overdominance.

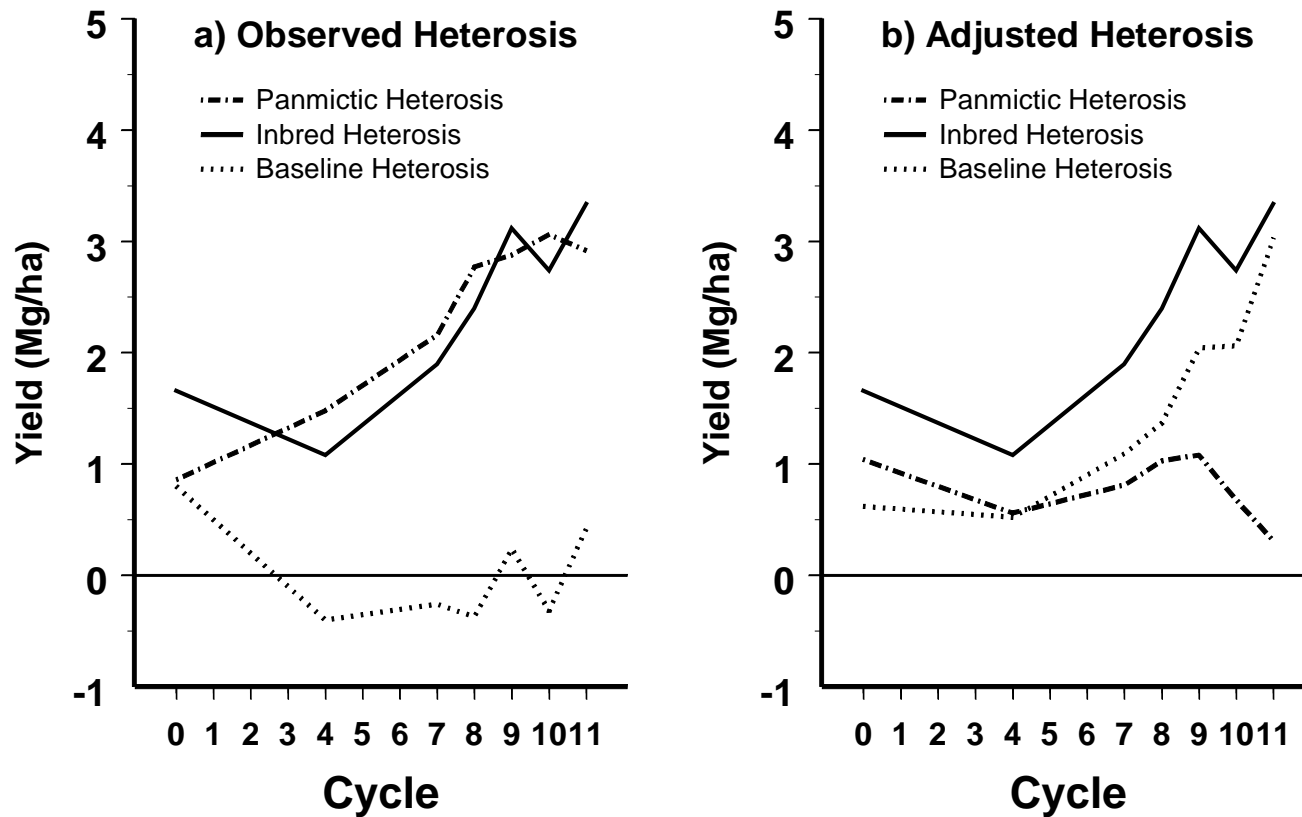


Fig. 4. Heterosis estimates from the BSSS and BSCB1 reciprocal recurrent selection program. Data taken from Keeratinijakal and Lamkey (1993a,b). Panmictic heterosis was calculated as the mean of the F_1 minus the mean of the parental populations. Inbred heterosis was calculated as the two times the F_1 mean minus the F_1 -selfed mean. Baseline heterosis was calculated as inbred heterosis minus panmictic heterosis. a) Panmictic midparent value was calculated using the observed data for the cycle means. b) Panmictic midparent value was calculated using cycle means adjusted for inbreeding depression due to random genetic drift.

