Prediction of maize single-cross performance by mixed linear models with microsatellite marker information

M. Balestre[1], R.G. Von Pinho[2] and J.C. Souza[1]

[1]Departamento de Biologia, Universidade Federal de Lavras, Lavras, MG, Brasil
[2]Departamento de Agricultura, Universidade Federal de Lavras, Lavras, MG, Brasil

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

ABSTRACT. We evaluated the potential of the best linear unbiased predictor (BLUP) along with the relationship coefficient for predicting the performance of untested maize single-cross hybrids. Ninety $S_{0:2}$ progenies arising from three single-cross hybrids were used. The 90 progenies were genotyped with 25 microsatellite markers, with nine markers linked to quantitative trait loci for grain yield. Based on genetic similarities, 17 partial inbred lines were selected and crossed in a partial diallel design. Similarity and relationship coefficients were used to construct the additive and dominance genetic matrices; along with BLUP, they provided predictions for untested single-crosses. Five degrees of imbalance were simulated (5, 10, 20, 30, and 40 hybrids). The correlation values between the predicted genotypic values and the observed phenotypic means varied from 0.55 to 0.70, depending on the degree of imbalance. A similar result was observed for the specific combining ability predictions; they varied from 0.61 to 0.70. It was also found that the relationship coefficient based on BLUP provided more accurate predictions than similarity-in-state predictions. We conclude

that BLUP methodology is a viable alternative for the prediction of untested crosses in early progenies.

Key words: BLUP; Molecular markers; Similarity-in-state; Untested hybrids

INTRODUCTION

In maize breeding programs, diallel cross designs are useful in selecting genotypes with good combining ability, since, in diallel combinations, it is possible to breakdown the cross values into general combining ability (GCA) and specific combining ability (SCA) (Griffing, 1956). GCA and SCA values support the breeder in the choice of genitors, given that these parameters are linked to the additive and dominant effects and the selection of elite lines for superior hybrid production, exploiting heterosis in an effective way (Vencovsky and Barriga, 1992).

However, due to the great number of crosses that can be obtained with few genotypes, diallel cross design is little used in the early stages of breeding programs. An alternative could be the use of partial diallels and partial circular diallel design (Kempthorne and Curnow, 1961; Ramalho et al., 1993). However, the restriction of these diallel designs results from the failure of SCA information due to missing crosses. To avoid that problem, the strategy usually employed is the prediction of the crosses based just on the GCA genitor information. Nevertheless, when the dominance effects are higher than the additive effects, that method becomes inefficient (Cruz et al., 1997).

Thus, some methodologies have been proposed for the prediction of untested crosses, i.e., the use of the genetic distance by molecular markers in the prediction of SCA and heterosis values (Melchinger, 1999; Reif et al., 2003; Balestre et al., 2008a). However, results have demonstrated little efficacy and repeatability (Dias et al., 2004).

An alternative would be the use of the best linear unbiased predictor (BLUP) for the prediction of missing crosses (Bernardo, 1994; Charcosset et al., 1998; Piepho et al., 2008; Schrag et al., 2009). For that, the availability of pedigree data or the relationship information between tested and untested genotypes is necessary (Bernardo, 1994). However, the breeder does not always have such information, which could limit the use of this methodology.

To get around this limitation, some authors have suggested the use of similarity-in-state by molecular markers, when pedigree data are unavailable (Bernardo, 1994, 1995; André, 1999; Balestre et al., 2008c). According to Nejati-Javaremi et al. (1997), similarity-in-state has the "advantage" of considering the entire allelic similarity existing among the assessed individuals. Identity by state has been defined as allelic similarity due to a locus transmitted without considering a genealogic history, just taking into account whether two loci have the same sequence and same length, i.e., if the alleles are distinguished by their sequence and if the alleles are distinguished by the number of repeats of a microsatellite (Rousset, 2007). Thus, all coefficients that take into account just co-occurrence of band (raw molecular data) are labeled a similarity-in-state coefficient.

Bernardo (1994) employed a coancestry coefficient in order to correct for possible bias that can arise from genetic similarity-in-state, because this tends to overestimate the relationship values. However, this method is applicable just for partial diallels in which endogamic crosses, theoretically, do not occur. In other words, it can just be used when prior information about heterotic groups is available.

An alternative to get around this restriction is the Lynch and Ritland (1999) relationship (LR) coefficient, proposed to infer about relationships between individuals by molecular markers when pedigree information is unavailable. Relationship coefficients are very useful for complex pedigrees or when there is little prior information about population structure, as well as in complex diallel designs. This coefficient uses moment estimators to estimate the relatedness between a pair of individuals, which is a continuous quantity defined in terms of probabilities of identity-by-descent (Ritland, 1996). In addition, this coefficient has been widely recommended, since it shows smaller sampling variance when compared to the usual coefficients (Lynch and Ritland, 1999; Bouvet et al., 2008). Based on these claims, similarity-in-state and similarity-by-descent must be better understood to be employed with BLUP in untested hybrid predictions.

The objective of this study was to evaluate the efficacy of BLUP for the prediction of missing crosses in maize hybrids, employing similarity-in-state, estimated by Roger's modified distance coefficient and similarity-by-descent estimated by the LR coefficient, in order to determine if predictions in untested hybrids are influenced by this approach.

