

# An approach for predicting heterosis based on an additive, dominance and additive $\times$ additive model with environment interaction

ZI CHENG XU† & JUN ZHU\*

Department of Agronomy, Zhejiang Agricultural University, Hangzhou 310029, China

A method is proposed for predicting potential heterosis of offspring of crop hybrids by an additive, dominance and additive  $\times$  additive model (ADAA). By using unbiased predictors of additive and dominance, as well as additive  $\times$  additive effects, general formulae for predicting heterosis over mid-parent and heterosis over the better parent are derived for different generations. When there exists genotype by environment (GE) interaction, formulae are also derived for predicting interaction heterosis. Heterosis in a specific environment is the sum of heterosis arising from the main genotypic effect and that arising from GE interaction deviation. The epistasis heterosis ( $\Delta_{AA}$ ) could play an important role in the use of heterosis for both an  $F_1$  hybrid and its later generations. In addition, a simple formula is given for predicting the number of generations of a cross that would still keep a certain level of heterosis over the better parent. Data from a diallel cross of cotton are analysed as a worked example for predicting genotypic value, heterosis, and the number of generations for each cross when heterosis over the better parent is larger than 5%.

**Keywords:** epistasis model, GE interaction, genetic effects, heterosis prediction.

## Introduction

Heterosis is an important way of increasing yield and improving quality in crops. Since 1914, when the term 'heterosis' was coined and first proposed by G. H. Shull (see Stuber *et al.*, 1992), genetic research on heterosis has been an important issue. Since heterosis over environments is variable (Virmani *et al.*, 1982; Young & Virmani, 1990) and environment-dependent (Knight, 1973), the genetic basis of heterosis *per se* is very complicated. Up to now a large number of methods addressing various aspects of heterosis have been developed based on biomathematical models (for a review see Schnell & Cockerham, 1992) and/or molecular techniques (Smith *et al.*, 1990; Bernardo, 1992; Stuber *et al.*, 1992; Zhang *et al.*, 1994). Numerous authors have reported inconsistent relationships between yield heterosis and: (i) geographical distance (Moll *et al.*, 1962); (ii) genealogical distance (Cowen & Fery, 1987); (iii) genetic divergence based on both

quantitative traits and pedigree relationships (Cox & Murphy, 1990); and (iv) genetic distance based on molecular markers (Smith *et al.*, 1990; Bernardo, 1992; Zhang *et al.*, 1994). Because many traits of agronomic importance are quantitative in nature, controlled by polygenes and affected by environments, prediction of heterosis might not be very effective.

However, most of the approaches for predicting heterosis are based on information from the parents only. From a practical point of view, another key issue is whether we can predict the number of generations maintaining mid-parent or the better parent heterosis. An additional key issue is how to remove the influence of environmental effects and how to predict the performance of crosses (or maintenance of heterosis) in later generations from phenotypic observations on parents, the  $F_1$  and the  $F_2$  generations. Zhu (1993, 1997) and Zhu *et al.* (1993) proposed methods for directly predicting genotypic values and heterosis with an additive–dominance model. But an additive–dominance model may not fit all quantitatively inherited traits, because epistatic effects can also be an important component of genetic effects. Genetic models neglecting epistasis may result in biased information (Ketata *et al.*, 1976).

\*Correspondence. E-mail: jzhu@zjau.edu.cn

†Permanent Address: Department of Agronomy, Henan Agricultural University, Zhengzhou 450002, China.

Schnell & Cockerham (1992) pointed out that heterosis without dominance arises from additive × additive epistasis alone. Zhu (1989) proposed an additive, dominance and additive × additive model (ADAA model) according to the principles of the general genetic model (Cockerham, 1980). By applying mixed-model approaches to analysing the ADAA model, variance components, heritability and genetic effects can be estimated (or predicted); then the genetic performance of parents and their crosses can be evaluated.

Based on the ADAA model with genotype × environment interaction, the objectives of present paper are: (i) to derive general formulae for predicting heterosis over mid-parent, heterosis over the better parent and deviation of heterosis caused by environment interaction of hybrids; (ii) to predict for how many generations heterosis of a cross could be maintained by using predicted genotypic values; and (iii) to show the importance of additive × additive effects in the utilization of heterosis for an F<sub>1</sub> hybrid and its progeny. An example of yield traits in cotton is used to demonstrate the prediction of heterosis under different environments and its application in a breeding programme.

### Genetic model

The assumptions for the ADAA model are: (i) normal diploid segregation; (ii) inbred parents in diallel mating are a random sample from a reference population; (iii) no additive × dominance epistasis and no dominance × dominance epistasis; (iv) linkage equilibrium. When genetic experiments are carried out with a randomized complete block design within multiple environments, a linear model can be written for the phenotypic mean value  $y_{hijkl}$  of the  $k$ th mating type ( $k = 0$  for parent,  $k = 1$  for F<sub>1</sub>,  $k = 2$  for F<sub>2</sub>) from lines  $i$  and  $j$  in the  $l$ th block within the  $h$ th environment:

$$Y_{hijkl} = \mu + E_h + G_{ijk} + GE_{hijk} + B_{l(h)} + \varepsilon_{hijkl}, \quad (1)$$

where  $\mu$  and  $E_h$  are the population mean and the environmental effect, respectively;  $G_{ijk}$  is the total genetic main effect, and  $GE_{hijk}$  is the total genotype × environment (GE) interaction;  $B_{l(h)}$  is the randomized complete block effect,  $B_{l(h)} \sim (0, \sigma_B^2)$ ;  $\varepsilon_{hijkl}$  is the residual effect,  $\varepsilon_{hijkl} \sim (0, \sigma_\varepsilon^2)$ .

