

CROP BREEDING, GENETICS & CYTOLOGY

Prediction of Maize Single-Cross Performance Using RFLPs and Information from Related Hybrids

Rex Bernardo*

ABSTRACT

Methods for predicting hybrid yield would facilitate the identification of superior maize (*Zea mays* L.) single crosses. Best linear unbiased prediction of the performance of single crosses, based on (i) restriction fragment length polymorphism (RFLP) data on the parental inbreds and (ii) yield data on a related set of single crosses, was evaluated. Yields of m single crosses were predicted as $y_M = C V^{-1} y_p$, where: $y_M = m \times 1$ vector of predicted yields of missing (i.e., no yield data available) single crosses; $C = m \times n$ matrix of genetic covariances between the missing and predictor hybrids; $V = n \times n$ matrix of phenotypic variances and covariances among predictor hybrids; and $y_p = n \times 1$ vector of predictor hybrid yields corrected for trial effects. From a set of 54 single crosses, made between six Iowa Stiff Stalk Synthetic (SSS) and nine non-SSS inbreds, 100 different sets of $n = 10, 15, 20, 25,$ or 30 predictor hybrids were chosen at random. Pooled correlations between predicted and observed yields of the remaining $(54 - n)$ hybrids ranged from 0.654 to 0.800. The correlations were slightly higher when dominance variance was included in the model or when coefficients of coancestry were determined from RFLP rather than pedigree data. The correlations remained relatively stable across different, arbitrary values of genetic variances. The results suggested that single-cross yield can be predicted effectively based on parental RFLP data and yields of a related set of hybrids.

IN COMMERCIAL MAIZE BREEDING PROGRAMS, the identification of pairs of inbreds with superior yield performance in single-cross combination is costly and time consuming. Because of strong dominance effects for maize grain yield (Hallauer and Miranda, 1981), hybrid performance cannot be predicted from inbred per se data (Smith, 1986). Thus, it is necessary to cross the inbreds and evaluate the hybrids themselves in extensive yield trials.

Commercial maize hybrids are typically made between inbreds from opposite, complementary heterotic groups. Unfortunately, maize breeders are unable to evaluate all possible single-cross combinations between inbreds from different heterotic groups because the number of possible hybrids is often prohibitive. To illustrate, there are 100 possible single crosses between 10 inbreds from Heterotic Group X and 10 inbreds from Heterotic Group Y. But if the breeder has 50 inbreds from X and 50 inbreds from Y, the number of possible single crosses increases to 2500. Whereas the breeder may be able to adequately test 100 hybrids in yield trials, evaluating 2500 hybrids would be much more costly.

The use of inbred per se molecular marker data has been suggested as a means to (i) determine the heterotic grouping and degree of genetic relationship among inbreds and (ii) predict hybrid performance based on molecular

marker dissimilarity between parents. Restriction fragment length polymorphisms have been found useful for assigning inbreds to heterotic groups as well as for determining relationships among inbreds in the same heterotic group (Smith et al., 1990; Melchinger et al., 1991; Dudley et al., 1991; Hogan and Dudley, 1992; Bernardo, 1993). But in theoretical (Bernardo, 1992; Charcosset et al., 1991) as well as empirical studies using RFLPs (Godshalk et al., 1990; Melchinger et al., 1990; Dudley et al., 1991), the correlations between single-cross yield and molecular marker dissimilarity between parents have been too low to be of any predictive value.

Although yield data may not be available for all possible single-cross combinations among available inbreds, some of these combinations already may have been evaluated by the breeder. For example, yield data may be available for 200 out of 2500 possible hybrids between 50 inbreds from X and 50 inbreds from Y. If information on the RFLP or pedigree relationships among the 100 parental inbreds is available, by best linear unbiased prediction (BLUP) (Henderson, 1975; 1985) the yield data on the 200 tested hybrids may be used to predict the yields of the remaining 2300 untested hybrids. The BLUP procedure, usually assuming an additive and intrapopulation genetic model, have been used extensively in dairy cattle (*Bos taurus*) evaluation (Henderson, 1988). However, the application of BLUP in crop plants has been very limited (White and Hodge, 1989).

The objectives of this paper are (i) to illustrate and evaluate the usefulness of BLUP for predicting single-cross performance using RFLPs and information from related hybrids, and (ii) to test the robustness of the method with regards to genetic model and measure of genetic relationship.

