

Prediction of maize double-cross hybrids using the best linear unbiased prediction with microsatellite marker information

M. Balestre¹, R.G. Von Pinho² and J.C. Souza¹

¹Departamento de Biologia, Universidade Federal de Lavras, Lavras, MG, Brasil

²Departamento de Agricultura, Universidade Federal de Lavras, Lavras, MG, Brasil

Corresponding author: M. Balestre
E-mail: marciobalestre@hotmail.com

Genet. Mol. Res. 10 (1): 25-35 (2011)
Received December 5, 2010
Accepted December 21, 2010
Published January 11, 2011
DOI 10.4238/vol10-1gmr985

ABSTRACT. We examined the usefulness of the best linear unbiased prediction associated with molecular markers for prediction of untested maize double-cross hybrids. Ten single-cross hybrids from different commercial backgrounds were crossed using a complete diallel design. These 10 single-cross hybrids were genotyped with 20 microsatellite markers. The best linear unbiased prediction associated with microsatellite information gave relatively good prediction ability of the double-cross hybrid performance, with correlations between observed phenotypic values and genotypic prediction values varying from 0.27 to 0.54. Taking into account the predictions of specific combining ability, the correlation between observed and predicted specific combining ability varied from 0.50 to 0.88. Based on these results, we infer that it is feasible to predict maize double-cross hybrids with different unbalance degrees without including any prior information about parental inbreed lines or single-cross hybrid performance.

Key words: BLUP; Similarity-by-descent; Specific combining ability; Molecular markers

INTRODUCTION

Double-cross hybrids can be considered the first commercial maize hybrid cultivar available to farmers. This fact was possible due to work developed by Jones in the early part of 20th century (Troyer, 2009). In Brazil, current studies indicate that roughly 20% of the seed market involves double-cross hybrid cultivars (Cruz and Pereira Filho, 2008). Various seed companies maintain this kind of cultivar in order to exploit the heterosis that exists among pairs of single-cross hybrids, mainly those with low heterosis *per se* (Jenkins, 1934).

Synthesis of double-cross hybrids depends on the number of single-cross hybrids involved in the crosses; it can become unfeasible due to an increase in the number of single-cross hybrids used as genitors. For instance, starting from 10 inbred lines the breeder might obtain 630 double-cross hybrids. Evaluation of all of them becomes impracticable; the cost of evaluation increases greatly when there are more hybrids. It is more efficient to evaluate only some genotypes and predict the other crosses.

In order to predict maize double-cross hybrid grain yields, Jenkins (1934) suggested four prediction models called Jenkins' methods A, B, C, and D. They can be used to predict performance of double-cross hybrids based on non-parental single-cross hybrids performance in order to evaluate only high-performance double-cross hybrids.

Although the prediction methods proposed by Jenkins are widely used, they have an important limitation, the need to evaluate all possible single-cross hybrids in diallel crosses. However, this situation might not occur, since the breeders regularly make use of incomplete diallel crosses; therefore, the number of single-crosses could be reduced for double-cross prediction, making the Jenkins method ineffective. An alternative could be to predict double-cross hybrids based on pedigree information; if it is possible to determine covariance measures or genetic similarity among double-cross hybrids, the information on inbred lines or parental single-cross performance is no longer necessary.

Bernardo (1994) proposed a method for maize single-cross prediction taking into account the genealogy estimated by molecular markers plus the best linear unbiased prediction (BLUP). He considered that, starting from a number of evaluated hybrids and prior genealogy information concerning these hybrids, it is possible to make measures of covariance among tested and untested hybrids based on probability of alleles inherited from common genitors (similarity-by-descent).

The efficiency of this approach has been demonstrated in several studies of maize (Bernardo, 1994, 1995, 1996; Schrag et al., 2009; Balestre et al., 2010). In all these studies, a relatively good correlation between phenotypic means and predicted means was observed, demonstrating acceptable accuracy. The prediction accuracy is dependent on the heritability of traits in the tested hybrids, adequate estimates of variance components, accurate approximation of relationship coefficients by molecular markers (Balestre et al., 2009), and number of predictor parentals in the analysis (Vuylsteke et al., 2000; Schrag et al., 2006). Some authors suggest the use of molecular marker information associated with similarity-by-state coefficients for genotypic value prediction (Bauer et al., 2006; Balestre et al., 2008). However, Balestre et al. (2009, 2010) observed problems with using similarity-by-state coefficients for prediction of genotypic values, mainly because information concerning unrelated genotypes is included in the relationship matrix, causing bias in the prediction of genotypic values.