If the experiment is conducted by a modified diallel mating with a set of parents and their F<sub>1</sub>s and F<sub>2</sub>s,  $G_{ijk}$  and  $GE_{hijk}$  can be partitioned into genetic components for different generations (Zhu, 1989):

for parent P<sub>*i*</sub> ( $i = j, k = 0$ ):

$$G_{ii0} = 2A_i + D_{ii} + 4AA_{ii},$$

$$GE_{hii0} = 2AE_{hi} + DE_{hii} + 4AAE_{hii};$$

for F<sub>1*ij*</sub> from line  $i \times$  line  $j$  ( $k = 1$ ):

$$G_{ij1} = A_i + A_j + D_{ij} + AA_{ii} + AA_{jj} + 2AA_{ij},$$

$$GE_{hij1} = AE_{hi} + AE_{hj} + DE_{hij} + AA_{E_{hii}} + AA_{E_{hjj}} + 2AAE_{hij};$$

and for F<sub>2*ij*</sub> derived from selfing of F<sub>1*ij*</sub> ( $k = 2$ ):

$$G_{ij2} = A_i + A_j + \frac{1}{4}D_{ii} + \frac{1}{4}D_{jj} + \frac{1}{2}D_{ij} + AA_{ii} + AA_{jj} + 2AA_{ij},$$

$$GE_{hij2} = AE_{hi} + AE_{hj} + \frac{1}{4}DE_{hii} + \frac{1}{4}DE_{hjj} + \frac{1}{2}DE_{hij} + AA_{E_{hii}} + AA_{E_{hjj}} + 2AAE_{hij};$$

where  $A_i$  and  $A_j$  are cumulative additive effects from line  $i$  and line  $j$ , respectively,  $A_i$  and  $A_j \sim (0, \sigma_A^2)$ ; the cumulative dominance effect is  $D_{ij}$  (or  $D_{ii}, D_{jj}$ )  $\sim (0, \sigma_D^2)$ ; the cumulative additive × additive effect is  $AA_{ij}$  (or  $AA_{ii}, AA_{jj}$ )  $\sim (0, \sigma_{AA}^2)$ ; the cumulative additive × environment interaction is  $AE_{hi}$  (or  $AE_{hj}$ )  $\sim (0, \sigma_{AE}^2)$ ; the cumulative dominance × environment interaction is  $DE_{hij}$  (or  $DE_{hii}, DE_{hjj}$ )  $\sim (0, \sigma_{DE}^2)$ ; the cumulative additive × additive × environment interaction is  $AAE_{hij}$  (or  $AAE_{hii}, AAE_{hjj}$ )  $\sim (0, \sigma_{AAE}^2)$ .

### Prediction methods

Genetic model (1) can be expressed in the form of a mixed linear model:

$$\mathbf{y} = \mathbf{Xb} + \sum_{u=1}^m \mathbf{U}_u \mathbf{e}_u + \mathbf{e}_m \sim \left( \mathbf{Xb}, \mathbf{V} = \sum_{u=1}^m \sigma_u^2 \mathbf{U}_u \mathbf{U}_u^T + \sigma_m^2 \mathbf{I} \right), \quad (2)$$

where  $\mathbf{y}$  is the vector of phenotype values with mean  $\mathbf{Xb}$  and variance  $\mathbf{V}$ ;  $\mathbf{b}$  is the vector of fixed effects;  $\mathbf{X}$  is the known incidence matrix relating to the fixed effects;  $\mathbf{e}_u$  is the vector of the  $u$ th random factor,  $\mathbf{e}_u \sim (\mathbf{0}, \sigma_u^2 \mathbf{I})$ ;  $\mathbf{U}_u$  is the known incidence matrix relating to the random vector  $\mathbf{e}_u$ .

Random effects in the genetic model are predictable without bias by the methods of linear unbiased prediction (LUP) (Zhu, 1992; Zhu & Weir, 1996a), and adjusted unbiased prediction (AUP) (Zhu, 1993; Zhu & Weir, 1996b). Because Monte Carlo simulation revealed that LUP could give prediction with unbiased mean but underestimated variance for random variables (Zhu & Weir, 1996b), the AUP method was suggested for predicting genetic effects:

$$\hat{\mathbf{e}}_{u(x)} = \kappa_u(\alpha_u \mathbf{U}_u^T \mathbf{Q}_x \mathbf{y}),$$

where  $\mathbf{Q}_x = \mathbf{V}_x^{-1} \mathbf{V}_x^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{V}_x^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_x^{-1}$ ,

$$\mathbf{V}_x = \sum_{u=1}^m \alpha_u \mathbf{U}_u \mathbf{U}_u^T + \alpha_m \mathbf{I} \text{ with inverse } \mathbf{V}_x^{-1}, \text{ and}$$

$$\kappa_u = \sqrt{(n_u - 1) \hat{\sigma}_u^2 / (\alpha_u^2 \mathbf{y}^T \mathbf{Q}_x \mathbf{U}_u \mathbf{U}_u^T \mathbf{Q}_x \mathbf{y})}.$$

$\alpha_u$  is the prior value of the  $u$ th random factor.  $\kappa_u$  is an adjusted coefficient to ensure,  $\hat{\mathbf{e}}_{u(x)}^T \hat{\mathbf{e}}_{u(x)} / (n_u - 1) = \hat{\sigma}_u^2$ , and set  $\hat{\sigma}_u^2 = 0$  when  $\hat{\sigma}_u^2 < 0$ .