THEORY

Genetic Covariance Between Single Crosses

Assume x and x' are inbreds from Heterotic Group X, whereas y and y' are inbreds from Heterotic Group Y. Inbreds from X and Y are unrelated. The genetic value (expressed as a deviation from the population mean) of hybrid (x,y) at the k th locus is:

$$G_{ij/k} = \alpha_{i/k}^x + \alpha_{j/k}^y + \delta_{i/kj/k}^{x,y}$$

where: $\alpha_{i/k}^x$ = testcross additive effect of Allele i_k^x (received from Heterotic Group X through Inbred x) at Locus k ; $\alpha_{j/k}^y$ = testcross additive effect of Allele j_k^y (received from Heterotic Group Y through Inbred y) at Locus k ; and $\delta_{i/kj/k}^{x,y}$ = dominance effect of allelic pair $i_k^x j_k^y$ at Locus k .

Assuming negligible epistasis, no linkage, and gametic equi-

Abbreviations: RFLPs, restriction fragment length polymorphisms; BLUP, best linear unbiased prediction; REML, restricted maximum likelihood; GCA, general combining ability; SCA, specific combining ability; SSS, Iowa Stiff Stalk Synthetic; NSSS, non-Iowa Stiff Stalk Synthetic

Limagrain Genetics, 4805 W. Old Church Rd., Champaign, IL 61821. Received 19 April 1993. *Corresponding author.

librium in X and Y, Melchinger's (1988) expression for the genetic covariance between hybrids (x.y) and (x'.y') reduces to:

$$\begin{aligned} \text{Cov}[(x.y), (x'.y')] &= \Sigma [f_{xx'} \Sigma p_{i/k}^X (\alpha_{i/k}^X)^2 + f_{yy'} \Sigma p_{j/k}^Y (\alpha_{j/k}^Y)^2 + f_{xx'} f_{yy'} \\ &\quad \Sigma \Sigma p_{i/k}^X p_{j/k}^Y (\delta_{i/kj/k}^X)^2] \\ &= f_{xx'} V_{A(X/Y)} + f_{yy'} V_{A(Y/X)} + f_{xx'} f_{yy'} V_{D(XY)} \quad [1] \end{aligned}$$

where $f_{xx'}$ = coefficient of coancestry between Inbreds x and x' ; $f_{yy'}$ = coefficient of coancestry between Inbreds y and y' ; $p_{i/k}^X$ = frequency of Allele i_k^X in Heterotic Group X; $p_{j/k}^Y$ = frequency of Allele j_k^Y in Heterotic Group Y; $V_{A(X/Y)}$ = pooled (across k loci) testcross additive variance of Allele i_k^X ; $V_{A(Y/X)}$ = pooled (across k loci) testcross additive variance of Allele j_k^Y ; and $V_{D(XY)}$ = pooled (across k loci) dominance variance of allelic pair $i_k^X j_k^Y$.¹

The coefficients of coancestry ($f_{xx'}$ and $f_{yy'}$) can be determined from pedigree records (Falconer, 1981) or from RFLP data as follows (Bernardo, 1993):

$$f_{xx'}^R = [S_{xx'} - 1/2(S_{xy} + S_{x'y})] / [1 - 1/2(S_{xy} + S_{x'y})] \quad [2]$$

where $f_{xx'}^R$ = RFLP-based estimate of coefficient of coancestry between x and x' ; $S_{xx'}$ = proportion of RFLP variants common to x and x' ; and S_{xy} (or $S_{x'y}$) = average proportion of RFLP variants common to x (or x') and a series of inbreds from Heterotic Group Y.

Estimation of Genetic Variances

Traditional mating designs for estimating genetic variances, such as the Design II mating scheme, are applicable if the parental inbreds are unrelated. Because the parental inbreds used in predicting single-cross performance are related, such mating designs cannot be used. In contrast, restricted maximum likelihood (REML) (Harville, 1977; Henderson, 1985) procedures for estimating genetic variances account for relationships among inbreds. Assume that n single crosses are made between n_X inbreds from Heterotic Group X and n_Y inbreds from Heterotic Group Y. The single crosses are evaluated in t different yield trials resulting in p total observations. If all possible single crosses between heterotic groups are evaluated, $n = n_X n_Y$, and if all single crosses are evaluated in each yield trial, $p = nt$. The linear model is:

$$y = X\beta + Z_1 a_X + Z_2 a_Y + Z d + e$$

where y = $p \times 1$ vector of observed yields; β = $t \times 1$ vector of trial effects; a_X = $n_X \times 1$ vector of general combining ability (GCA) effects of inbreds in Heterotic Group X; a_Y = $n_Y \times 1$ vector of GCA effects of inbreds in Heterotic Group Y; d = $n \times 1$ vector of specific combining ability (SCA) effects; e = vector of residuals; and X , Z_1 , Z_2 , and Z are incidence matrices of 1s and 0s relating y to β , a_X , a_Y , and d , respectively.