As an alternative, Balestre et al. (2009, 2010) suggest the use of Lynch and Ritland's (1999) coefficient, which is a coefficient of similarity-by-descent obtained by multiallelic mo-

lecular markers. The advantage of this coefficient is its flexibility, as it can be used in diallel or complex pedigrees when genealogical data are unavailable.

In an attempt to refine this methodology, we evaluated the efficiency of the BLUP with molecular markers in order to obtain estimates of prediction of untested double-cross maize hybrids.

MATERIAL AND METHODS

Ten single-cross hybrids from different backgrounds were used. Starting with these hybrids, a complete diallel was performed to obtain 45 double-cross hybrids, which were assessed in a randomized complete block design with three repetitions. The treatments were assessed in 15 environments distributed in the States of Minas Gerais, Bahia, and Goiás, which correspond to the Southeast, Northeast and Center-West areas of Brazil. The plots consisted of two 4-m rows, with a population density of 55,000 plants per hectare.

Each single-cross hybrid was genotyped with 20 microsatellite markers, with nine linked to quantitative trait loci for components of grain yield in tropical maize (Balestre et al., 2008). Starting from the plots dataset, diallel analysis was conducted using the IV method, proposed by Griffing (1956). Analyses were performed using the SAS® System Software, IML Module (SAS Institute, 2000). Estimates of fixed effects and BLUPs for general combining ability (GCA) and specific combining ability (SCA) were obtained in a manner similar to that proposed by Bernardo (1995) for partial diallels.

The linear model considered was the following:

$$y = X\beta + Z_1g + Z_2s + Z_3\zeta + Z_4\delta + \eta \tag{Equation 1}$$

where y is the vector of the plot means of the hybrid combinations; β is the vector of fixed effects (block within sites, general and local mean); g is the vector of effects of GCA of the single-cross hybrids; s is the vector of SCA of the hybrids, and X , Z_1 and Z_2 are the incidence matrices of the effects β , g and s , respectively. The residues vector can be decomposed into $e = \zeta + \delta + \eta$, where ζ , δ and η are the interaction effects of GCA by environments, SCA by environments and experimental error, respectively.

The joint solution for fixed and random effects was obtained by the following system of equations, conforming to Henderson (1984):

$$\begin{bmatrix} \beta^0 \\ \hat{g} \\ \hat{s} \\ \xi \\ \hat{\delta} \end{bmatrix} = \begin{bmatrix} X'X & X'Z_1 & X'Z_2 & X'Z_3 & X'Z_4 \\ Z_1'X & Z_1'Z_1 + A_1^{-1}\gamma_1 & Z_1'Z_2 & Z_1'Z_3 & Z_1'Z_4 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + A_2^{-1}\gamma_2 & Z_2'Z_3 & Z_2'Z_4 \\ Z_3'X & Z_3'Z_1 & Z_3'Z_2 & Z_3'Z_3 + I_1\gamma_3 & Z_3'Z_4 \\ Z_4'X & Z_4'Z_1 & Z_4'Z_2 & Z_4'Z_3 & Z_4'Z_4 + I_2\gamma_4 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \\ Z_3'y \\ Z_4'y \end{bmatrix} \tag{Equation 2}$$

$$\gamma_1 = \frac{\sigma_e^2}{\sigma_{GCA}^2} \quad \gamma_2 = \frac{\sigma_e^2}{\sigma_{SCA}^2} \quad \gamma_3 = \frac{\sigma_e^2}{\sigma_{GCAE}^2} \quad \gamma_4 = \frac{\sigma_e^2}{\sigma_{SCAIE}^2}$$

where A_1 is the matrix of additive genetic similarity among the partial inbred lines.