MATERIAL AND METHODS

Description of materials

Seventeen $S_{0:2}$ progenies from three single-cross hybrid backgrounds were used. These populations were labeled as: population A - raised from hybrid P30F45, population B - from hybrid Dow 657, and population C - from hybrid DKB333B. These three populations (A, B and C) are separately preserved since good combining ability among them is obtained (Amorin et al., 2006). The 17 partial inbred lines were used in order to recover the best of 256 $S_{0:1}$ intra- and interpopulation crosses obtained by Amorim et al. (2006).

All 17 progenies were genotyped with 25 microsatellite markers (simple sequence repeats), with 9 markers linked to quantitative trait loci for grain yield. The other markers were selected according to the results from "bootstrap" analysis for those same populations (Amorim et al., 2006). The extraction of DNA, reaction preparation and estimates of genetic distances were described earlier by Balestre et al. (2008b). The Roger's modified distance (RMD) coefficient was utilized as the similarity-in-state coefficient because it is the expression best suited for this type of analysis; it provides a coefficient recommended for prearranging heterotic groups (Reif et al., 2005), besides showing acceptable values of distortion and stress for these populations in genetic divergence analysis (Balestre et al., 2008b).

A partial diallel (11 x 6) was performed with the 17 selected partial inbred lines, where, within each group, the lines were used to obtain the intra- and interpopulation crosses. This design was chosen due to evidence for good combining ability within and between those populations (Amorin et al., 2006).

Experimental design and correction of means by mixed model

Due to the small number of seeds, 60 crosses were evaluated together with four checks in two locations in Lavras County (Campus of the Universidade Federal de Lavras and Vitorinha Farm), in the agricultural year of 2007/2008. One of the checks was the hybrid DKB199, and the other three checks were single-cross hybrids that originated populations A

(P30F45), B (Dow 657) and C (DKB333B). The experimental design used was an 8 x 8-lattice design with three repetitions. The hybrids were grown in two-row plots, each row 3 m long and spaced 0.80 m apart, at a plant population density of 60,000 plants/ha, fertilized with 450 kg/ha NPK (8-28-16) at sowing and 100 kg/ha nitrogen as side dressing. The plots were harvested and grain yields (t/ha at 155 g H_2O/kg) were recorded.

The linear mixed model, in the matrix form, used for data analysis, was expressed as:

$$y = X\beta + Z_1 a + Z_2 b + Z_3 i + \delta \qquad \text{(Equation 1)}$$

where y is the vector of observed grain yields, β is the vector of the fixed effects (overall mean, locations and repetitions inside locations), a is the random vector of the genotypic values, b is the random vector of the block effects, i is the random vector of genotype by environment interactions, δ is the random error, and X, Z_1 Z_2, and Z_3 are the matrices of incidence related to β, a, b, and i, respectively.

Diallel design analysis by mixed models

The genotypic effects were partitioned into GCA and SCA by diallel analysis, using the method IV proposed by Griffing (1956). The analyses were done using the SAS® PROC IML System (SAS Institute, 2000). The estimates of fixed effects and BLUPs of the combination abilities (SCA and GCA) were obtained in a manner similar to that presented in Bernardo's (1995) proposal for partial diallels. However, in this study, the complete diallel model was chosen, considering all the genotypes within the same group.

The linear model considered was given by:

$$y_c = X\beta + Z_1 g + Z_2 s + e \qquad \text{(Equation 2)}$$

where: y_c is the vector of observed grain yield corrected by inter-block recovery (i.e., hybrid by two location trial mean); β is the vector of the fixed effects (overall mean and location); g is the vector of effects of GCA; s is the vector of SCA; e is the error vector, and X, Z_1 and Z_2 are the incidence matrices of effects β, g and s, respectively.

The joint solution for the fixed and random effects was obtained by the equation system, according to Henderson (1984):

$$\begin{bmatrix} \beta^0 \\ \hat{g} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} X'X & X'Z_1 & X'Z_2 \\ Z_1'X & Z_1Z_1 + A_1^{-1}\gamma_1 & Z_1'Z_2 \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + A_2^{-1}\gamma_2 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \end{bmatrix} \qquad \text{(Equation 3)}$$

with $\gamma_1 = \frac{\sigma_e^2}{\sigma_{GCA}^2}$ $\gamma_2 = \frac{\sigma_e^2}{\sigma_{SCA}^2}$, 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:

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

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 if contrary; S_{ac}: assigned 1 if a and c are identical and 0 if contrary; S_{ad}: assigned l if a and d are identical and 0 if contrary; S_{bc}: assigned 1 if b and c are identical and 0 if contrary; S_{bd}: assigned 1 if b and d are identical and 0 if contrary; 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 by 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 lines derived from single-cross hybrids as employed in this study.

