Theor Appl Genet (1997) 95:655–659

R. Bernardo

RFLP markers and predicted testcross performance of maize sister inbreds

Received: 17 March 1997 / Accepted: 18 April 1997

Abstract Inbreds selfed from the same F_2 or backcross population are referred to as sister inbreds. In some situations, maize (*Zea mays* L.) sister inbreds may not have testcross data available for best linear unbiased prediction (BLUP) of single-cross performance. This study evaluated the usefulness of BLUP and restriction fragment length polymorphism (RFLP)-based coefficients of coancestry (f) in predicting the testcross performance of sister inbreds. Parental contributions (p) were estimated from 70 RFLP loci for 15 inbreds that comprised three sister inbreds selfed from each of five F_2 populations. Estimates of p were subsequently used to calculate RFLP-based f. Grain yield, moisture, and stalk lodging data were obtained for 2265 single crosses tested by Limagrain Genetics in multilocation trials from 1990 to 1995. Performance of the sister inbreds when crossed to several inbred testers was predicted from the performance of the tested single crosses and RFLP-based f. Correlations between predicted and observed performance, obtained with a delete-one cross-validation procedure, were erratic and mostly low for all three traits. Correspondence was poor between ranks for predicted and observed general combining ability of the sister inbreds. The results suggested that the proportion of the genome derived by a sister inbred from a given parental inbred does not solely determine its testcross performance. The failure of BLUP and RFLP-based f to consistently predict testcross performance indicated that actual field testing will continue to be necessary for preliminary evaluation of sister inbreds.

Key words Best linear unbiased prediction · General combining ability · Maize · Restriction fragment length polymorphism · Sister inbreds

Introduction

Best linear unbiased prediction (BLUP) has been found useful for identifying superior maize (*Zea mays* L.) single crosses prior to field testing (Bernardo 1996a). In BLUP, the performance of untested single crosses is predicted from known genetic relationships among parental inbreds and performance data for crosses between inbreds. Performance data are usually available because maize inbreds in one heterotic group [e.g., Iowa Stiff Stalk Synthetic (BSSS)] are testcrossed to one or more inbreds in an opposite heterotic group [e.g., Lancaster Sure Crop (LSC)] during inbred development (Hallauer 1990).

There may be situations, however, wherein pertinent testcross data are not available for a given inbred (Bernardo 1996b). For example, a breeder may be interested in crosses between BSSS and LSC inbreds, but a particular BSSS inbred may have been tested only in combination with non-LSC inbreds. Also, testcross data may not be readily available for inbreds licensed from foundation seed or university sources. Some inbreds may have been tested only in environments that are not within the breeder's target population of environments. Bernardo (1996b) found that when an inbred has not been tested in any hybrid combination, its testcross performance could still be predicted if the coefficient of coancestry (f, Falconer 1981) between the untested inbred and other inbreds with testcross data is known.

Inbreds developed from the same F_2 or backcross population (Hallauer 1990) are referred to as sister inbreds. In calculating f from pedigree records, the actual parental contributions to inbred progeny are

Communicated by P. M. A. Tigerstedt

R. Bernardo (✉)
Limagrain Genetics, 4805 W. Old Church Road, Champaign, IL 61821, USA

Present address:
Department of Agronomy, Purdue University, 1150 Lilly Hall of Life Sciences, West Lafayette, IN 47907-1150, USA

assumed equal to their expected values (Falconer 1981), e.g., 0.5 for an F_2-derived inbred and one of its parents and 0.75 for a BC_1-derived inbred and the recurrent parent. Pedigree-based estimates of f are identical among sister inbreds. Consequently, if the sister inbreds have not been tested in any hybrid combination, their predicted testcross performance will be identical and BLUP cannot be used to identify the best sister inbred to cross to a specific inbred tester.

Bernardo et al. (1997) used restriction fragment length polymorphism (RFLP) markers to study differences in parental contribution to maize sister inbreds. The largest difference was observed for two F_2-derived inbreds that inherited 0.67 and 0.25 of their RFLP alleles, respectively, from the same parent. Variation in testcross performance of sister inbreds may be partly due to variation in the proportion of alleles received by each sister inbred from a given parent. If a parent contributes 0.67 of its alleles to sister inbred A and 0.25 to sister inbred B, then that parent is likely to be more similar to A than to B in terms of testcross performance. Marker-based estimates of f would reflect such variability in parental contribution to sister inbreds and may be useful in BLUP of testcross performance of new, untested sister inbreds.

The objective of this study was to evaluate the usefulness of RFLP-based f in BLUP when sister inbreds have not been tested in hybrid combination within a particular heterotic pattern.