The additive relationship matrix was obtained according to Lynch and Ritland (1999) by the following expression.

$$\hat{r}_{xy(k)} = \frac{p_a(S_{bc} + S_{bd}) + p_b(S_{ac} + S_{ad}) - 4p_a p_b}{(1 + S_{ab})(p_a + p_b) + 2p_a p_b} \quad \text{(Equation 3)}$$

where $\hat{r}_{xy(k)}$ is the additive relationship estimate between individual X (taken as a reference) with alleles *a* and *b* and individual Y with alleles *c* and *d* in locus *k*; S_{ab} : assigned 1 if *a* and *b* are identical and 0 otherwise; S_{ac} : assigned 1 if *a* and *c* are identical and 0 otherwise; S_{ad} : assigned 1 if *a* and *d* are identical and 0 otherwise; S_{bc} : assigned 1 if *b* and *c* are identical and 0 otherwise; S_{bd} : assigned 1 if *b* and *d* are identical and 0 otherwise; p_a and p_b are the frequencies of alleles *a* and *b* along the lines for a given locus *k*.

This coefficient was derived from Lynch and Ritland (1999) in order to obtain pairwise relatedness between two individuals considering molecular markers with two alleles per locus. This is a common condition in single-cross hybrids derived from inbred lines, as employed in this study.

Considering that the relationship estimates are obtained for many loci, the use of weights for the estimates is suggested in order to reduce sample variance that can arise from differences in reference genotypes and in levels of variation (Lynch and Ritland, 1999). In addition, these authors propose reciprocal estimation of $\hat{r}_{xy(k)}$, that is, initially the individual X is taken as a reference and then the individual Y is used (Lynch and Ritland, 1999).

The multilocus expression, taking into account all loci, the weights attributed to each locus, and reciprocal relationship estimates, is given by:

$$\hat{r}_{xy} = \frac{\sum_1^L w_{r,x(k)} \hat{r}_{x,y(k)} + \sum_1^L w_{r,y(k)} \hat{r}_{y,x(k)}}{W_{r,x} + W_{r,y}} \quad \text{(Equation 4)}$$

with

$$w_{r,x(k)} = \left(\frac{1}{\text{Var}[\hat{r}_{x,y(k)}]} \right) = \frac{(1 + S_{ab})(p_a + p_b) - 4p_a p_b}{2p_a p_b} \quad \text{(Equation 5)}$$

where $w_{r,x(k)}$ and $w_{r,y(k)}$ are the weights for the *k*th locus of the estimate of $\hat{r}_{xy(k)}$ and reciprocal estimates and $W_{r,x}$ and $W_{r,y}$ the sum of the weights attributed to all loci.

The dominance relationship matrix (A_2) was constructed according to Henderson (1984) and Van Vleck (1993). The estimates of variance components were obtained by restricted maximum likelihood by expectancy maximization algorithm (EM-REML).

The prediction of the genotypic values and the SCA of the unanalyzed crosses were made based on simulated unbalance or cross-validation. Thus, of a total of 45 double-cross hybrids analyzed, six situations were simulated considering random unbalance of 5, 10, 15, 20, 25, and 30 hybrids. For each situation, unrestricted unbalance was used, repeating each process 2000 times; i.e., in some cases all 10 genitors could be among the predictor hybrids, in others cases only six.

Since the information on parental inbred lines of the 10 single-cross hybrids is un-

known for use in double-cross predictions, it becomes necessary to develop a mathematical function minimizing the sum square of unobserved genotypic values and future genotypic values that can be predicted.

For the joint probability vector (θ, \hat{y}_p) , where θ is the unobserved genotypic values of untested hybrids and \hat{y}_p is the genotypic values of the evaluated double-cross hybrids corrected by fixed effects, given by $\hat{y}_p = (Z'R^{-1}Z)^{-1}Z'R^{-1}(y-X\beta)$, the function $t(\hat{y}_p)$ that minimizes the sum square deviations $E[\theta - t(\hat{y}_p)]^2$ is given by:

$$E[\theta - t(\hat{y}_p)]^2 = \iint [\theta - t(\hat{y}_p)]^2 g(\theta, \hat{y}_p) d\theta d\hat{y}_p \quad \text{(Equation 6)}$$

$$E[\theta - t(\hat{y}_p)]^2 = \iint [\theta - t(\hat{y}_p)]^2 f(\theta | \hat{y}_p) h(\hat{y}_p) d\theta d\hat{y}_p \quad \text{(Equation 7)}$$