When obtaining unbiased prediction of additive effects, dominance effects, additive  $\times$  additive effects and their environment interaction effects, the genotypic values ( $\mu + G$ ) of parents and their offspring could be predicted without bias. Zhu (1993) proposed an approach to predict heterosis for an additive–dominance model and found that general heterosis over mid-parent for an  $F_1$  generation  $H_M(F_1)$  can be expressed as a function of dominance heterosis ( $\Delta_D$ ) which is the difference between the heterozygote dominance effect and the average of the homozygote dominance effects ( $D_{ij} - \frac{1}{2}(D_{ii} + D_{jj})$ ). Expected heterosis over mid-parent for the  $F_n$  generation  $H_M(F_n)$  can be expressed as a function of  $H_M(F_1)$  accordingly, i.e.  $H_M(F_n) = (\frac{1}{2})^{n-1} H_M(F_1) = (\frac{1}{2})^{n-1} \Delta_D$ . General heterosis over the better parent for the  $F_n$  generation  $H_B(F_n)$  can also be expressed as a function of  $H_M(F_1)$  and parental genetic difference ( $\omega_G = |2(A_i - A_j) + (D_{ii} - D_{jj})|$ ), i.e.  $H_B(F_n) = (\frac{1}{2})^{n-1} H_M(F_1) - \frac{1}{2} \omega_G$ . When there exists GE interaction, formulae for predicting interaction heterosis are derived in a similar way (Zhu, 1997). Heterosis in a specific environment therefore consists of two components: general heterosis arising from genetic main effects, and interaction heterosis arising from GE interaction effects.

The total heterosis over mid-parent for  $F_n = H_M(F_n) + H_{ME}(F_n)$  and the total heterosis over the better parent for  $F_n = H_B(F_n) + H_{BE}(F_n)$ , where general heterosis is the performance of heterosis expected across different environments, whereas interaction heterosis is the deviation from general heterosis in a specific environment.

When the additive–dominance model is expanded by including an additive  $\times$  additive effect, formulae for predicting general heterosis and interaction heterosis of each generation in different environments can be derived as follows.

General heterosis and interaction heterosis over mid-parent for the  $F_1$  are:

$$\begin{aligned} H_M(F_1) &= G(F_{1ij}) - \frac{1}{2}[G(P_i) + G(P_j)] \\ &= \left[ D_{ij} - \frac{1}{2}(D_{ii} + D_{jj}) \right] \\ &\quad + 2 \left[ AA_{ij} - \frac{1}{2}(AA_{ii} + AA_{jj}) \right] \\ &= \Delta_D + 2\Delta_{AA} \text{ and} \end{aligned}$$

$$\begin{aligned} H_{ME}(F_1) &= GE(F_{1ij}) - \frac{1}{2}[GE(P_i) + GE(P_j)] \\ &= \left[ DE_{hij} - \frac{1}{2}(DE_{hii} + DE_{hjj}) \right] \\ &\quad + 2 \left[ AAE_{hij} - \frac{1}{2}(AAE_{hii} + AAE_{hjj}) \right] \\ &= \Delta_{DE} + 2\Delta_{AAE}, \end{aligned}$$

setting dominance heterosis  $\Delta_D = D_{ij} - \frac{1}{2}(D_{ii} + D_{jj})$ , epistasis heterosis  $\Delta_{AA} = AA_{ij} - \frac{1}{2}(AA_{ii} + AA_{jj})$ , dominance  $\times$  environment heterosis  $\Delta_{DE} = DE_{hij} - \frac{1}{2}(DE_{hii} + DE_{hjj})$  and epistasis  $\times$  environment heterosis  $\Delta_{AAE} = AAE_{hij} - \frac{1}{2}(AAE_{hii} + AAE_{hjj})$ .

General heterosis and interaction heterosis over mid-parent for the  $F_2$  are:

$$\begin{aligned} H_M(F_2) &= \frac{1}{2} \left[ D_{ij} - \frac{1}{2}(D_{ii} + D_{jj}) \right] \\ &\quad + 2 \left[ AA_{ij} - \frac{1}{2}(AA_{ii} + AA_{jj}) \right] \\ &= \frac{1}{2} \Delta_D + 2\Delta_{AA} \text{ and} \\ H_{ME}(F_2) &= \frac{1}{2} \left[ DE_{hij} - \frac{1}{2}(DE_{hii} + DE_{hjj}) \right] \\ &\quad + 2 \left[ AAE_{hij} - \frac{1}{2}(AAE_{hii} + AAE_{hjj}) \right] \\ &= \frac{1}{2} \Delta_{DE} + 2\Delta_{AAE}. \end{aligned}$$

Similarly, general heterosis and interaction heterosis over mid-parent for the  $F_n$  are:

$$\begin{aligned} H_M(F_n) &= (1/2)^{n-1} \Delta_D + 2\Delta_{AA} \text{ and} \\ H_{ME}(F_n) &= (1/2)^{n-1} \Delta_{DE} + 2\Delta_{AAE}. \end{aligned}$$

The predicted results obtained from the above formulae are the real values of heterosis. Because different traits have different units of measurement, heterosis based on the population mean ( $\mu$ ) can be used for comparing among different traits:

$$\begin{aligned} H_{PM}(F_n) &= (1/\mu)H_M(F_n) = (1/\mu)[(1/2)^{n-1} \Delta_D + 2\Delta_{AA}] \text{ and} \\ H_{PME}(F_n) &= (1/\mu)H_{ME}(F_n) = (1/\mu)[(1/2)^{n-1} \Delta_{DE} + 2\Delta_{AAE}]. \end{aligned}$$

It is thus clear that additive effects are not included in general heterosis over mid-parent of the different generations, but that  $\frac{1}{2}\Delta_D$  is deducted for each succeeding generation. The amount of epistasis heterosis  $\Delta_{AA}$ , however, does not vary with generations. Similarly, interaction heterosis over mid-parent is not affected by additive  $\times$  environment interaction effects. The  $F_n$  generation will be  $\frac{1}{2}\Delta_{DE}$  less than the  $F_{n-1}$  generation.  $2\Delta_{AAE}$  remains constant for all generations.