Decomposition of Eq. [1] leads to the covariances between GCA and between SCA effects. The covariance matrix of a_X is $A_1 V_{A(X/Y)}$. The diagonal elements of A_1 are equal to 1 and the off-diagonal elements are equal to $f_{xx'}$ between the corresponding inbreds. Similarly, the covariance matrix of a_Y is

¹ Using the notation of Hallauer and Miranda (1981) for a single-locus, two-allele model:
 $\text{Cov}[(x.y), (x'.y')] = f_{xx'} p(1-p)[a+d(1-2r)]^2 + f_{yy'} r(1-r)[a+d(1-2p)]^2 + f_{xx'} f_{yy'} 4p(1-p)r(1-r)d^2$
 where: p and r are the frequencies of the + allele in X and Y, respectively; a is half the difference between genetic values of the homozygotes; and d is the value of the heterozygote.

$A_2 V_{A(Y/X)}$. The diagonal elements of A_2 are equal to 1 and the off-diagonal elements are equal to $f_{yy'}$. The covariance matrix of d is $DV_{D(XY)}$, with the diagonal elements of D equal to 1 and the off-diagonal elements equal to $f_{xx'} f_{yy'}$.

Suppose $n = 5$ hybrids between $n_X = 2$ inbreds from X and $n_Y = 3$ inbreds from Y are evaluated in $t = 2$ sets of yield trials:

Trial	Hybrid	Yield
1	Inbred X1 · Inbred Y1	10
1	Inbred X1 · Inbred Y2	15
1	Inbred X1 · Inbred Y3	12
1	Inbred X2 · Inbred Y3	20
2	Inbred X1 · Inbred Y1	15
2	Inbred X2 · Inbred Y2	20

The system of equations is:

$$y = X \beta + Z_1 a_X + Z_2 a_Y + Z d$$

10	1 0	β_1	1 0	a_{X1}	1 0 0	a_{Y1}	1 0 0 0 0	d_1
15	1 0	β_2	1 0	a_{X2}	0 1 0	a_{Y2}	0 1 0 0 0	d_2
12	1 0		1 0		0 0 1	a_{Y3}	0 0 1 0 0	d_3
20	1 0		0 1		0 0 1		0 0 0 1 0	d_4
15	0 1		1 0		1 0 0		1 0 0 0 0	d_5
20	0 1		0 1		0 1 0		0 0 0 0 1	

Assuming nongenetic effects are uncorrelated, β , a_X , a_Y , and d can be calculated from the following set of equations:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a}_X \\ \hat{a}_Y \\ \hat{d} \end{bmatrix} = \begin{bmatrix} X'X & X'Z_1 & X'Z_2 & X'Z \\ Z_1'X & Z_1'Z_1 + A_1^{-1} \Phi_1 & Z_1'Z_2 & Z_1'Z \\ Z_2'X & Z_2'Z_1 & Z_2'Z_2 + A_2^{-1} \Phi_2 & Z_2'Z \\ Z'X & Z'Z_1 & Z'Z_2 & Z'Z + D^{-1} \Phi_3 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \\ Z'y \end{bmatrix}$$

$$= \begin{bmatrix} C_{00} & C_{01} & C_{02} & C_{03} \\ C_{10} & C_{11} & C_{12} & C_{13} \\ C_{20} & C_{21} & C_{22} & C_{23} \\ C_{30} & C_{31} & C_{32} & C_{33} \end{bmatrix} \begin{bmatrix} X'y \\ Z_1'y \\ Z_2'y \\ Z'y \end{bmatrix}$$

where V_E = nongenetic variance; $\Phi_1 = V_E/V_{A(X/Y)}$; $\Phi_2 = V_E/V_{A(Y/X)}$; and $\Phi_3 = V_E/V_{D(XY)}$.

REML estimates of nongenetic and genetic variances can be obtained by iterating on (Henderson, 1985):

$$V_E = [y'y - (\text{solution vector})'(\text{right-hand side vector})] / (p - t);$$

$$V_{A(X/Y)} = (a_X' A_1^{-1} a_X + V_E \text{tr} A_1^{-1} C_{11}) / n_X;$$

$$V_{A(Y/X)} = (a_Y' A_2^{-1} a_Y + V_E \text{tr} A_2^{-1} C_{22}) / n_Y;$$

$$V_{D(XY)} = (d' D^{-1} d + V_E \text{tr} D^{-1} C_{33}) / n$$

where tr = trace operator, i.e., sum of the diagonal elements of a matrix.