Isolating the marginal function $h(\hat{y}_p)$ and deriving in relation to $t(\hat{y}_p)$, we obtain:

$$2 \int h(\hat{y}_p) [E(\theta | \hat{y}_p) - t(\hat{y}_p)] d\hat{y}_p = 0 \quad \text{(Equation 8)}$$

Thus, this function will give the minor sum square if $t(\hat{y}_p) = E(\theta | \hat{y}_p)$. Considering that the genotypic values of evaluate hybrids presented a multivariate normal distribution with unobserved genotypic values of unevaluated hybrids, it might be assumed that $E(\theta | \hat{y}_p)$ is a linear function of \hat{y}_p . Therefore, the function of $t(\hat{y}_p)$ minimizing $E[\theta - t(\hat{y}_p)]^2$ can be given by:

$$t(\hat{y}_p) = \beta_0 + \beta_1 \hat{y}_p \quad \text{(Equation 9)}$$

Deriving $E[\theta - \beta_0 + \beta_1 \hat{y}_p]^2$ in relation to β_0 and β_1 , and making it equal to zero one has:

$$\beta_0 = \mu_\theta - C_{\theta\hat{y}_p} V_{\hat{y}_p}^{-1} \mu_{\hat{y}_p} \quad \text{(Equation 10)}$$

$$\beta_1 = C_{\theta\hat{y}_p} V_{\hat{y}_p}^{-1} \quad \text{(Equation 11)}$$

which can be written as: $t(\hat{y}_p) = \mu_\theta + C_{\theta\hat{y}_p} V_{\hat{y}_p}^{-1} (\hat{y}_p - \mu_{\hat{y}_p})$.

Since the expectations of μ_θ and $\mu_{\hat{y}_p}$ are equal to zero because the corrected means of genotypic values are taking as deviates from fixed effects, the BLUP of genotypic values of untested hybrids is given by:

$$\hat{y}_{nt} = C_{\theta\hat{y}_p} V_{\hat{y}_p}^{-1} \hat{y}_p \quad \text{(Equation 12)}$$

in which \hat{y}_{nt} is the vector of the hybrid grain yield with simulated loss; $C_{\theta\hat{y}_p}$ is the matrix of genetic covariance between the crosses with simulated loss and the analyzed crosses, and

$V_{\hat{y}_p}^{-1}$ is the inverted matrix of phenotypic covariance between the analyzed crosses.

Similarly, the SCA values of the crosses with simulated loss were predicted using the following expression:

$$d_{\pi} = SV^{-1}\hat{s}_p \quad (\text{Equation 13})$$

in which d_{π} is the vector of the SCA of the crosses with simulated loss; S is the matrix of genetic covariance between the crosses with simulated loss and the analyzed crosses; V^{-1} is the inverted matrix of genetic variance and covariance between the analyzed crosses, and \hat{s}_p is the vector of the SCA of the analyzed hybrids.

The efficiency of BLUP for predicting the values of SCA and the genotypic values was assessed by the magnitude of the correlation values. In addition, selection efficiency of Hamblin and Zimmermann (1986) was also applied, comparing the nine largest predicted values of SCA and genotypic values, with the nine largest values of SCA and observed yield. Means of all genotypic values of the untested hybrids were considered in all combinations and conditions of unbalance.

RESULTS

The estimates of similarity-by-descent shown in Table 1 demonstrate that the 10 single-cross hybrids used as genitors probably came from different backgrounds; i.e., the single-cross hybrids genotyped by microsatellite markers presented only 10 pairs of relationship measures of 45 possible estimates. Furthermore, SCA variance was roughly five times greater than GCA variance. The heritability in the wide sense was high, reflecting good accuracy of genotypic values of the hybrids (Table 1).

Table 1. Similarity-by-descent estimates obtained among 10 commercial single-cross hybrids by microsatellite markers.