If we set parent  $i$  as the better one, then the parental genetic difference  $\omega_G$  and parental interaction difference  $\omega_{GE}$  can be defined as:

$$\omega_G = |2(A_i - A_j) + (D_{ii} - D_{jj}) + 4(AA_{ii} - AA_{jj})|,$$

$$\omega_{GE} = |2(AE_{hi} - AE_{hj}) + (DE_{hii} - DE_{hjj}) + 4(AAE_{hii} - AAE_{hjj})|,$$

which stand for the genetic difference and GE interaction difference between parent  $i$  and parent  $j$ , respectively.

General heterosis and interaction heterosis over the better parent for the  $F_1$  are:

$$H_B(F_1) = G(F_{1ij}) - G(P_i)$$

$$= (A_j - A_i) + (D_{ij} - D_{ii}) + (2AA_{ij} + AA_{jj} - 3AA_{ii})$$

$$= \Delta_D + 2\Delta_{AA} - \frac{1}{2}\omega_G$$

$$= H_M(F_1) - \frac{1}{2}\omega_G \text{ and}$$

$$H_{BE}(F_1) = GE(F_{1ij}) - GE(P_i)$$

$$= (AE_{hj} - AE_{hi}) + (DE_{hij} - DE_{hii}) + (2AAE_{hij} + AAE_{hjj} - 3AAE_{hii})$$

$$= \Delta_{DE} + 2\Delta_{AAE} - \frac{1}{2}\omega_{GE}$$

$$= H_{ME}(F_1) - \frac{1}{2}\omega_{GE}.$$

General heterosis and interaction heterosis over the better parent for the  $F_2$  are:

$$H_B(F_2) = (A_j - A_i) + \frac{1}{2} \left[ D_{ij} + \frac{1}{2}D_{jj} - 3(D_{ii}/2) \right]$$

$$+ (2AA_{ij} + AA_{jj} - 3AA_{ii})$$

$$= H_M(F_2) - \frac{1}{2}\omega_G \text{ and}$$

$$H_{BE}(F_2) = (AE_{hj} - AE_{hi}) + \frac{1}{2} \left[ DE_{hij} + \frac{1}{2}DE_{hjj} - 3(DE_{hii}/2) \right]$$

$$+ (2AAE_{hij} + AAE_{hjj} - 3AAE_{hii})$$

$$= H_{ME}(F_2) - \frac{1}{2}\omega_{GE}.$$

Similarly, general heterosis and interaction heterosis over the better parent for the  $F_n$  are:

$$H_B(F_n) = (1/2)^{n-1} \Delta_D + 2\Delta_{AA} - \frac{1}{2}\omega_G$$

$$= H_M(F_n) - \frac{1}{2}\omega_G \text{ and}$$

$$H_{BE}(F_n) = (1/2)^{n-1} \Delta_{DE} + 2\Delta_{AAE} - \frac{1}{2}\omega_{GE}$$

$$= H_{ME}(F_n) - \frac{1}{2}\omega_{GE}.$$

General heterosis and interaction heterosis over the better parent based on the population mean for  $F_n$  can be expressed as:

$$H_{PB}(F_n) = (1/\mu) \left[ H_M(F_n) - \frac{1}{2}\omega_G \right]$$

$$= H_{PM}(F_n) - \frac{1}{2}\delta_G \text{ and}$$

$$H_{PBE}(F_n) = (1/\mu) \left[ H_{ME}(F_n) - \frac{1}{2}\omega_{GE} \right]$$

$$= H_{PME}(F_n) - \frac{1}{2}\delta_{GE},$$

where  $\delta_G = \omega_G/\mu$  and  $\delta_{GE} = \omega_{GE}/\mu$ .

Thus it can be seen that the formulae for predicting heterosis over the better parent are different from those for predicting heterosis over mid-parent. The former is related not only to dominance effects and additive  $\times$  additive effects, but also to additive effects ( $\omega_G$  including the difference of additive effects between the two parents). General heterosis over the better parent for each generation can be expressed as functions of  $\Delta_D$ ,  $\Delta_{AA}$  and  $\omega_G$ , where  $\Delta_{AA}$  and  $\omega_G$  remain unchanged from  $F_1$  to  $F_n$ , and  $\Delta_D$  decreases by half with every selfing generation. In a similar way, interaction heterosis over the better parent is the functions of  $\Delta_{DE}$ ,  $\Delta_{AAE}$  and  $\omega_{GE}$ , and the  $F_n$  generation will decrease by  $\frac{1}{2}\Delta_{DE}$  from the  $F_{n-1}$  generation, but  $\Delta_{AAE}$  and  $\omega_{GE}$  will not change. It should be especially pointed out that a cross will show high heterosis in both  $F_1$  hybrids and their later generations when  $\Delta_{AA}$  is large, even if  $\Delta_D$  is small or less than zero.

If we set heterosis over the better parent based on the population mean  $H_{PB}(F_n)$  equal to a prior value  $\alpha$  (usually 0.10 or 0.05 for positive heterosis), the number of generations of a cross with  $H_{PB}(F_n) = \alpha$  is calculated by

$$n = \{ \log[\Delta_D / (2\mu\alpha - 4\Delta_{AA} + \omega_G)] / \log(2) \} + 2,$$

when  $\Delta_D / (2\mu\alpha - 4\Delta_{AA} + \omega_G)$  is larger than zero.

The jackknife resampling method (Miller, 1974) is appropriate for calculating the predictors of genetic merits and their standard errors for the  $t$ -test (Zhu, 1993; Zhu & Weir, 1996b). The null hypotheses of no

difference between genotypic value and population mean ( $H_0: G(F_n) = 0$ ), and no heterosis ( $H_0: H(F_n) = 0$ ) can both be tested by  $t$ -tests.