Best Linear Unbiased Prediction of Single-Cross Performance

The yield performance of a set of m missing single crosses can be predicted based on yields of a set of n predictor hybrids. Let y_p be a $n \times 1$ vector of average yields of predictor hybrids, corrected for yield trial effects (β is):

$$y_p = (Z'Z)^{-1} Z'(y - X\beta) \quad [3]$$

where Z , y , X , and β are defined in the preceding section. Yields of the missing single crosses can be predicted as:

$$y_M = C V^{-1} y_p \quad [4]$$

where: $\mathbf{y}_M = m \times 1$ vector of predicted yields of missing single crosses; $\mathbf{C} = m \times n$ matrix of genetic covariances between the i th missing single cross and j th predictor hybrid; and $\mathbf{V} = n \times n$ phenotypic variance-covariance matrix among the predictor hybrids. Assuming nongenetic effects are uncorrelated, the elements of \mathbf{C} as well as the off-diagonal elements of \mathbf{V} are calculated using Eq. [1]. The i th diagonal element of \mathbf{V} is equal to $V_{A(X/Y)} + V_{A(Y/X)} + V_{D(XY)} + V_E/$ (number of observations for the i th predictor hybrid).

MATERIALS AND METHODS

Data Set

Fifty-four maize single crosses were obtained by crossing six inbreds related to Iowa Stiff Stalk Synthetic (SSS) to nine inbreds unrelated to SSS (designated NSSS, and predominantly Lancaster Surecrop). The 54 hybrids along with two checks were evaluated in 1992 at Kirkland and Somonauk, IL, and Clarence and Manchester, IA, using a 7×8 rectangular lattice design with two replicates. The trial at Somonauk was discarded because of poor seedling emergence. The hybrids were grown in two-row plots, each row 5 m long and spaced 0.76 m apart, at a plant population density of 73 700 plants ha^{-1} . The plots were machine-harvested and grain yields (t ha^{-1} at 155 $\text{g H}_2\text{O kg}^{-1}$) were recorded. Lattice analyses of variance were performed on yield data from each location. Adjusted entry means and average effective error variances were used in a combined analysis of variance across locations.

Coefficients of coancestry among the six SSS and nine NSSS were determined from pedigree records and from RFLP data (Eq. [2]). When pedigree relationships between ancestors of two inbreds were unknown, the coefficient of coancestry between the ancestral inbreds was assumed zero. Patterns of hybridization fragments (bands) were determined using 110 well-dispersed probes and restriction digests of genomic DNA from each of the 15 parental inbreds. The restriction enzymes *EcoRI* and *HindIII* were used in combination with each probe. Each of the 220 probe-enzyme combinations was considered an RFLP locus, and each unique banding pattern an RFLP variant. DNA extraction, restriction enzyme digestion, gel electrophoresis, Southern blotting, and probe hybridization were done as described by Murigneux et al. (1993). Equation [2] may give negative estimates of coefficients of coancestry due to sampling error (Bernardo, 1993). Because coefficients of coancestry are nonnegative by definition, negative estimates were set equal to zero. Confidence intervals ($\alpha = 0.05$) on RFLP-based estimates of coefficients of coancestry were obtained as the variance of a ratio (Kempthorne, 1969) assuming that the proportions of shared RFLP variants followed normal approximations of a binomial distribution (Steel and Torrie, 1980).

Restriction fragment length polymorphism- and pedigree-based coefficients of coancestry were used to construct two sets of \mathbf{A}_1 , \mathbf{A}_2 , and \mathbf{D} matrices. Correspondingly, two different sets of REML estimates of β , $V_{A(X/Y)}$, $V_{A(Y/X)}$, $V_{D(XY)}$, and V_E

were obtained. A computer program for estimating genetic variances by REML was written in QBasic and run on a PC-compatible 386 microcomputer. Tolerances equal to 0.0001 were specified during iteration.

Method Validation

Sets of $n = 10, 15, 20, 25,$ and 30 predictor hybrids were selected at random from the 54 hybrids. Yields of each set of n predictor hybrids were obtained according to Eq. [3] and using trial effects (β) estimated from the analysis of all 54 single crosses. Based on the yields of the n predictor hybrids, yields of the remaining $(54 - n)$ hybrids were predicted using Eq. [4]. Simple correlations between the predicted and observed yields of the $(54 - n)$ missing single crosses were calculated. For each value of n , the sampling process was repeated for a total of 100 different sets of n predictor and $(54 - n)$ missing hybrids. Across the 100 repetitions, pooled correlation coefficients were calculated for each value of n , and 95% confidence intervals were obtained if the correlation coefficients were found homogeneous (Steel and Torrie, 1980).

Two genetic models were used: (i) full model and (ii) GCA model, wherein dominance variance [$V_{D(XY)}$] was excluded from Eq. [1]. For each genetic model, prediction of single-cross performance was done using coefficients of coancestry determined from (i) RFLP data and (ii) pedigree records.