Table 1 RFLP-based parental contribution (p) among five sets of maize sister inbreds

Sister inbred[a]	p[b]	90% confidence interval on p
C1	0.596 A	(0.397, 0.788)
C2	0.694 A*	(0.568, 0.793)
C3	0.435 A	(0.236, 0.644)
E1	0.595 AB	(0.393, 0.760)
E2	0.801 A*	(0.622, 0.944)
E3	0.590 B	(0.358, 0.780)
G1	0.595 A	(0.452, 0.768)
G2	0.248 B*	(0.061, 0.465)
G3	0.667 A	(0.432, 0.867)
H1	0.489 A	(0.353, 0.621)
H2	0.481 A	(0.363, 0.613)
H3	0.774 B*	(0.625, 0.889)
I1	0.763 A*	(0.620, 0.906)
I2	0.351 B	(0.145, 0.622)
I3	0.655 A	(0.469, 0.849)

[a] Sets of sister inbreds (each set selfed from a separate F_2 population) are identified by their heterotic group (C, E, G, H, and I) and a numeric designation; e.g., C1, C2, and C3 were the three sister inbreds selfed from the same F_2 population in heterotic group C

[b] p is the proportion of RFLP alleles inherited from the first parent in the sister inbred's pedigree record. Estimates of p, within each set of three sister inbreds, followed by the same letter were not significantly different at $P = 0.10$. Estimates of p followed by $*$ were significantly different from the expected value of 0.5

Materials and methods

Sister inbreds, RFLP analysis, and coefficient of coancestry

Three sister inbreds selfed from each of five F_2 populations were studied (Table 1). The 15 inbreds were identified by their heterotic group (arbitrarily designated C, E, G, H, and I) and a numeric designation; e.g., C1, C2, and C3 were the three sister inbreds in heterotic group C. Banding patterns were determined for 70 well-dispersed, single-locus probes and restriction digests of genomic DNA from the sister inbreds and their parents. Either *Eco*RI, *Eco*RV, or *Hin*dIII was the restriction enzyme used in combination with each probe. Each of the 70 probe-enzyme combinations was considered to be a RFLP locus and each unique banding pattern a RFLP allele.

Assume inbreds i and j were the parents of inbred k. The symbols $_nS_{ij}$ and $_nS_{ik}$ denoted the proportion of RFLP loci in the nth linkage group ($n = 1$–10 in maize) with alleles common to the inbreds in subscript. The parental contribution of i to k was denoted p. Analysis was limited to RFLP loci for which alleles in k can be traced to either or both of its parents, and the parental contribution of j to k was $(1 - p)$. For each inbred, i was the first parent listed in the pedigree record. The RFLP-based estimate of parental contribution of i to k was:

$$p = (S_{ik} - S_{ij})/(1 - S_{ij})$$

where $S_{ik} = (1/10) \sum {}_nS_{ik}$ and $S_{ij} = (1/10) \sum {}_nS_{ij}$, i.e., the arithmetic average of the ten independent estimates (i.e., corresponding to each linkage group) of proportion of RFLP loci with alleles common to the inbreds in subscript.

Bootstrap resampling (Efron 1981) was used to obtain 90% confidence intervals (CI's) on (1) estimates of p for a sister inbred and its parents and (2) differences in p among the three sister inbreds derived from each F_2 population. From the $_nS_{ij}$ and $_nS_{ik}$ values for the ten linkage groups, ten random $_nS_{ij}$ and $_nS_{ik}$ values (i.e., both values for the same linkage group) were drawn with replacement. Values of S_{ij} and S_{ik} for the ten random samples were used to calculate p. Similarly, values of S_{ij} and S_{ik} for pairs of sister inbreds were used to calculate pairwise differences in p. The resampling procedure was repeated 10 000 times. The 10 000 p values were sorted in ascending order, and the 500th p value represented the lower limit whereas the 9500th p value represented the upper limit of a 90% CI. The difference between estimated and expected p was declared significant when the value of 0.5 was not within the bounds of the CI of p. The difference in p between two sister inbreds was declared significant when the value of zero was not within the bounds of the CI of the difference.

Values of f among inbreds were calculated by tabular analysis (Emik and Terrill 1949). Estimates of p were used to calculate RFLP-based f between each sister inbred and other inbreds in the same heterotic group. For the non-sister inbreds, values of p were determined from pedigrees.

BLUP of testcross performance of sister inbreds

Hybrid performance data were obtained from multilocation yield trials conducted by Limagrain Genetics from 1990 to 1995. Each multilocation yield trial was conducted at 2 to 12 locations in Illinois, Iowa, Wisconsin, Minnesota, and Indiana (USA), Ontario (Canada), southwest France, or northern Italy in a randomized complete block design with one or two replicates at each location.

The resulting data set was highly unbalanced across multilocation trials but, disregarding occasional missing plots at individual locations, balanced within multilocation trials. The performance at each individual location was not considered. Rather, each data point was the average performance of a single cross or check hybrid at several locations in a multilocation yield trial. Data were recorded for grain yield (quintals ha^{-1} at 15.5% moisture) and the percentage of grain moisture and stalk lodging.