Considering that the relationship estimates are obtained along 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 reference and then the individual Y (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:

$$\hat{r}_{xy} = \frac{\sum_1^L w_{r,x(k)} \hat{r}_{x,y(k)} + \sum_1^L w_{r,y(k)} \hat{r}_{y,x(k)}}{W_{r,x} + W_{r,y}} \qquad \text{(Equation 5)}$$

with

$$w_{r,x(k)} = \left(\frac{1}{Var[\hat{r}_{x,y(k)}]}\right) = \frac{(1 + S_{ab})(p_a + p_b) - 4 p_a p_b}{2 p_a p_b} \qquad \text{(Equation 6)}$$

where: $w_{r,x(k)}$ and $w_{r,y(k)}$ are the weights for the kth 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 being attributed to all loci.

The dominance relationship matrix (A_2) was constructed according to Henderson (1984) and Van Vleck (1993).

Also, Roger's distances were used to make the matrices A_1 and A_2 using the expression $s_{ij} = 1 - d_{ij}$, in which d_{ij} is RMD. Thus, the LR coefficient and RMD were compared by accuracy of untested single-cross predictions.

Prediction of untested single-cross hybrids and cross-validation procedure

The predictions of untested crosses for phenotypic values and SCA were performed using unbalanced simulations. Thus, using 60 evaluated hybrids, five degrees of imbalance were performed considering 5, 10, 20, 30, and 40 missing crosses. The process was repeated 100 times. A cross-validation procedure and hybrid prediction was performed as proposed by Bernardo (1994, 1995).

Considering that $\hat{y}_p$ are the averages of the evaluated hybrids corrected for β effects, one obtains:

$$\hat{y}_p = \left(Z'R^{-1}Z\right)^{-1} Z'R^{-1}(y - X\beta) \qquad \text{(Equation 7)}$$

Predictions of untested hybrids for grain yield were made using the following expression:

$$\hat{y}_{rt} = CV^{-1}\hat{y}_p \qquad \text{(Equation 8)}$$

in which: $\hat{y}_{nt}$ is the performance vector of the unbalanced crosses; C is the genetic covariance matrix between the unbalanced crosses and evaluated crosses; V is the variance and covariance matrix among the evaluated crosses.

In a similar way, the SCA values were predicted by the following expression:

$$d_{nt} = SV^{-1}\hat{s}_p \qquad \text{(Equation 9)}$$

in which: d_{nt} is the SCA vector of the unbalanced crosses, S is the genetic covariance matrix between the unbalanced crosses and evaluated crosses; V is the variance and covariance matrix among the evaluated crosses, $\hat{s}_p$ is the SCA vector of the evaluated hybrids.

The efficiency of BLUP in the prediction of the SCA and genotypic values was evaluated by correlation between observed values and predicted values (cross-validation procedure).

The Hamblin and Zimmermann (1986) selection efficiency index was applied to compare the 12 highest SCA and predictions of genotypic values, with the 12 highest SCA and phenotypic observed means. For that, the averages of untested single-cross hybrids, along all predictions, were considered for ranking the 12 highest values and compared with their respective observed values.

The RMD similarity coefficient and the LR coefficient were used to predict the untested crosses. Thus, the observed SCA and phenotypic means obtained by RMD and LR were correlated with the predicted values of the untested hybrid (cross-validation).

In addition, the prediction sum square (PRESS) was applied, using the expression given by:

$$PRESS = \frac{1}{I}\sum_{i=1}^{I}\frac{1}{J_i}\sum_{j=1}^{J_i}\left(\hat{\theta}_{ij} - \theta_i\right)^2 \qquad \text{(Equation 10)}$$

in which: θ_i is the observed values (for SCA and phenotypic cross values) of the ith untested hybrid; I is the 60 x 1 vector of hybrids to be predicted; J_i is the number of times that hybrid i was predicted, and $\hat{\theta}_{ij}$ is the jth prediction of θ_i. This expression measures the prediction sum square or the discrepancy between observed phenotypic means and predicted means.

RESULTS

The 17 genotypes showed a genetic similarity mean of 0.817 and a relationship mean of 0.077 (Table 1). This average relationship value, in spite of being relatively lower compared to the genetic similarity estimate, can be considered overestimated, because the negative values obtained by the Lynch and Ritland coefficient were considered equal to zero.

There were null relationship values between genotype groups (30B, 27B, 1B, 28B, and 29A), in relation to other genotypes (Table 1). Thus, one can infer that the LR coefficient provided a robust estimate of relationship between the genotypes, because it did not show any relationship value between the progenies derived from populations B and C, although it did show a small relationship value between the 30B and 29A progenies, which were derived from different populations.

This small relationship value between progenies was expected due to lack of major heterotic groups, resulting in a low relationship index among commercial hybrids. In Brazil, maize breeding programs do not preserve two-key heterotic groups as is frequent in other countries, i.e., dent and flint groups. Thus, the heterotic groups are established within breeding programs based on populations with high combining ability. These populations can be de-

Table 1. Estimates of similarity-in-state (above diagonal), obtained by Roger's modified distance (RMD) and relationship estimates (below diagonal), obtained by the Lynch and Ritland (LR) coefficient, in 17 $S_{0:2}$ maize progenies.