### A worked example

Yield data from an experiment on cotton (*Gossypium hirsutum* L.) conducted in two years is used as an example to illustrate the application of the methodology for predicting genotypic value and heterosis. The experiments with a randomized complete block design were carried out in the experimental station of Zhejiang Agricultural University in 1992 and 1993. There were three blocks in each year and 50 plots in each block. Five female parents, (1) A226, (2) A160, (3) A17, (4) Lumian 6 and (5) Zhongmian 12, and five male parents, (6) Zhongmian 13, (7) Xuzhou 184, (8) Shimian 2, (9) 4305 and (10) 4318, were used to produce 20 crosses of  $F_1$ s and  $F_2$ s. Cotton yield (kg/plot), boll number per plant, boll size (g), and lint percentage (%) were investigated. Genetic effects and genotype  $\times$  environment interaction effects were predicted by the AUP method (Zhu, 1993; Zhu & Weir, 1996b).

The average genotypic values ( $\mu + G$ ) and general heterosis of the 20 crosses are listed in Table 1 for yield traits of  $F_1$ s and  $F_2$ s. The totals of genotypic values and general heterosis were less in the  $F_2$  than in the  $F_1$ . The  $F_1$  genotypic value of each trait was not significantly different from the population mean. As far as cotton yield was concerned, general heterosis of the  $F_1$  and  $F_2$  over the mid-parent was 26.9% and 16.2%, respectively, and reached significance. On the other hand, general heterosis of the  $F_1$  over the better parent was 13.5% and significant at the 1% level, and decreased to as low as 2.8% in the  $F_2$ . Boll number per plant showed significant positive  $H_{PM}$  and nonsignificant negative  $H_{PB}$ ; lint percentage had highly significant negative  $H_{PB}$  but nonsignificant  $H_{PM}$ ; both  $H_{PM}$  and  $H_{PB}$  for boll size were not significant at the 5% level.

Predicted interaction heterosis for the two years is presented in Table 2. Heterosis arising from GE interaction was found to varying degrees for different traits. Under two environments there was positive  $H_{PME}$  and negative  $H_{PBE}$  for both cotton yield and boll number per plant;  $H_{PME}$  and  $H_{PBE}$  were positive for boll size but negative for lint percentage with the exception of  $H_{PME}$  ( $F_2$ ) in 1992.

It is important in practice to understand the performance of heterosis of each specific cross. Taking cotton yield as an example, the results of six crosses are given in Table 3. General heterosis over the mid-parent of these crosses, except for cross  $4 \times 10$ , was significant ( $P < 5\%$ ) or highly significant ( $P < 1\%$ ). Significant  $H_{PB}$  of the  $F_1$  was 40.4% for cross  $5 \times 9$  and 24.4% for cross  $5 \times 10$ , and decreased to 9.8% and 6.6% in the  $F_2$ , respectively. This was because both crosses had large  $\Delta_D$  and small  $\Delta_{AA}$  (cross  $5 \times 9$  having a negative  $\Delta_{AA}$  and cross  $5 \times 10$  having only 1.9% of  $\Delta_{AA}$ ). If the positive population heterosis over the better parent were set at 5%, the expected number of generations for  $5 \times 9$  and  $5 \times 10$  would be larger than two. However, there existed large negative heterosis deviations caused by GE interaction. From the sum of  $H_{PB}$  and  $H_{PBE}$ , the final heterosis over the better parent of  $5 \times 9$  in the  $F_2$  was  $-3.1\%$  in 1992 and  $3.4\%$  in 1993; similarly, that of  $5 \times 10$  in the  $F_2$  was  $1.9\%$  in 1992 and  $-6.4\%$  in 1993. Cross  $1 \times 7$  also shared large  $\Delta_D$  and small  $\Delta_{AA}$ , and its expected number of generations was only one.  $H_{PBE}$  ( $F_1$ ) and  $H_{PBE}$  ( $F_2$ ) of cross  $1 \times 7$  were negative in 1992 but positive in 1993. Because both  $\Delta_D$  and  $\Delta_{AA}$  were not large, cross  $4 \times 10$  can not be used even for the  $F_1$  in practice. Heterosis over the better parent of cross  $4 \times 10$  was negative so that zero was used as its expected number of generations.

We did not list the expected number of generations for crosses  $2 \times 9$  and  $3 \times 9$  for the following reasons. Both crosses were rather special, and it was necessary to analyse them in detail. For cross  $3 \times 9$ ,  $H_{PB}$  of the  $n$ th generation

**Table 1** Predicted genotypic values and general heterosis for yield traits†

Trait	Cotton yield		Boll number		Boll size		Lint%	
	Predictor	SE	Predictor	SE	Predictor	SE	Predictor	SE
$\mu + G(F_1)$	1.569	0.044	12.933	0.400	4.849	0.145	37.841	0.646
$\mu + G(F_2)$	1.409*	0.031	12.280	0.274	4.614	0.042	37.432*	0.175
$H_{PM}(F_1)$	0.269**	0.030	0.118*	0.031	0.069	0.055	0.021	0.020
$H_{PM}(F_2)$	0.162**	0.025	0.065**	0.015	0.019	0.022	0.010	0.014
$H_{PB}(F_1)$	0.135*	0.040	-0.012	0.044	0.008	0.052	-0.060*	0.019
$H_{PB}(F_2)$	0.028	0.023	-0.064	0.030	-0.043	0.020	-0.070*	0.020

†Difference between genetic value and population mean and difference between heterosis and zero are compared in deciding on significance, respectively.

\* $P < 0.05$ , \*\* $P < 0.01$ .