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
H1	1.000	0.000	0.000	0.136	0.000	0.000	0.000	0.000	0.000	0.009
H2		1.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000
H3			1.000	0.000	0.023	0.000	0.000	0.000	0.000	0.000
H4				1.000	0.000	0.000	0.000	0.000	0.000	0.056
H5					1.000	0.000	0.000	0.000	0.081	0.000
H6						1.000	0.000	0.054	0.083	0.000
H7		symmetric					1.000	0.000	0.000	0.036
H8								1.000	0.004	0.000
H9									1.000	0.000
H10										1.000
σ_{GCA}^2	0.049									
σ_{SCA}^2	0.223									
σ_e^2	0.816									
h^2	0.937									

Numbers in bold mean non-zero similarity-by-decent estimates.

The mean of the correlations between observed phenotypic values and predicted values was low independent of unbalance degree, varying from 0.30 to 0.38 (Table 2). Identical inferences can be made considering the median of the correlation probability distribution obtained with 2000 cycles of cross-validation; it varied from 0.30 to 0.39. Furthermore, it was observed that the standard deviation was smaller when there was a low degree of unbalance degree.

Table 2. Parameters related to correlation values between observed and predicted performance of double-cross hybrids obtained along 2000 simulations under different unbalance levels.

Parameters	Number of unbalanced hybrids					
	Grain yield					
	5	10	15	20	25	30
Mean	0.38	0.36	0.36	0.34	0.32	0.30
Median	0.39	0.39	0.37	0.35	0.33	0.30
Standard deviation	0.39	0.24	0.18	0.18	0.18	0.17
Skewness	-0.87	-0.55	-0.51	-0.21	-0.16	-0.06
Mode	0.54	0.46	0.39	0.40	0.36	0.27

Parameters	Specific combining ability					
	5	10	15	20	25	30
Mean	0.54	0.56	0.54	0.51	0.45	0.39
Median	0.63	0.60	0.57	0.53	0.49	0.44
Standard deviation	0.37	0.23	0.18	0.15	0.16	0.18
Skewness	-1.15	-0.84	-0.84	-0.75	-1.07	-0.85
Mode	0.88	0.71	0.61	0.55	0.52	0.50

During the cross-validation process, high skewness of the probability distribution or sampling (normality rejected by the Shapiro-Wilk test) was observed, suggesting that the mean and median might not be a good measure of position, since the most plausible values are away from these estimators; i.e., the mean and median might not reflect the values of maximum probability when a high degree of skewness occurs in the distribution due to outliers. Wright et al. (2000) argued that under these circumstances the mode is the most recommended measure of position when the aim of the researcher is to seek the most probable values in the frequency distribution. The sampling densities and the mode of the distributions were determined in order to obtain the most probable values of correlation during the cross-validation process. Taking into account the mode of the distributions, the correlations between untested and tested hybrids ranged from 0.27 to 0.54 (Figure 1).

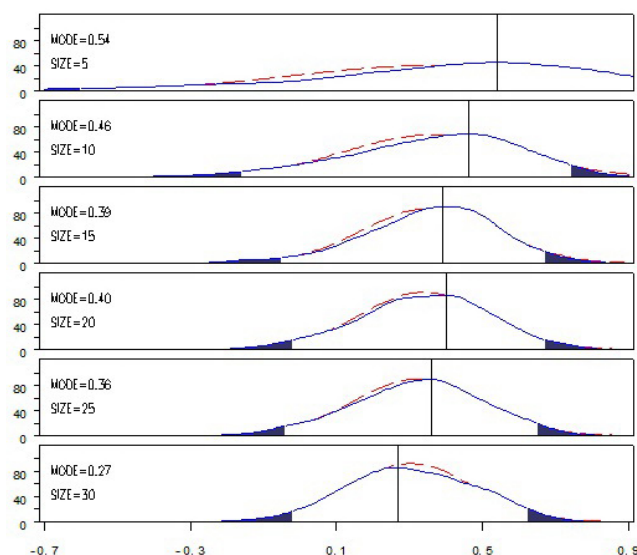


Figure 1. Probability distribution of correlation values between observed phenotypic means and predicted means obtained along 2000 simulations in maize double-cross hybrids.