Relationship (LR)								Similarity-in-state (RMD)									
	30B	27B	19C	22C	1B	1C	14C	29C	5C	12C	15C	23C	25C	28B	7C	28C	29A
30B	1.00	0.93	0.73	0.73	0.90	0.73	0.69	0.74	0.74	0.76	0.74	0.73	0.68	1.00	0.80	0.73	0.76
27B	0.21	1.00	0.74	0.74	0.88	0.74	0.70	0.75	0.74	0.76	0.74	0.74	0.69	0.93	0.80	0.74	0.74
19C	0.00	0.00	1.00	0.84	0.73	0.88	0.84	0.90	0.88	0.93	0.88	0.84	0.81	0.73	0.97	0.88	0.74
22C	0.00	0.00	0.08	1.00	0.73	0.88	0.78	0.90	0.85	0.88	0.85	0.97	0.80	0.73	0.97	0.90	0.72
1B	0.10	0.12	0.00	0.00	1.00	0.76	0.71	0.72	0.73	0.75	0.73	0.73	0.66	0.90	0.80	0.75	0.74
1C	0.00	0.00	0.04	0.07	0.00	1.00	0.85	0.86	0.90	0.98	0.90	0.88	0.77	0.73	0.97	0.94	0.74
14C	0.00	0.00	0.00	0.03	0.00	0.18	1.00	0.81	0.88	0.85	0.88	0.78	0.78	0.69	0.97	0.86	0.69
29C	0.00	0.00	0.09	0.00	0.00	0.20	0.09	1.00	0.88	0.94	0.88	0.90	0.84	0.74	0.97	0.88	0.73
5C	0.00	0.00	0.13	0.04	0.00	0.29	0.20	0.12	1.00	0.90	0.97	0.85	0.85	0.74	0.97	0.88	0.77
12C	0.00	0.00	0.00	0.01	0.00	0.09	0.14	0.14	0.15	1.00	0.90	0.88	0.88	0.76	0.97	0.94	0.77
15C	0.00	0.00	0.07	0.00	0.00	0.21	0.18	0.21	0.18	0.29	1.00	0.85	0.85	0.74	0.97	0.88	0.77
23C	0.00	0.00	0.07	0.24	0.00	0.09	0.00	0.15	0.17	0.00	0.02	1.00	0.80	0.73	0.97	0.90	0.72
25C	0.00	0.00	0.00	0.00	0.00	0.01	0.13	0.33	0.07	0.28	0.16	0.04	1.00	0.68	0.97	0.78	0.67
28B	0.22	0.27	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.80	0.73	0.76
7C	0.00	0.00	0.07	0.03	0.00	0.50	0.50	0.10	0.37	0.39	0.39	0.24	0.36	0.00	1.00	0.97	0.80
28C	0.00	0.00	0.05	0.00	0.00	0.25	0.12	0.37	0.14	0.04	0.15	0.06	0.12	0.00	0.24	1.00	0.71
29A	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

rived from landraces, single-cross hybrids originating from elite lines (breeding population), or occasionally it is feasible use commercial hybrids. In new breeding programs, commercial single-cross hybrids have been frequently used for new population formation, given that the companies keep their heterotic groups disconnectedly, leading to good combining ability among commercial single-cross hybrids, making them potentially viable for the extraction of lines.

Figure 1 shows that RMD, which measures the similarity-in-state, clustered the genotypes in a way similar to that seen with the LR coefficient. In spite of the slight modifications within the groups, two major groups, which are represented by populations B and C, are evident. It should be pointed out that although the 29A genotype belongs to population A, this partial inbred line was allocated within the genotype group derived from population B, due to the small relationship value estimated with the 30B genotype.

The overall mean grain yield for the 60 evaluated hybrids was 8.962 t/ha, with a heritability value of 0.741. GCA variance was of low magnitude (0.002) in relation to SCA variance (1.622).

The correlation values, depending on the number of unbalanced hybrids, varied from 0.555 to 0.704 when the LR coefficient was used (Table 2). This value can be considered of moderate magnitude, because the accuracy obtained in this experiment was 0.86.

Nevertheless, considering RMD, the correlation values were inferior to that obtained by the LR coefficient, i.e., varying from 0.351 to 0.60 (Table 2). It was also observed that the correlation was higher when the number of predictor hybrids was high. This finding was also reported in previous studies (Bernardo, 1994, 1995).