**Table 2** Predicted genotype × environment values and interaction heterosis for yield traits in 1992 and 1993

Traits	Genetic parameter	1992		1993	
		Predictor	SE	Predictor	SE
Cotton yield	<i>GE</i> (F <sub>1</sub> )	-0.007	0.050	0.009	0.042
	<i>GE</i> (F <sub>2</sub> )	-0.039	0.052	-0.005	0.045
	H <sub>PME</sub> (F <sub>1</sub> )	0.071*	0.020	0.042*	0.015
	H <sub>PME</sub> (F <sub>2</sub> )	0.049	0.022	0.033	0.016
	H <sub>PBE</sub> (F <sub>1</sub> )	-0.062	0.032	-0.063*	0.023
	H <sub>PBE</sub> (F <sub>2</sub> )	-0.084*	0.031	-0.072*	0.023
Boll number	<i>GE</i> (F <sub>1</sub> )	0.068	0.258	0.101	0.192
	<i>GE</i> (F <sub>2</sub> )	0.006	0.292	-0.156	0.189
	H <sub>PME</sub> (F <sub>1</sub> )	0.009	0.026	0.050	0.026
	H <sub>PME</sub> (F <sub>2</sub> )	0.004	0.028	0.030	0.026
	H <sub>PBE</sub> (F <sub>1</sub> )	-0.092*	0.033	-0.023	0.030
	H <sub>PBE</sub> (F <sub>2</sub> )	-0.097*	0.037	-0.043	0.031
Boll size	<i>GE</i> (F <sub>1</sub> )	0.081	0.061	0.164*	0.062
	<i>GE</i> (F <sub>2</sub> )	-0.068	0.029	-0.038	0.035
	H <sub>PME</sub> (F <sub>1</sub> )	0.063**	0.015	0.072**	0.015
	H <sub>PME</sub> (F <sub>2</sub> )	0.031	0.012	0.028*	0.010
	H <sub>PBE</sub> (F <sub>1</sub> )	0.033	0.016	0.050*	0.015
	H <sub>PBE</sub> (F <sub>2</sub> )	0.001	0.014	0.006	0.011
Lint%	<i>GE</i> (F <sub>1</sub> )	-1.102	0.471	0.832	0.489
	<i>GE</i> (F <sub>2</sub> )	-0.144	0.260	0.595	0.277
	H <sub>PME</sub> (F <sub>1</sub> )	-0.020	0.015	-0.037	0.016
	H <sub>PME</sub> (F <sub>2</sub> )	0.005	0.009	-0.043**	0.009
	H <sub>PBE</sub> (F <sub>1</sub> )	-0.054*	0.017	-0.061*	0.015
	H <sub>PBE</sub> (F <sub>2</sub> )	-0.029*	0.010	-0.067**	0.009

\* *P* < 0.05, \*\* *P* < 0.01.

was calculated as  $\frac{1}{\mu}[(\frac{1}{2})^n \hat{\Delta}_D + 2\hat{\Delta}_{AA} - \frac{1}{2}\hat{\omega}_G] = (\frac{1}{2})^n \times 0.276 + 0.157$ . It is clear that as *n* becomes large enough,  $(\frac{1}{2})^n \hat{\Delta}_D$  approaches zero, and even under this circumstance the cross should still have heterosis of 15.7% over the better parent because of the large value for  $\hat{\Delta}_{AA}$ . Hence, when we predict the expected number of generations for this cross, the lower limit should be approximately larger than 16%. Meanwhile, it was indicated that additive × additive effects, which are the heritable genetic component, play an important role in hybrid breeding. In addition, this cross also maintained a large H<sub>PBE</sub> value in the F<sub>2</sub>, which was 4.6% in 1992 and -18.4% in 1993. Therefore this cross could make full use of yield potential in suitable environments.

Cross 2 × 9 showed the same trend as cross 3 × 9. H<sub>PB</sub> (F<sub>1</sub>) and H<sub>PB</sub> (F<sub>2</sub>) of cross 2 × 9 were 27.2% and 22.7%, respectively, and deviated from the population mean at *P* < 0.05. In further generations, Δ<sub>D</sub> would drop gradually and the smallest value (lower limit) of H<sub>PB</sub> would be  $\frac{1}{\mu}[2\hat{\Delta}_{AA} - \frac{1}{2}\hat{\omega}_G] = 18.2\%$ . Because cross 2 × 9 had similar Δ<sub>AA</sub> and smaller Δ<sub>D</sub> and ω<sub>G</sub> compared with cross 3 × 9, the lower limit of H<sub>PB</sub> of cross 2 × 9

was higher than that of cross 3 × 9. Moreover, the small interaction heterosis of cross 2 × 9 in both 1992 and 1993 indicated that this cross was hardly influenced by environments. From the above analysis we know that a cross might attain a high level of heterosis when Δ<sub>AA</sub> is large enough even though Δ<sub>D</sub> is small.

### Discussion

F<sub>1</sub> hybrids in maize, rice and sorghum have been successfully developed and cultivated. However, primarily because of the difficulty of producing F<sub>1</sub> seed economically, utilization of F<sub>1</sub> heterosis in many other crops has been limited. Significant heterosis of agronomic traits in the F<sub>2</sub> generation has been reported in cotton (Meredith, 1990), winter wheat (Cox & Murphy, 1990), rapeseed (Engqvist & Becker, 1991), sesame (Ding *et al.*, 1987), etc. Meredith (1990) indicated that F<sub>2</sub> hybrids had the genetic potential for increasing cotton (*Gossypium hirsutum*) yields and fibre quality. When compared with the parents in spring oilseed rape (*Brassica napus* L.), the F<sub>2</sub> was 11% higher in yield and