RESULTS AND DISCUSSION

Means, Genetic Variances, and Relationships among Parental Inbreds

Single-cross yields, averaged across three locations, ranged from 9.61 t ha^{-1} for SSS1 \times NSSS9 to 13.04 t ha^{-1} for SSS6 \times NSSS4 (Table 1). The estimate of average performance of the single crosses was 11.07 t ha^{-1} when RFLP-based coefficients of coancestry were used in BLUP, and 11.03 t ha^{-1} when pedigree relationships were used in BLUP. Nongenetic and genetic variances as well as entry-mean heritability, estimated using RFLP and pedigree (in parentheses) relationships in BLUP, were $V_E = 0.5537$ (0.5520); $V_{A(X/Y)} = 0.5409$ (0.8186); $V_{A(Y/X)} = 0.1783$ (0.2033); $V_{D(XY)} = 0.0757$ (0.0769); and $h^2 = 0.812$ (0.857).

Restriction fragment length polymorphism-based estimates of coefficients of coancestry among SSS inbreds (f_{xx}^R) ranged from 0.05 for (SSS2, SSS5) to 0.65 for three different pairs of SSS inbreds (Table 2). Pedigree-based coefficients of coancestry among SSS inbreds (f_{xx}) ranged from 0.04 for three different pairs of inbreds to 0.77 for (SSS1, SSS6) (Table 2). Among the 15 pairwise combinations of SSS inbreds, the average f_{xx}^R was 0.35 and the average f_{xx} was 0.33. The two estimates of relationship among SSS inbreds were highly correlated ($r = 0.92^*$, $\alpha = 0.05$). The f_{xx}^R was significantly different

Table 1. Average yields of maize single crosses between Iowa Stiff Stalk Synthetic (SSS) inbreds and non-SSS (NSSS) inbreds evaluated at three locations in 1992.

Inbreds	SSS1	SSS2	SSS3	SSS4	SSS5	SSS6
NSSS1	10.82	12.45	10.97	10.94	9.98	11.73
NSSS2	11.66	11.02	10.14	10.05	10.02	12.31
NSSS3	11.32	10.68	11.99	11.65	10.04	12.10
NSSS4	12.30	12.54	12.50	11.68	10.28	13.04
NSSS5	12.79	12.81	11.82	12.06	10.50	13.00
NSSS6	11.49	11.84	11.54	11.74	10.30	12.48
NSSS7	11.45	11.23	10.98	10.89	9.81	12.41
NSSS8	10.73	11.00	10.96	11.49	9.94	11.67
NSSS9	9.61	11.08	10.60	10.72	9.97	11.09

LSD ($\alpha = 0.05$) = 1.21

Table 2. Coefficients of coancestry based on 220 RFLPs (above diagonal) and pedigree (below diagonal) among Iowa Stiff Stalk Synthetic (SSS) maize inbreds.

Inbreds	SSS1	SSS2	SSS3	SSS4	SSS5	SSS6
SSS1	1	0.65	0.65*	0.15	0.14	0.65
SSS2	0.61	1	0.47	0.12	0.05	0.63*
SSS3	0.40	0.39	1	0.22	0.26	0.57
SSS4	0.04	0.04	0.19	1	0.48*	0.12
SSS5	0.06	0.05	0.28	0.69	1	0.06
SSS6	0.77	0.76	0.52	0.04	0.06	1

* Significantly different ($\alpha = 0.05$) from corresponding pedigree-based estimate.

Table 3. Coefficients of coancestry based on 220 RFLPs† (above diagonal) and pedigree (below diagonal) among maize inbreds unrelated to Iowa Stiff Stalk Synthetic (NSSS).

Inbreds	NSSS1	NSSS2	NSSS3	NSSS4	NSSS5	NSSS6	NSSS7	NSSS8	NSSS9
NSSS1	1	0.69	0.34	0.11	0.07	0	0.07	0.08	0.02
NSSS2	0.73	1	0.25	0.06	0.03	0	0.03	0.03	0
NSSS3	0.38	0.36	1	0.05	0	0	0.09	0.01	0.06
NSSS4	0	0	0	1	0.77	0.45	0.06	0	0.18
NSSS5	0	0	0	0.88	1	0.42	0.02	0	0.14
NSSS6	0	0	0	0.44	0.50	1	0.03	0	0.05
NSSS7	0	0	0	0	0	0	1	0.03	0.05
NSSS8	0	0	0	0	0	0	0	1	0
NSSS9	0	0	0	0.22	0.25	0.13	0	0	1

† Negative estimates were set equal to zero. None of the RFLP-based estimates was significantly different ($\alpha = 0.05$) from the corresponding pedigree-based estimates.

from $f_{xx'}$ for the (SSS1, SSS3), (SSS2, SSS6), and (SSS4, SSS5) pairwise combinations.

Among NSSS inbreds, coefficients of coancestry based on RFLPs ($f_{yy'}^R$) and based on pedigree ($f_{yy'}$) were zero for several pairs of inbreds (Table 3). The inbred pair (NSSS4, NSSS5) had the highest values of both $f_{yy'}^R$ (0.77) and $f_{yy'}$ (0.88). The average $f_{yy'}^R$ was 0.12 and the average $f_{yy'}$ was 0.11. The two estimates of relationship among NSSS inbreds were highly correlated ($r = 0.98^*$, $\alpha = 0.05$). None of the pairwise combinations of NSSS inbreds had significantly different estimates of $f_{yy'}^R$ and $f_{yy'}$.