The standard deviation of the correlation decreased when the number of predictor hybrids also decreased; in other words, the standard deviation of the correlation values was lower when the imbalance was higher (Table 2). In addition, it was observed that the standard deviation of the correlations using RMD was higher than with the LR coefficient across all levels of imbalance presented in this study.
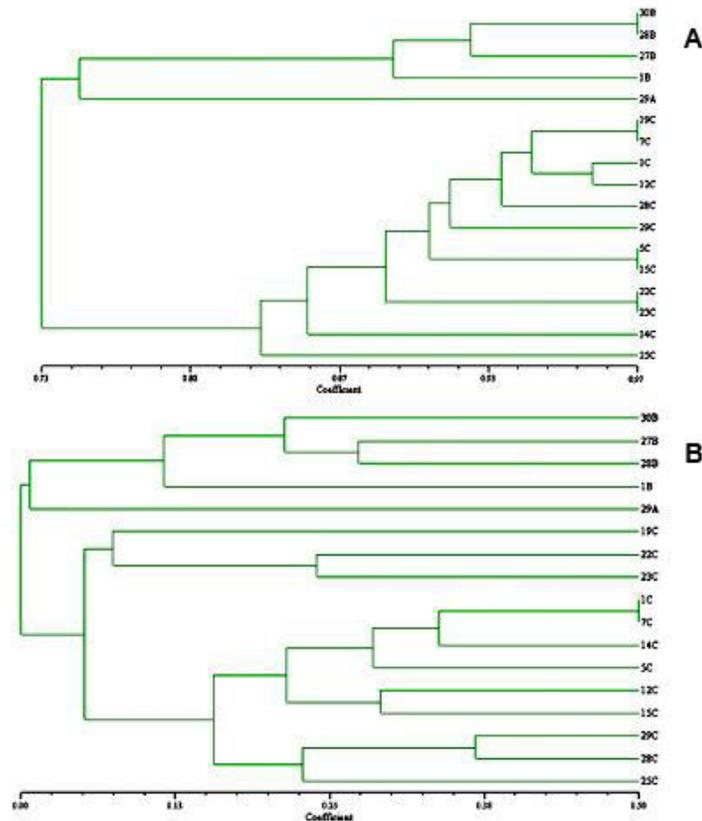
Figure 1. Dendrogram of genetic similarity (A) and relationship (B) estimated in 17 maize $S_{0:2}$ progenies. Progenies coded with the letter A originated from hybrid P30F45, and B and C progenies originated from the hybrids Dow 657 and DBB333B, respectively.

Table 2. Minimum, mean, maximum, and standard deviation of correlation values between the phenotypic means and predicted value across 100 simulations, considering the Lynch and Ritland (LR) estimated relationship coefficient and genetic similarity obtained by Roger's modified distance (RMD), in 17 $S_{0:2}$ maize progenies.

| | LR | | | | |
| | Number of missing hybrids | | | | |
	5	10	20	30	40
Minimum	0.219	-0.029	0.007	0.006	0.013
Average	0.702	0.643	0.601	0.560	0.555
Maximum	0.961	0.945	0.825	0.763	0.722
Standard deviation	0.208	0.203	0.173	0.173	0.112
	RMD				
	Number of missing hybrids				
	5	10	20	30	40
Minimum	-0.702	-0.025	0.070	0.012	-0.299
Average	0.600	0.562	0.503	0.459	0.351
Maximum	0.983	0.919	0.783	0.686	0.632
Standard deviation	0.382	0.261	0.189	0.175	0.172

The same results can be observed for the correlation between SCA of the untested hybrids and observed SCA values (Table 3). The correlation values varied from 0.617 to 0.704 when the LR coefficient was used and from 0.507 to 0.632 when RMD was used. It was also observed that when there was a high degree imbalance, the correlation between the predicted and observed values was smallest.

Table 3. Minimum, mean, maximum, and standard deviation of correlation between the observed specific combining ability (SCA) and the predicted value across 100 simulations, considering the Lynch and Ritland (LR) estimated relationship coefficient and genetic similarity obtained by means of Roger's modified distance (RMD), in 17 $S_{0:2}$ maize progenies.

	LR Number of missing hybrids				
	5	10	20	30	40
Minimum	-0.093	0.402	0.398	0.528	0.347
Average	0.704	0.701	0.647	0.644	0.617
Maximum	0.967	0.958	0.831	0.839	0.750
Standard deviation	0.300	0.136	0.108	0.082	0.090
	RMD Number of missing hybrids				
	5	10	20	30	40
Minimum	-0.698	-0.103	-0.233	0.227	-0.253
Average	0.632	0.573	0.525	0.524	0.507
Maximum	0.975	0.949	0.851	0.711	0.761
Standard deviation	0.381	0.269	0.233	0.129	0.201

From Table 4, one can note that RMD showed a higher predicted sum square error when compared to the LR relationship coefficient. Thus, we can infer that the prediction values obtained by RMD when compared with the LR coefficient, show a larger prediction error. In addition, one can observe that when the LR coefficient was used in the predictions for all degrees of imbalance, the correlation between the 60 observed values and the 60 predicted hybrid means was higher than with the RMD similarity coefficient (Table 4).

Table 4. Prediction sum square error (PRESS) between predicted hybrid values and observed phenotypic values taking into account 500 unbalanced simulations by BLUP associated with the Lynch and Ritland (LR) relationship coefficient and genetic similarity obtained by Roger's modified distance (RMD), in 17 $S_{0:2}$ maize progenies.

	LR		RMD	
	Genotypic value	SCA	Genotypic value	SCA
PRESS	1.057	0.839	1.149	0.978

SCA = specific combining ability.

The 12 superior untested hybrid means obtained for all degrees of imbalance are presented in Table 5. The observed phenotypic values associated with the 12 superior untested hybrids were of high magnitude (10.218 t/ha), showing an advantage of 1.26 t/ha in relation to the overall mean. The selection efficiency, taking into account the 12 highest observed phenotypic values, was 40% with a selection coincidence of 50% (Table 5). However, this result was only demonstrated when the LR coefficient was used.