The empirical correlations between predicted and observed values for SCA can be considered high ($r = 0.88$) when using five unbalanced hybrids (Table 2). A high degree of skewness of the sampling probability distribution obtained during cross-validation process for SCA prediction was observed (Figure 2); consequently, the mean and median might not describe the most probable values obtained from 2000 simulations. The correlation values for SCA predictions varied from 0.50 to 0.88. In addition, the correlation magnitude varied according to the level of unbalance. The mean and median for SCA predictions ranged from 0.39 to 0.54 and 0.44 to 0.63, respectively, within 2000 simulations.

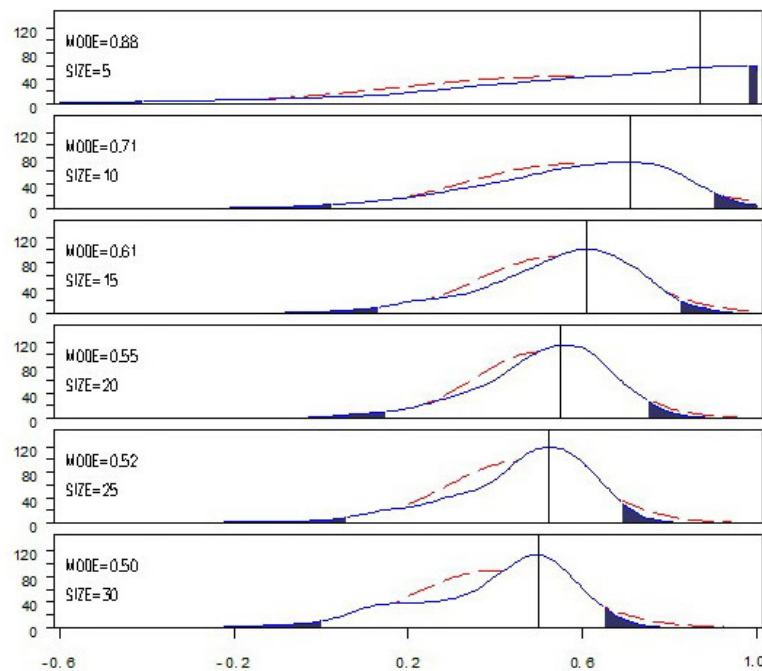


Figure 2. Probability distribution of correlation values between observed specific combining ability and predicted specific combining ability obtained along 2000 simulations in maize double-cross hybrids.

The double-cross hybrids with the greatest predicted means obtained throughout the cross-validation process were coincident with the superior observed hybrids considering a selection intensity of 20% (Figure 3); i.e., among nine predicted hybrid candidates for selection, four were coincident with the nine best hybrids, including the first and second most productive hybrids (36 and 5). This finding supports Bernardo's (1996) argument about magnitude of correlation in the prediction process and probability of selection of the best genotype. The best hybrids in this case would be selected, even with 31% of selection efficiency, according to Hamblin and Zimmermann's (1986) coefficient. In addition, it was noted that among the selected hybrids, the observed phenotypic means were outside of the confidence interval, considering a 5% probability. An exception was noted for hybrid 14; the predicted mean was identical to the observed mean, demonstrating that it is possible to obtain predictions very close to the true values.