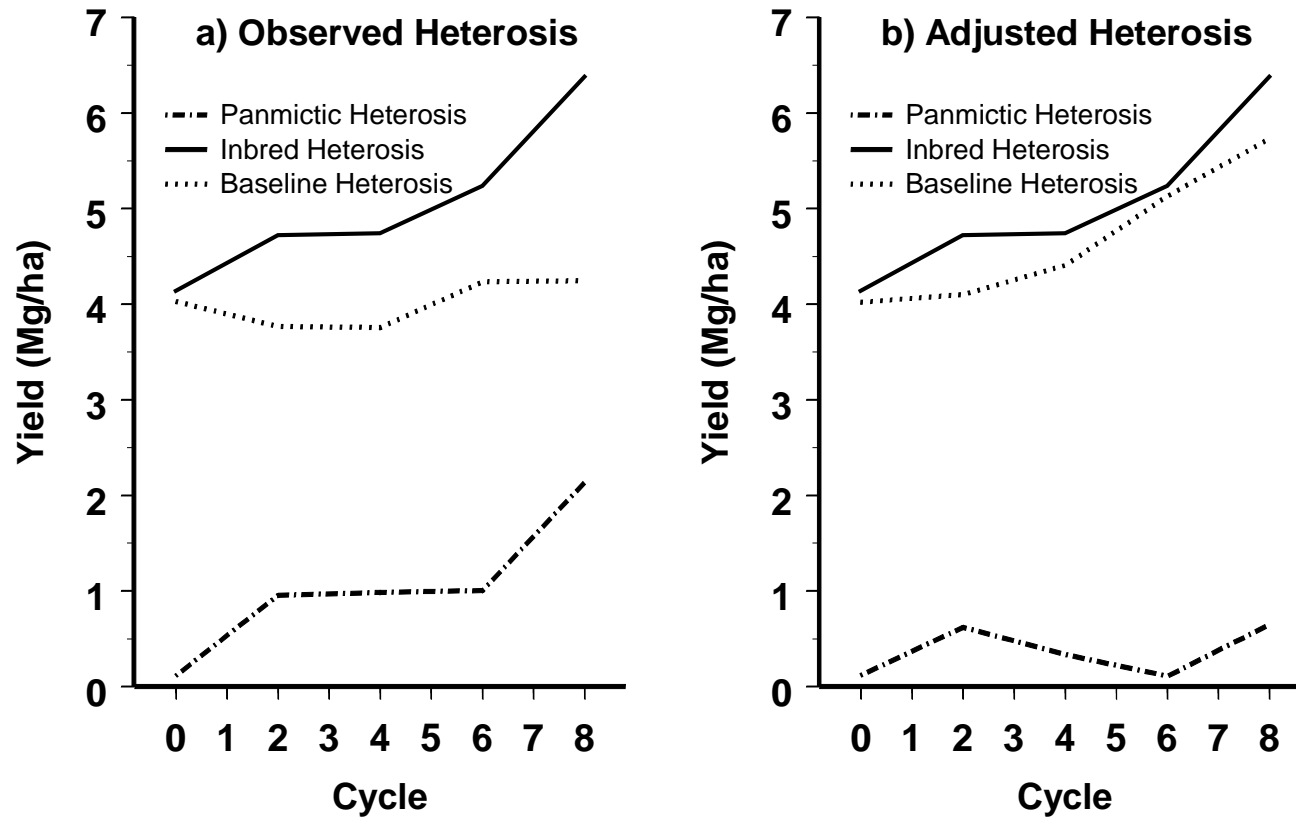


Fig. 5. Heterosis estimates from the BS10 and BS11 full-sib reciprocal recurrent selection program. Data taken from Eyherabide and Hallauer (1991). Panmictic heterosis was calculated as the mean of the F_1 minus the mean of the parental populations. Inbred heterosis was calculated as the two times the F_1 mean minus the F_1 -selfed mean. Baseline heterosis was calculated as inbred heterosis minus panmictic heterosis. a) Panmictic midparent value was calculated using the observed data for the cycle means. b) Panmictic midparent value was calculated using cycle means adjusted for inbreeding depression due to random genetic drift.