Table 5. Selection efficiency index (SE), coincidence percentage (C) of the 12 best predicted genotypic values and their respective phenotypic means, in relation to the 12 superior observed phenotypic means obtained in the evaluation of 60 maize hybrids by LR (above) and RMD (below) coefficients.

Order	Cross	GV[1]	Grain yield
1	30BX7C	9.644	10.777
2	27BX7C	9.636	9.762
3	27BX28C	9.615	10.162
4	1CX28B	9.613	9.239
5	1BX25C	9.582	7.667
6	30BX28C	9.580	11.347
7	5CX28B	9.548	10.108
8	29CX28B	9.510	11.110
9	15CX28B	9.505	10.504
10	30BX23C	9.456	11.211
11	14CX28B	9.413	10.102
12	12CX28B	9.356	10.629
Average		9.538	10.218
C%	50.00		
SE%	40.00		
Order	Cross	GV[1]	Grain yield
1	27BX28B	10.949	4.457
2	1BX28B	10.295	6.752
3	1BX29A	10.204	9.622
4	30BX29A	10.201	8.415
5	30BX28C	9.932	11.347
6	1CX29A	9.836	8.169
7	27BX29A	9.784	9.080
8	1CX28B	9.769	9.239
9	27BX28C	9.718	10.162
10	1BX23C	9.681	9.771
11	30BX23C	9.635	10.108
12	5CX28B	9.621	10.777
Average		9.969	9.027
C%	16.70		
SE%	0.00		

[1]Prediction deviation associated with each prediction considering 500 simulations of unbalanced simulations. GV = genotypic values obtained across all 500 simulations.

On the other hand, the selection efficiency index was equal to zero when RMD was utilized (Table 5). Although it showed a moderate correlation between the observed values and predicted values, the 12 highest predicted hybrids did not coincide with the top observed phenotypic hybrid means.

Table 6 shows that the 12 untested genotypes selected for SCA predicted by BLUP with the LR coefficient had high observed SCA values (0.968) and moderate efficiency of selection (52%). In addition, there was 58% coincidence between the 12 highest selected untested SCA values and the 12 highest observed SCA values.

The same is not applicable to the RMD predictions, because the selection coincidence and selection efficiency were lowest (Table 6). In contrast, RMD selected the high SCA observed value, which was not attained using the LR coefficient (Table 6). However, on average, the LR coefficient was superior to RMD in the selection of the best individuals.

Table 6. Selection efficiency index (SE), coincidence percentage (C) of the 12 superior predicted specific combining ability values (SCA[1]) and their respective observed values (SCA[2]) , considering 500 simulations, in relation to the 12 superior specific ability values obtained in the evaluation of 60 maize hybrids by LR (above) and RMD (below) coefficients.

Order	Cross	SCA[1]	SCA[2]
1	30BX7C	0.285	1.386
2	1CX28B	0.278	0.248
3	27BX7C	0.267	0.633
4	30BX28C	0.254	1.776
5	1BX25C	0.235	-0.930
6	15CX28B	0.232	1.169
7	1BX7C	0.231	1.625
8	5CX28B	0.218	0.867
9	30BX23C	0.211	1.661
10	12CX28B	0.209	1.240
11	27BX28C	0.197	0.910
12	1BX23C	0.188	0.603
Average		0.234	0.933
C%	58.00		
SE%	50.00		
Order	Cross	SCA[1]	SCA[2]
1	27BX28B	0.388	-3.399
2	30BX29A	0.297	-0.457
3	30BX28C	0.227	1.776
4	1BX28B	0.180	-1.721
5	1BX29A	0.168	0.432
6	1BX23C	0.120	0.604
7	1CX28B	0.097	0.248
8	30BX23C	0.088	1.661
9	1BX28C	0.087	3.339
10	27BX28C	0.078	0.911
11	27BX29A	0.074	0.034
12	12CX28B	0.068	1.240
Average		0.156	0.389
C%	33.33		
SE%	20.00		

Thus, it is possible to infer that RMD, which is considered a similarity-in-state coefficient, shows lower efficiency in the selection of superior genotypes, when compared to the LR coefficient.

DISCUSSION

The use of BLUP, for the prediction of untested hybrids has been suggested by several authors (Bernardo, 1994, 1995, 1997; Charcosset et al., 1998; Piepho et al., 2008; Schrag et al., 2009). In all these works, a moderate correlation was found to be between the untested hybrids and their observed phenotypic values. In our study, similar results were also obtained, i.e., the correlation between the untested hybrids and their observed values was moderate, demonstrating that BLUP approach shows a good predictive efficiency in untested hybrids. This efficiency can be quantified by considering the square root of heritability (0.861), which delimits the maximum correlation value of the predictions (Bernardo, 1994). In our results, the correlation magnitudes retained were from 64.5 to 81.4% of the maximum correlation limits, which is bordered by the square root of heritability.

According to Bernardo (1996), these correlation values can be considered sufficiently high, because they provide a high probability (±80%) of selecting the best hybrid in a

superior group considering a selection intensity of 20%. However, the best genotype was not placed within the group of selected hybrids, although there was 50% coincidence and 40% selection efficiency between the selected hybrids and the highest observed phenotypic means.