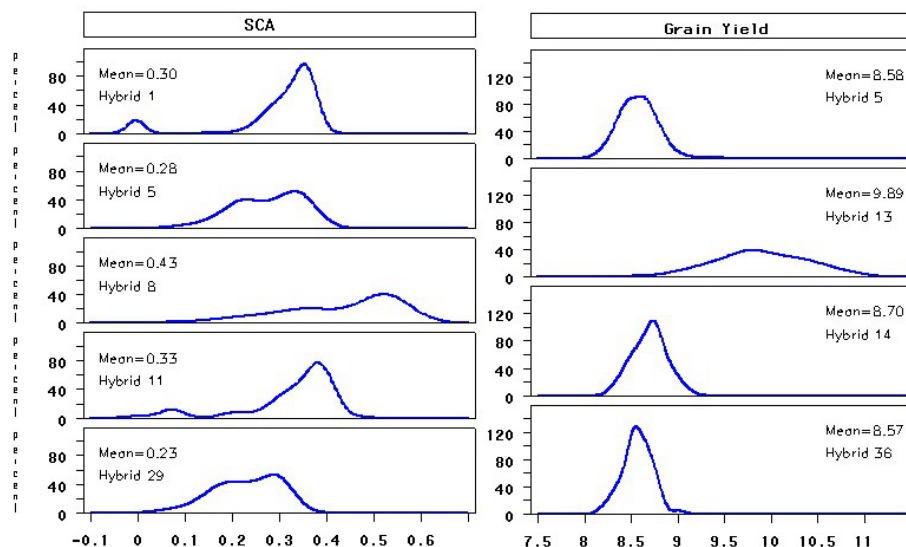


Figure 3. Probability distribution of the predicted values for the unbalanced double-cross hybrids obtained along 2000 simulations and coincident with the nine larger values of grain yield and specific combining ability (SCA) considering 20% selection intensity.

The hybrids with the greatest predicted SCA obtained during the cross-validation process and that were coincident with the highest observed SCA are shown in Figure 3. Although the selection efficiency was 44% for these hybrids, the highest SCA was not observed among them; the SCA between single-crosses SC6 and SC7, which gave rise to double-cross 36, was not among the selected hybrids. However, the second most productive hybrid (DC5) was included among the hybrid candidates for evaluation. Two of five hybrids had observed SCAs outside the predicted confidence interval.

DISCUSSION

Schrag et al. (2009) using several prediction methods, including what we used here, observed that predictions explained from 16 to 46% of the genetic variance. We also found that taking into account the genetic variance, the predictions explained from 8 to 31% of the genetic variance; in some cases, the R^2 was larger than that obtained by Schrag et al. (2009), using BLUP + pedigree + phenotypic values ($R^2 = 24\%$).

High prediction efficiency for SCA was observed mainly when a low level of unbalance was applied ($R^2 = 77.4\%$). Consequently, the prediction method proposed by Bernardo (1994) could be also applied to double-cross maize hybrids even if there is no prior information about inbred lines or single-cross hybrid performance. This method presented satisfactory prediction ability based merely on the double-cross hybrid performance and measures of similarity-by-descent obtained with microsatellite markers.

In the prediction process, the breeder seeks prediction values with maximum likelihood of future values, namely, those with the greatest probability of occurrence. The probability distribution of futures values can be obtained through a cross-validation process, as we

demonstrated here. Throughout this process, we observed that the expected values were not coincident with the most probable future values due to skewness of the probability distribution. In these cases, some authors have proposed the use of medians or modes of the frequency distribution. According to Dalenius (1965) the mode is applied in some situations, such as the measure of position in asymmetric distributions, in order to express the most probable value of a probability distribution, especially when the goal of the inference is related to the prediction of future values. Likewise, Bickel (2002, 2003) argues that the mode is more robust than the mean and the median in cases of asymmetric probability distributions. Since in hybrid prediction, one of the breeder's objectives is to suggest genotype candidates for future evaluations, the use of the mode for a one-modal frequency distribution might be more reliable.

When we look at the number of parental predictors, neither restriction was applied during the cross-validation process; in some situations the whole set of 10 single-cross hybrids was used as a parental predictor and in others only six single-cross hybrids were used as predictors, mainly when 30 double-cross hybrids were unbalanced. This kind of simulation could reflect real situations of random unbalance obtained in field crosses. Nonetheless, some recent papers have proposed using more predictor parentals in the cross-validation process in order to increase the prediction efficiency (Vuylsteke et al., 2000; Schrag et al., 2006). Although this procedure increases the prediction efficiency, it does not resolve the breeder's need to predict non-evaluated crosses even when their genitors are not among the set of appraised hybrids. Schrag et al. (2009) observed that the predicted difference between considering all inbred lines or not in the cross-validation process was marginal.