Correlations Between Predicted and Observed Single-Cross Yield

Among the different combinations of (i) number of predictor hybrids, (ii) genetic model, and (iii) estimate of genetic relationship, the pooled (across 100 repetitions) correlation between predicted and observed single-cross yields ranged from 0.654 to 0.800 (Table 4). These high correlations indicate that single-cross performance can be predicted effectively using information from related hybrids, especially when one considers that the theoretical maximum value of these correlations was < 1 . The predicted genetic values were not compared to known genetic values, but to observed phenotypic values of the single crosses. The correlation between genetic and phenotypic value is the square root of heritability (h^2). Because h^2 was 0.812 using RFLP relationships and 0.857 using pedigrees, the theoretical maximum values of the correlation between predicted and observed single-cross yields were $(0.812)^{0.5} = 0.901$ and $(0.857)^{0.5} = 0.926$. However, the correlations between predicted and observed yields were obtained from a relatively small data set, and extrapolation of the results to different and larger data sets must be done with caution.

Correlations between predicted and observed single-cross yields increased as the number of predictor hybrids (n) increased. The largest increments in the correlations occurred when n increased from 10 to 15, but the correlations did not increase substantially beyond $n = 25$. However, conclusions regarding the optimum number of predictor hybrids are difficult to make. The optimum number of predictor hybrids is likely to vary among data sets, and larger numbers of predictor hybrids may be needed as the diversity among inbreds in each heterotic group increases.

The correlations between predicted and observed single-cross yields were consistently greater for the full model than for the GCA (dominance variance excluded) model

(Table 4). The correlations ranged from 0.661 to 0.800 for the full model and from 0.654 to 0.793 for the GCA model. Based on empirical data, dominance variance [$V_{D(XY)}$] constitutes approximately 40% of the total genetic variance for maize grain yield (Hallauer and Miranda, 1981). Hence, correlations between predicted and observed single-cross yields were expected to be greater for the full model than for the GCA model. However, the differences in the correlations using the two models were very small. This result may be specific to the data set used in this study. The covariance between hybrids (x, y) and (x', y') includes $V_{D(XY)}$ only when the coefficients of coancestry between x and x' and between y and y' are both > 0 (Eq. [1]). In many instances, particularly for the NSSS inbreds, coefficients of coancestry were near-zero or equal to zero (Table 3). Thus, the contribution of $V_{D(XY)}$ in the full model was small for many of the pairs of single crosses in this study. Also, the estimate of $V_{D(XY)}$ was small compared to $V_{A(X/Y)}$ and $V_{A(Y/X)}$. Consequently, attempts to predict SCA between pairs of inbreds were not successful (results not shown), with $r \leq 0.15$ between predicted and observed SCA. The advantage of the full model over the GCA model in predicting single-cross yield is expected to increase as the degree of relationship among inbreds within heterotic groups increases.

Although the two estimates of genetic relationship were highly correlated, RFLP-based coefficients of coancestry (f^R) gave consistently better predictions of single-cross yield than pedigree-based coefficients of coancestry (f).

Table 4. Pooled (across 100 repetitions) correlations between predicted and observed maize single-cross yields for different combinations of number of predictor hybrids, genetic model, and estimate of coefficient of coancestry.

Model	Number of predictor hybrids	Genetic relationship based on:	
		RFLPs	Pedigree
Full	10	0.695	0.661
	15	0.740†	0.718†
	20	0.770†	0.755†
	25	0.795†	0.786†
	30	0.800†	0.793†
GCA	10	0.688†	0.654
	15	0.733†	0.712†
	20	0.764†	0.749†
	25	0.788†	0.779†
	30	0.793†	0.786†

† Sample correlations were homogeneous at $\alpha = 0.05$. Approximate confidence intervals on the pooled correlation coefficients were ($r \pm 0.015$).

Table 5. Pooled (across 100 repetitions) correlations between predicted and observed maize single-cross yield using arbitrary estimates of genetic variances, full model, and RFLP-based estimates of coefficient of coancestry.