The SCA correlation values showed a moderate magnitude (0.617 to 0.704), yet were higher than those reported by Bernardo (1994, 1995), André (1999) and Iemma (2003). Bernardo (1994, 1995) attributed the low correlation between untested SCA predictions with observed SCA (r ≤ 0.15) to bias in the SCA variance. This author observed that SCA variance was extremely low when compared to the GCA variance. Thus, the crosses were strongly influenced by GCA, converging to more accurate predictions, due to low SCA effects. In other words, in the study by Bernardo (1994), the predictions considering the additive model (reduced model) showed almost the same results in relation to the full model, demonstrating low SCA influence in the predictions of the untested hybrids.

In this study, SCA variance (V_{SCA}) was higher than GCA variance, capturing about 99% of the total genetic variance. Therefore, it can be inferred that the predictions obtained in this study are of great importance, considering that the hybrid combinations were highly influenced by SCA. The hypothesis of possible bias in the V_{SCA} estimate should be taken into account, considering the additive and dominance variances usually reported in the maize crop (Hallauer and Miranda Filho, 1981) and mainly due to inbred and outbred combinations obtained in the complete diallel (Boer and Hoeschele, 1993). Thus, the results can be considered restricted to the group of hybrids evaluated in this study.

Considering the probability of selecting the best specific combinations (SCA), it can be demonstrated that the selection efficiency index among the 12 observed superior hybrids and 12 predicted superior hybrids was 50%, with 58% coincidence between them. Although the SCA values were obtained by means of several predictions with different combinations and predictor hybrids, it is seen that there is reasonable probability of selection of the best hybrids without being necessary to assess them.

In the comparison between similarity-in-state (RMD) and the LR coefficient, we found that in all analyses, the use of RMD showed inferior results. That is, RMD showed lower correlation between predicted and observed grain yield means (cross-validation) and higher PRESS when compared with the LR coefficient, demonstrating that the predictions obtained with the RMD coefficient were biased in relation to the LR coefficient. In addition, this inferiority can also be verified considering the selection efficiency of the superior SCA and hybrid combinations.

This larger prediction error (PRESS) caused by RMD on SCA and genotypic mean predictions was expected, due to the recovery of genetic information that takes into consideration only the evolutionary aspect (similarity-in-state) of the individuals and not their relationship. Thus, with greater or lesser weight, all genetic covariance values are recovered by the linear predictor (BLUP) for each SCA and GCA value, including those individuals that do not share alleles inherited from common progenitors. This fact results in low constant predictions, since all hybrids, even those that not share common alleles, can contribute to the prediction of unrelated hybrids.

However, this bias is eliminated by the use of Lynch and Ritland's (1996) coefficient, since despite using molecular data (similarity-by-state), it takes into consideration the conditional probability that two individuals have alleles inherited from common progenitors (similarity-by-descent). Thus, only individuals having common alleles, with some probabil-

ity level, take part in the prediction of the untested hybrids. Therefore, those pairwise relationships equal to zero will not contribute to the prediction, resulting in more reliable results.

The use of similarity-in-state does not limit just the RMD coefficient, but also the coefficients that take into account merely the co-occurrence of bands (0 and 1), i.e., Dice, Jaccard, simple method, etc. Accordingly, Bernardo (1994) suggested an expression to correct this bias. However, as already commented, in complete diallels the coancestry coefficient is not applied because it does not contemplate all pairwise relationships between the genotypes under evaluation. That is, the coancestry expression considers that every and any similarity between two groups in partial diallels is due to similarity-in-state. Thus, hypothetical similarity-in-state is removed from genotypes placed within each group. On the other hand, the LR coefficient can be used in complete diallels or complex crosses, because the relationship values obtained by this estimator consider all individuals under evaluation.

Thus, it can be inferred that the use of similarity-in-state, even when using further coefficients that exploit all genetic covariance, without pondering the probability that those alleles are inherited from common genitors, can show lower efficiency in the prediction of the untested hybrids.

Another important aspect to be highlighted is the reason for using an unbalanced complete diallel instead of a partial diallel model for the analysis. As stated in Material and Methods, the crosses were performed in an 11 x 6 scheme, which would lead to a partial diallel to evaluate GCA and SCA. However, these crosses were obtained in order to recover the intra- and interpopulation combinations, which led to the allocation of related individuals outside the groups, i.e., not necessarily in distinct groups.

In this study, the use of a partial diallel loses part of the relationship in matrix A because kinship individuals were allocated outside of groups, while in complete diallel, even when highly unbalanced, can recover this information, leading to greater accuracy in predicting GCA and SCA values. This fact became evident, when the GCA and SCA level of accuracy, considering the unbalanced complete diallel design (0.99 and 0.85 for GCA and SCA, respectively) was compared with the partial diallel design exploiting relationship information (0.988 and 0.844) and without relationship information (0.987 and 0.833). Therefore, the preference for the use of an unbalanced complete diallel instead of a partial one can be justified. Although the results obtained in this study were restricted to two test environments and limited genotype numbers, the magnitude of the predictions reached was similar to reports of previous studies, in addition, there was low correlation among hybrid performance in the two environments and high dominance effect, showing that BLUP is an effective methodology for the prediction and selection of the untested crosses, even when there is a predominance of non-additive effects and a high degree of imbalance.