In our study, the correlation between predicted and observed means was similar to that reported by Jenkins (1934) ($r = 0.61-0.76$). However, the simulation was more restrictive than that used by Jenkins. We utilized a set of disconnected commercial single-cross hybrids derived from different backgrounds and employed with an unbalanced level of up to 67%. Even under these circumstances, the method proposed by Bernardo (1994) was robust in the prediction of double-cross hybrids mainly for specific combining ability predictions. We conclude that it is possible to predict double-cross hybrids performance under different levels of unbalance without any prior information about parental inbred lines or single-cross hybrid performance.

REFERENCES

- Balestre M, Von Pinho RG, Souza JC and Machado JC (2008). Potential of maize single-cross hybrids for extraction of inbred lines using the mean components and mixed models with microsatellite marker information. *Genet. Mol. Res.* 7: 1106-1118.
- Balestre M, Von Pinho RG, Souza JC and Oliveira RL (2009). Potential use of molecular markers for prediction of genotypic values in hybrid maize performance. *Genet. Mol. Res.* 8: 1292-1306.
- Balestre M, Von Pinho RG and Souza JC (2010). Prediction of maize single-cross performance by mixed linear models with microsatellite marker information. *Genet. Mol. Res.* 9: 1054-1068.
- Bauer AM, Reetz TC and Leon J (2006). Estimation of breeding values of inbred lines using best linear unbiased prediction (BLUP) and genetic similarities. *Crop Sci.* 46: 2685-2691.
- Bernardo R (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Sci.* 34: 20-25.
- Bernardo R (1995). Genetic models for predicting maize single-cross performance in unbalanced yield trial data. *Crop Sci.* 35: 141-147.
- Bernardo R (1996). Best linear unbiased prediction of maize single-cross performance. *Crop Sci.* 36: 50-56.
- Bickel DR (2002). Robust estimators of the mode and skewness of continuous data. *Comput. Stat. Data Anal.* 39: 153-163.

- Bickel DR (2003). Robust and efficient estimation of the mode of continuous data: the mode as a viable measure of central tendency. *J. Stat. Comput. Simulat.* 73: 899-912.
- Cruz JC and Pereira Filho IA (2008). Cultivares de Milho Disponíveis no Mercado de Sementes do Brasil para a Safra 2008/2009. Available at [<http://www.cnpms.embrapa.br/publicacoes/milho/cultivares.htm>]. Accessed November 1, 2009.
- Dalenius T (1965). The mode - a neglected statistical parameter. *J. R. Stat. Soc. Ser. A* 128: 110-117.
- Griffing B (1956). Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* 9: 463-493.
- Hamblin J and Zimmermann MJ (1986). Breeding common bean for yield in mixtures. *Plant Breed. Rev.* 4: 245-272.
- Henderson CR (1984). Applications of Linear Models in Animal Breeding. University of Guelph Press, Guelph.
- Jenkins MT (1934). Methods of estimating the performance of double crosses in corn. *J. Am. Soc. Agron.* 26: 199-204.
- Lynch M and Ritland K (1999). Estimation of pairwise relatedness with molecular markers. *Genetics* 152: 1753-1766.
- SAS Institute (2000). User's Guide. Version 8. SAS Institute Inc, Cary.
- Schrag TA, Melchinger AE, Sorensen AP and Frisch M (2006). Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. *Theor. Appl. Genet.* 113: 1037-1047.
- Schrag TA, Mohring J, Maurer HP, Dhillon BS, et al. (2009). Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. *Theor. Appl. Genet.* 118: 741-751.
- Troyer AF (2009). Development of Hybrid Corn and the Seed Corn Industry. In: Handbook of Maize: Genetics and Genomics (Bennetzen JL and Hake S, eds.). Springer, New York, 87-114.
- Van Vleck LD (1993). Selection Index in Introduction to Mixed Models Methods for Genetic Improvement of Animals. The Green Book Press, Boca Raton.
- Vuylsteke M, Kuiper M and Stam P (2000). Chromosomal regions involved in hybrid performance and heterosis: their AFLP(R)-based identification and practical use in prediction models. *Heredity* 85 (Pt 3): 208-218.
- Wright DR, Stern HS and Berger J (2000). Comparing traditional and Bayesian analyses of selection experiments in animal breeding. *J. Agric. Biol. Environ. Stat.* 5: 240-256.