**Table 3** Genotypic values, heterosis and their interaction with environment for cotton yield of six crosses

Crosses	1 × 7	2 × 9	3 × 9	4 × 10	5 × 9	5 × 10
<i>Main genotype effect</i>						
$\mu + G(F_1)$	1.632	1.469	1.711	1.348	1.813*	1.573
$\mu + G(F_2)$	1.462	1.401	1.504	1.324	1.356	1.307
$H_{PM}(F_1)$	0.278*	0.282*	0.473**	0.132	0.453**	0.381**
$H_{PM}(F_2)$	0.165**	0.237*	0.335*	0.117	0.147*	0.203**
$H_{PB}(F_1)$	0.050	0.272*	0.433*	-0.103	0.404*	0.244**
$H_{PB}(F_2)$	-0.063	0.227*	0.295*	-0.119	0.098	0.066
$n$	1.000	—	—	0.000	2.247**	2.132**
$\Delta_D$	0.339	0.136	0.414	0.046	0.916	0.533
$\Delta_{AA}$	0.039	0.144	0.148	0.076	-0.119	0.019
$\omega_G$	0.684	0.030	0.120	0.705	0.147	0.411
$(2\Delta_{AA} - 0.5\omega_G)/\mu$	-0.177	0.182	0.157	-0.134	-0.208	-0.113
<i>Environment 1 (1992)</i>						
$GE(F_1)$	0.300*	-0.339**	0.063	-0.169	-0.101	-0.077
$GE(F_2)$	0.295*	-0.345**	-0.002	-0.204	-0.241	-0.118
$H_{PME}(F_1)$	0.113	0.064	0.189	-0.038	0.072	0.063
$H_{PME}(F_2)$	0.110	0.060	0.146	-0.062	-0.021	0.036
$H_{PBE}(F_1)$	-0.204*	0.021	0.089	-0.162**	-0.035	-0.020
$H_{PBE}(F_2)$	-0.207**	0.017	0.046	-0.185*	-0.129	-0.047
<i>Environment 2 (1993)</i>						
$GE(F_1)$	-0.222	0.344	-0.031	-0.031	0.139	-0.150
$GE(F_2)$	-0.288*	0.369*	-0.028	0.000	0.152*	-0.154*
$H_{PME}(F_1)$	0.033	0.021	0.063	0.134	-0.004	0.007
$H_{PME}(F_2)$	-0.011	0.037	0.065	0.154	0.005	0.004
$H_{PBE}(F_1)$	0.023	-0.023	-0.186*	0.047	-0.073	-0.127
$H_{PBE}(F_2)$	0.609	-0.006	-0.184*	0.068	-0.064	-0.130

\*  $P < 0.05$ , \*\*  $P < 0.01$ .

earlier in flowering time; the descent lines of the  $F_6$  derived from random single seeds had an 8% lower yield and were later flowering (Engqvist & Becker, 1991).

The general formulae proposed in this paper might be useful for predicting heterosis in the offspring of crop hybrids. If we define the lowest positive value of heterosis over the better parent as  $\alpha$  (e.g. 0.10 or 0.05), it is convenient to calculate the number of generations for which offspring of hybrids could maintain 10% or 5% heterosis in practical production. The predicted number of generations has close relationships with  $\Delta_D$ ,  $\Delta_{AA}$  and  $\omega_G$ . The positive heterosis over the better parent will be maintained longer if  $\Delta_D$  and  $\Delta_{AA}$  are larger and  $\omega_G$  smaller. In order to use hybrids for longer, the difference between the two parents should not be too large (a small  $\omega_G$  might cause decreased heterosis of  $F_1$  hybrids). When  $\Delta_D/(2\mu\alpha - 4\Delta_{AA} + \omega_G)$  reaches 1, 2, 4 and 8, respectively, the number of generations for positive heterosis over the better parent will be 2, 3, 4 and 5 accordingly.

Because GE interaction of heterosis is widely observed, it is reasonable to define the total heterosis in a specific environment as the sum of  $H_{PB}$  (or  $H_{PM}$ ) and

$H_{PBE}$  (or  $H_{PME}$ ). Here, heterosis arising from genetic main effects allows estimation of the stability of heterosis across different environments; heterosis arising from GE interaction reflects the deviation of heterosis in specific environments. In general, the genotypic value and heterosis of yield traits will show a declining trend with the increase of generations, but different characters behave differently. When both  $\Delta_D$  and  $\Delta_{AA}$  have large positive values, a cross (e.g. cross 3 × 9) could express high heterosis and maintain it for several generations. When  $\Delta_D$  is small but  $\Delta_{AA}$  is large enough, heterosis of a cross (e.g. 2 × 9) could remain relatively stable over generations; and particularly, when  $\Delta_D$  is less than zero, the upper limit of total heterosis could reach  $(1/\mu)[2(\Delta_{AA} + \Delta_{AAE}) - \frac{1}{2}(\omega_G + \omega_{GE})]$ . When  $\Delta_D$  is large enough but  $\Delta_{AA}$  is small, these kinds of crosses (e.g. crosses 5 × 9 and 5 × 10) could be used mainly in the  $F_1$ , and their offspring would have little genetic merit because heterosis would decline quickly in further generations. Hybrids with small values for both  $\Delta_D$  and  $\Delta_{AA}$  (e.g. cross 4 × 10) are not useful in practice. Furthermore, the above analysis indicates that maintainable heterosis cannot be caused by dominance, but

rather results from epistatic effects like additive  $\times$  additive effects. The additive  $\times$  additive effects play an important role in the utilization of heterosis for  $F_1$  hybrids, and particularly for their later generations.