Furthermore, there was evidence that the Lynch and Ritland relationship coefficient associated with BLUP provided more accurate predictions when compared to similarity-in-state, obtained by Roger's modified distances.

ACKNOWLEDGMENTS

REFERENCES

Amorim EP, Amorim UBO, Dos Santos JB, Pereira De Souza A, et al. (2006). Genetic distance based on SSR and grain yield of inter and intrapopulational maize single cross hybrids. *Maydica* 51: 507-513.

André CMG (1999). Avaliação da Melhor Predição Linear não Tendenciosa (BLUP) Associada ao Uso de Marcadores Moleculares na Análise Dialélica. Master's thesis, Universidade Federal de Lavras, Lavras.

Balestre M, Machado JC, Lima JL, Souza JC, et al. (2008a). Genetic distance estimates among single cross hybrids and correlation with specific combining ability and yield in corn double cross hybrids. *Genet. Mol. Res.* 7: 65-73.

Balestre M, Von Pinho RG, Souza JC and Lima JL (2008b). Comparison of maize similarity and dissimilarity genetic coefficients based on microsatellite markers. *Genet. Mol. Res.* 7: 695-705.

Balestre M, Von Pinho RG, Souza JC and Machado JC (2008c). 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.

Bernando 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 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.

Bernardo R (1997). RFLP markers and predicted testcross performance of maize sister inbreds. *Theor. Appl. Genet.* 95: 655-659.

Boer IJM de and Hoeschele I (1993). Genetic evaluation methods for populations with dominance and inbreeding. *Theor. Appl. Genet.* 86: 245-258.

Bouvet JM, Kelly B, Sanou H and Allal F (2008). Comparison of marker- and pedigree-based methods for estimating heritability in an agroforestry population of *Vitellaria paradoxa* C.F. Gaertn. (shea tree). *Genet. Resour. Crop Evol.* 55: 1291-1301.

Charcosset A, Bonnisseau B, Touchebeuf O, Burstin J, et al. (1998). Prediction of maize hybrid silage performance using marker data: comparison of several models for specific combining ability. *Crop Sci.* 38: 38-44.

Cruz CD, Regazzi AJ and Carneiro PCS (1997). Modelos Biométricos Aplicados ao Melhoramento Genético. Editora UFV, Viçosa.

Dias LA, Picoli EA, Rocha RB and Alfenas AC (2004). *A priori* choice of hybrid parents in plants. *Genet. Mol. Res.* 3: 356-368.

Griffing B (1956). Concept of general and specific combining ability in relation to diallel crossing system. *Aust. J. Biol. Sci.* 9: 463-493.

Hallauer AR and Miranda Filho JB (1981). Quantitative Genetics in Maize Breeding. Iowa State University Press, Ames.

Hamblin J and Zimmermann MJ (1986). Breeding common bean for yield in mixtures. *Plant Breed. Rev.* 4: 245-272.

Henderson CR (1984). Applications of Liner Models in Animal Breeding. University of Guelph Press, Guelph.

Iemma M (2003). Uso do Melhor Predictor Linear Não Viesado (BLUP) em Análise Diallélica e Predição de Híbridos. Master's thesis, Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

Kempthorne O and Curnow RN (1961). The partial diallel cross. *Biometrics* 17: 229-250.

Lynch M and Ritland K (1999). Estimation of pairwise relatedness with molecular markers. *Genetics* 152: 1753-1766.

Melchinger AE (1999). Genetic Diversity and Heterosis. In: The Genetics and Exploitation of Heterosis in Crops (Coors JG and Pandey S, eds.). American Society of Agronomy-Crop Science Society of America, Madison, 99-118.

Nejati-Javaremi A, Smith C and Gibson JP (1997). Effect of total allelic relationship on accuracy of evaluation and response to selection. *J. Anim. Sci.* 75: 1738-1745.

Piepho HP, Möhring J, Melchinger AE and Büchse A (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161: 209-228.

Ramalho MAP, Santos JB dos and Zimmermann MJ de O (1993). Genética Quantitativa em Plantas Autógamas, Aplicações ao Melhoramento do Feijoeiro. Editora UFG, Goiânia.

Reif JC, Melchinger AE, Xia XC, Warburton ML, et al. (2003). Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. *Crop Sci.* 43: 1275-1282.

Reif JC, Melchinger AE and Frisch M (2005). Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Sci.* 45: 1-7.

Ritland K (1996). Estimators for pairwise relatedness and inbreeding coefficients. *Genet. Res.* 67: 175-186.

Rousset F (2007). Inferences from Spatial Population Genetics. In: Handbook of Statistical Genetics (Balding DJ, Bishop M and Cannings C, eds.) 3rd edn. Wiley, Chichester, 945-979.

SAS Institute (2000). Version 8. SAS Institute Inc., Cary.

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.

Van Vleck LD (1993). Selection Index in Introduction to Mixed Models Methods for Genetic Improvement of Animals. CRC Press, Boca Raton.

Vencovsky R and Barriga P (1992). Genética Biométrica no Fitomelhoramento. Sociedade Brasileira de Genética, Ribeirão Preto.