$V_{A(XY)}:V_{A(XY)}:V_{D(XY)}^\dagger$	Number of predictor hybrids				
	10	15	20	25	30
REML ($\approx 15:5:2$)	0.695	0.740‡	0.770‡	0.795‡	0.800‡
1:1:1	0.661	0.727‡	0.761‡	0.772‡	0.789‡
1:1:2	0.646	0.715‡	0.754‡	0.765‡	0.778‡
1:2:1	0.660‡	0.703‡	0.747‡	0.772‡	0.785‡
1:2:2	0.634	0.700‡	0.744‡	0.766‡	0.781‡
2:1:1	0.712	0.760‡	0.776‡	0.793‡	0.789‡
2:1:2	0.694	0.749‡	0.774‡	0.784‡	0.803‡
2:2:1	0.681	0.732‡	0.771‡	0.786‡	0.804‡

$^\dagger V_{A(XY)}$ = testcross additive variance among SSS lines when crossed to NSSS lines; $V_{A(YX)}$ = testcross additive variance among NSSS lines when crossed to SSS lines; and $V_{D(XY)}$ = dominance variance in the SSS \times NSSS cross population.

‡ Sample correlations were homogeneous at $\alpha = 0.05$. Approximate confidence intervals on the pooled correlation coefficients were ($r \pm 0.015$).

Correlations between predicted and observed single-cross yield ranged from 0.688 to 0.800 using f^R and from 0.654 to 0.793 using f . The f^R values may be more accurate estimators of relationship than f if (i) effects of selection and/or genetic drift were present during inbred development, (ii) inbreds were developed using recurrent selection, or (iii) pedigree information was unavailable or unreliable (Bernardo, 1993). The calculation of f assumes no selection during inbred development. However, intense selection for yield and other agronomic characteristics is practiced during inbred development (Hallauer, 1990) such that unequal parental contributions to the inbred progeny may occur. No assumption regarding RFLP-quantitative trait loci (QTL) linkage was made in this study. But if some of the RFLPs were linked to QTL, then f^R would, to some extent, have accounted for deviations from f due to selection. Even when the RFLPs are unlinked to QTL, f^R may provide estimates of genetic relationship among inbreds derived from recurrent selection or among inbreds with unknown or unreliable pedigrees. For inbreds licensed from foundation seed companies or selfed from commercial hybrids, incomplete or unknown pedigrees are not uncommon. In this study, all 15 parental inbreds were developed using selection during inbreeding. Also, SSS6 was developed using recurrent selection whereas pedigree relationships between some ancestral inbreds of NSSS3, NSSS7, and NSSS8 were unknown and assumed zero. The consistently higher correlations between predicted and observed yields using f^R rather f suggests that f^R may have provided better estimates of the true degree of relationship, i.e., probability that inbreds share alleles that are identical by descent. Nevertheless, the magnitude of the correlations between predicted and observed yields obtained using f indicates that pedigree-based coefficients of coancestry may be used to predict single-cross yields when RFLP data are unavailable.

One of the drawbacks of BLUP methodology has been the need for precise estimates of genetic variances (White and Hodge, 1989). But the correlations between predicted and observed yields remained relatively stable across different, arbitrary proportions of genetic variances (Table 5). Some of the correlations obtained using arbitrary proportions of $V_{A(XY)}$, $V_{A(YX)}$, and $V_{D(XY)}$ were higher than those obtained using REML estimates of genetic variances. These results indicate that, at least for the data set used in this study, precise estimates of genetic variances are not necessary for effective prediction of single-cross performance and approximations of these variances may be sufficient.

Bias in the expression for covariance between single crosses is caused by gametic disequilibrium in the parental populations, epistasis, or both (Melchinger, 1988). The simplifying assumptions of gametic equilibrium in the parental populations and negligible epistasis may not be valid in actual breeding populations. Also, BLUP methodology requires the assumption that inbreds are random members of the parental populations, whereas the inbreds used in this study were highly selected lines. But the magnitudes of the correlations between predicted and observed single-cross yield, as well as their robustness with regards to genetic variances, suggest that the simplified model assuming gametic equilibrium and negligible epistasis is sufficient for the prediction of single-cross performance.

Application in Breeding Programs

The BLUP approach used in this study requires information on (i) the relationships among inbreds within each heterotic group and (ii) yield performance of a related set of single crosses. The method has potential application in commercial maize breeding programs because (i) in the future, new inbreds likely will be routinely fingerprinted for RFLPs, at least for plant protection purposes; and (ii) inbreds are invariably tested for yield performance in hybrid combination at some stage during the breeding process. If RFLP data are not available, relationships among inbreds can be determined from pedigree records.

Single crosses are often evaluated in different sets of yield trials, locations, or years. The BLUP method permits the analysis of such unbalanced data. But instead of utilizing all available data (most likely unbalanced) from different yield trials to estimate yields of predictor hybrids, an alternative approach is to select a reference set of inbreds from each heterotic group (e.g., X and Y). The reference inbreds from X may be crossed to the reference inbreds from Y, and the resulting single crosses may be evaluated in the same yield trials at several locations for several years. Single-cross yields and genetic variances can be estimated from this balanced data set. This reference population of predictor hybrids may then be used to predict the performance of future single crosses between inbreds from X and Y.