When the ADAA model is employed to predict heterosis for agronomic traits of crop hybrids, the formulae for predicting heterosis could be extended or reduced with flexibility depending on practical circumstances. If a genetic experiment is conducted only in one environment, and there is no GE interaction, the analysis of interaction heterosis is not necessary. If we do not consider epistatic effects, the prediction formulae for heterosis are the same as those proposed by Zhu (1993, 1997) based on an additive–dominance model. In an analysis of heterosis, if we do not know prior genetic information for some of the traits studied, the ADAA model can be used for an initial analysis. When the variance of additive  $\times$  additive effects is not significant, the data can then be re-analysed by an AD model instead of the ADAA model. The ADAA model can give unbiased estimates of variance components even if these additive  $\times$  additive effects do not exist (Zhu, 1992). In comparison with other genetic components, additive  $\times$  dominance and dominance  $\times$  dominance effects are very complicated, generally negligible and will decline quickly with increasing generations, so the ADAA model does not include additive  $\times$  dominance epistasis and dominance  $\times$  dominance epistasis. If we need to consider AD and DD interactions, the ADAA model should be expanded according to the principles of general genetic models (Cockerham, 1980), and the formulae for predicting heterosis should be derived accordingly. Because all kinds of gene effects are predictable by the AUP or LUP methods (Zhu, 1992, 1993; Zhu & Weir, 1996b), the predicted values of heterosis are also unbiased in theory. In addition, it is effective and feasible for the ADAA model to deal with balanced or unbalanced data from a diallel mating design.

### Acknowledgements

This work was supported in part by ‘The Trans-Century Training Program Foundation for Outstanding Individuals in Science and Technology’ of the State Education Commission of China. The authors are very grateful to Terry Crawford and two anonymous referees for their helpful and valuable suggestions.

### References

- BERNARDO, R. 1992. Relationship between single-cross performance and molecular marker heterozygosity. *Theor. Appl. Genet.*, **83**, 628–634.
- COCKERHAM, C. C. 1980. Random and fixed effects in plant genetics. *Theor. Appl. Genet.*, **56**, 119–131.

- COWEN, N. M. AND FERY, K. J. 1987. Relationships between three measures of genetic distance and breeding behavior in oats (*Avena sativa* L.). *Genome*, **29**, 97–106.
- COX, T. S. AND MURPHY, J. P. 1990. The effect of parental divergence on heterosis in winter wheat crosses. *Theor. Appl. Genet.*, **79**, 241–250.
- DING, F. Y., JIANG, J. P. AND ZHANG, D. X. 1987. A study on  $F_1$ ,  $F_2$  heterosis and the correlation between parents and hybrids in sesame. *Sci. Agr. Sinica*, **20**, 70–76.
- ENGQVIST, G. M. AND BECKER, H. C. 1991. Heterosis and epistasis in rapeseed estimated from generation means. *Euphytica*, **58**, 31–35.
- KETATA, H., SMITH, E. L., EDWARDS, L. H. AND MCNEW, R. W. 1976. Detection of epistatic, additive, and dominance variation in winter wheat (*Triticum aestivum* L. em Thell.). *Crop Sci.*, **16**, 1–4.
- KNIGHT, R. 1973. The relation between hybrid vigour and genotype  $\times$  environment interaction. *Theor. Appl. Genet.*, **43**, 318–322.
- MEREDITH, W. R., JR. 1990. Yield and fiber-quality potential for second-generation cotton hybrids. *Crop Sci.*, **30**, 1045–1048.
- MILLER, R. G. 1974. The jackknife: a review. *Biometrika*, **61**, 1–15.
- MOLL, R. H., SALHUANA, W. S. AND ROBINSON, H. F. 1962. Heterosis and genetic diversity in variety crosses of maize. *Crop Sci.*, **2**, 197–198.
- SCHNELL, F. W. AND COCKERHAM, C. C. 1992. Multiplicative vs. arbitrary gene action in heterosis. *Genetics*, **131**, 461–469.
- SMITH, O. S., SMITH, J. C. S., BOWEN, S. L., TENBORG, R. A. AND WALL, S. J. 1990. Similarities among a group of elite maize inbreds as measured by pedigree,  $F_1$  grain yield, grain yield heterosis, and RFLPs. *Theor. Appl. Genet.*, **80**, 833–840.
- STUBER, C. W., LINCOLN, S. E., WOLFF, D. W., HELENTJARIS, T. AND LANDER, E. S. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics*, **132**, 823–839.
- VIRMANI, S. S., AQUINO, R. C. AND KHUSH, G. S. 1982. Heterosis breeding in rice (*Oryza sativa* L.) *Theor. Appl. Genet.*, **63**, 373–380.
- YOUNG, J. B. AND VIRMANI, S. S. 1990. Heterosis in rice over environments. *Euphytica*, **51**, 87–93.
- ZHANG, Q., GAO, Y. J., YANG, S. H., RAGAB, R. A., SAGHAI MAROOF, M. A. AND LI, Z. B. 1994. A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. *Theor. Appl. Genet.*, **89**, 185–192.
- ZHU, J. 1989. *Estimation of Genetic Variance Components in the General Mixed Model*. Ph.D. Dissertation, North Carolina State University, Raleigh, NC.
- ZHU, J. 1992. Mixed model approaches for estimating genetic variances and covariances. *J. Biomath.*, **7**, 1–11.
- ZHU, J. 1993. Methods of predicting genotype value and heterosis for offspring of hybrids. *J. Biomath.*, **8**, 32–44.
- ZHU, J. 1997. *Analysis Methods for Genetic Models*. Agricultural Publishing House of China, Peking.
- ZHU, J. AND WEIR, B. S. 1996a. Mixed model approaches for diallel analysis based on a bio-model. *Genet. Res.*, **68**, 233–240.
- ZHU, J. AND WEIR, B. S. 1996b. Diallel analysis for sex-linked and maternal effects. *Theor. Appl. Genet.*, **92**, 1–9.
- ZHU, J., JI, D. F. AND XU, F. H. 1993. A genetic approach for analyzing intra-cultivar heterosis in crops. *Acta Genetica Sinica*, **20**, 262–271.