The results in this study were based on a relatively small data set of 54 single crosses. Further research with larger data sets is needed to assess the utility of BLUP for predicting single-cross yield. Research is currently being conducted on (i) the prediction of single cross yield

using a larger set of single crosses evaluated in different locations and years and (ii) using independent sets of predictor and missing hybrids. Nevertheless, the high correlations between predicted and observed yields obtained in this study suggest that BLUP may be useful for identifying superior single-cross combinations.

ACKNOWLEDGMENTS

The author thanks (i) Pierre Sourdille, Patricia Bonnefoy, Christophe Johnsson, Philippe Leroy, and Zivan Karaman for providing the RFLP data; (ii) Steve Schuetz and Bill Koehring for help in conducting the field experiments; and (iii) anonymous reviewers for very helpful comments and suggestions.

REFERENCES

- Bernardo, R. 1992. Relationship between single-cross performance and molecular marker heterozygosity. *Theor. Appl. Genet.* 83:628–634.
- Bernardo, R. 1993. Estimation of coefficient of coancestry using molecular markers in maize. *Theor. Appl. Genet.* 85:1055–1062.
- Charcosset, A., Lefort-Buson, M., and A. Gallais. 1991. Relationship between heterosis and heterozygosity at marker loci: a theoretical computation. *Theor. Appl. Genet.* 81:571–575.
- Dudley, J.W., M.A. Saghai-Maroo, and G.K. Rufener. 1991. Molecular markers and grouping of parents in maize breeding programs. *Crop Sci.* 31:718–723.
- Falconer, D.S. 1981. *Introduction to quantitative genetics*. 2nd ed. Longman, London.
- Godshalk, E.B., M. Lee, and K.R. Lamkey. 1990. Relationship of restriction fragment length polymorphisms to single-cross hybrid performance in maize. *Theor. Appl. Genet.* 80:273–280.
- Hallauer, A.R. 1990. Methods used in developing maize inbreds. *Maydica* 35:1–16.
- Hallauer, A.R., and J.B. Miranda, Fo. 1981. *Quantitative genetics in maize breeding*. Iowa State Univ. Press, Ames, IA.
- Harville, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* 72:320–340.
- Henderson, C.R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447.
- Henderson, C.R. 1985. Best linear unbiased prediction of non-additive genetic merits in noninbred populations. *J. Anim. Sci.* 60:111–117.
- Henderson, C.R. 1988. Progress in statistical methods applied to quantitative genetics since 1976. p. 85–90. *In* B.S. Weir et al. (ed.) *Proc. 2nd Int. Conf. Quantit. Genet.* Raleigh, NC. 31 May–5 June 1987. Sinauer Assoc., Sunderland, MA.
- Hogan, R.M., and J.W. Dudley. 1992. Cluster analysis of 148 U.S. maize inbreds using RFLP-based estimates of genetic relationship. p. 112–136. *In* *Proc. 28th Annu. Illinois Corn Breeders School*. Champaign, IL. 2–3 Mar. 1992. Univ. Illinois Dep. of Agronomy, Champaign, IL.
- Kempthorne, O. 1969. *An introduction to genetic statistics*. Iowa State Univ. Press, Ames, IA.
- Melchinger, A.E. 1988. Means, variances, and covariances between relatives in hybrid populations with disequilibrium in the parent populations. p. 400–415. *In* B.S. Weir, E.J. Eisen et al. (ed.) *Proc. 2nd Int. Conf. Quantit. Genet.* Raleigh, NC. 31 May–5 June 1987. Sinauer Assoc., Sunderland, MA.
- Melchinger, A.E., M. Lee, K.R. Lamkey, and W.L. Woodman. 1990. Genetic diversity for restriction fragment length polymorphisms: Relation to estimated genetic effects in maize inbreds. *Crop Sci.* 30:1033–1040.
- Melchinger, A.E., M.M. Messmer, M. Lee, W.L. Woodman, and K.R. Lamkey. 1991. Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphisms. *Crop Sci.* 31:669–678.
- Murigneux, A., D. Barloy, P. Leroy, and M. Beckert. 1993. Molecular and morphological evaluation of doubled haploid lines in maize. 1. Homogeneity within DH lines. *Theor. Appl. Genet.* 86:837–842.
- Smith, O.S. 1986. Covariance between line per se and testcross performance. *Crop Sci.* 26:540–543.
- Smith, O.S., J.S.C. Smith, S.L. Bowen, R.A. Tenborg, and S.J. Wall. 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.
- Steel, R.G.D., and J.H. Torrie. 1980. *Principles and procedures of statistics*. 2nd ed. McGraw-Hill, New York.
- White, T.L., and G.R. Hodge. 1989. *Predicting breeding values with applications in forest tree improvement*. Kluwer Academic Publ., Dordrecht, Netherlands.