Each of the three sister inbreds selfed from a given F_2 population was crossed with $n_T = 5$ to 14 tester inbreds from an opposite heterotic group (Table 2). The testcross performance of sister inbreds belonging to a given heterotic group was predicted from the performance of crosses between other inbreds in the same heterotic group and inbreds in the opposite heterotic group. For example, if F_j was the jth inbred in heterotic group F, the performance of $C1 \times F_j$, $C2 \times F_j$, and $C3 \times F_j$ was predicted from the performance of different crosses in the C × F heterotic pattern. Testcross data of the sister inbreds themselves were not used in BLUP (Bernardo 1996b). Five heterotic patterns were studied: C × F; E × D; G × B; H × A; and I × B. Groups A, B, C, and D were related to the BSSS population, whereas groups E, F, G, H, and I were unrelated to BSSS. The total number of inbreds in each heterotic group was 35 in A, 89 in B, 50 in C, 65 in D, 65 in E, 71 in F, 67 in G, 62 in H, and 37 in I.

Data analyses were performed separately for each of the five heterotic patterns. Assume n single crosses were made between n_1 inbreds from Group 1 and n_2 inbreds from Group 2. The single crosses, along with n_C check hybrids, were evaluated in b different yield trials resulting in q total data points. For each trait, the linear model assuming negligible epistasis was:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_0\mathbf{c} + \mathbf{Z}_1\mathbf{a}_1 + \mathbf{Z}_2\mathbf{a}_2 + \mathbf{Z}\mathbf{d} + \mathbf{e}$$

where: $\mathbf{y} = q \times 1$ vector of observed performance for a given trait (i.e., hybrid by multilocation trial means); $\boldsymbol{\beta} = b \times 1$ vector of yield trial effects; $\mathbf{c} = n_C \times 1$ vector of check hybrid effects; $\mathbf{a}_1 = n_1 \times 1$ vector of testcross additive effects of Group 1 inbreds; $\mathbf{a}_2 = n_2 \times 1$ vector of testcross additive effects of Group 2 inbreds; $\mathbf{d} = n \times 1$ vector of dominance effects; $\mathbf{e} = q \times 1$ vector of residual effects; and $\mathbf{X}$, $\mathbf{Z}_0$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}$ were incidence matrices of 1's and 0's relating $\mathbf{y}$ to $\boldsymbol{\beta}$, $\mathbf{c}$, $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{d}$, respectively. Multilocation yield trial effects were considered fixed, whereas all other effects in the model were considered random.

Solutions to $\boldsymbol{\beta}$, $\mathbf{c}$, $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{d}$ were obtained by solving the mixed-model equations for single-cross performance [see Bernardo (1996a) for details]. Estimates of residual and genetic variances were obtained iteratively by restricted maximum likelihood (Henderson 1985). Assume there were w predictor single crosses that did not have any of the three sister inbreds as one of the parents, i.e., $w = n - 3n_T$. After estimates of $\boldsymbol{\beta}$ and residual and genetic variances were obtained, the performance of sister inbreds when crossed to the n_T testers was predicted as:

$$\mathbf{y}_T = \mathbf{C}_{TP}\,\mathbf{C}_{PP}^{-1}\,\mathbf{y}_P$$

where: $\mathbf{y}_T = 3n_T \times 1$ vector of predicted testcross performance of sister inbreds; $\mathbf{C}_{TP} = 3n_T \times w$ matrix of genetic covariances, calculated with RFLP-based f, between the testcrosses of the sister inbreds and the predictor single crosses; $\mathbf{C}_{PP} = w \times w$ phenotypic variance-covariance matrix among the predictor single crosses; and $\mathbf{y}_P = w \times 1$ vector of average performance, corrected for $\boldsymbol{\beta}$, of the predictor single crosses. The reader is referred to Bernardo (1996a) for details on the calculation of $\mathbf{C}_{TP}$, $\mathbf{C}_{PP}$, and $\mathbf{y}_P$.

For each set of sister inbreds, the covariance between predicted and observed testcross performance was pooled across the n_T testers. Likewise, pooled variances of predicted and observed performance were calculated. The correlation between predicted and observed testcross performance, pooled across testers, was obtained as a measure of the usefulness of RFLP-based f in BLUP of testcross performance of sister inbreds. All the necessary computations were done with a modified version of lgHYPER©, a proprietary program written by the author.

Results and discussion

Estimates of p among sister inbreds

Estimates of p varied within each of the five sets of sister inbreds (Table 1). The range in p was largest for G1, G2, and G3 (range = 0.419) and smallest for E1, E2, and E3 (range = 0.211). Each of the five sets had 1 inbred (i.e., C2, E2, G2, H3, and I1) with an estimate of p that deviated significantly from the expected value of 0.5. Except for the C1, C2, and C3 sister inbreds, significant differences among estimates of p were found within each set.

The variation in p among sister inbreds may be due to random genetic drift (Falconer 1981). The effects of random genetic drift increase as effective population size decreases, and selfing utilizes the smallest possible effective population size (Li 1976). Selection during inbreeding may also contribute to variation in p (St. Martin 1982). Strong selection for yield performance and plant and ear characteristics is usually performed within and/or among families during maize inbred development (Hallauer 1990).

The variation in p among maize sister inbreds implies that pedigree-based f, which assumes that the actual parental contributions to progeny are equal to

Table 2 Correlations, pooled across n_T testers, between predicted and observed testcross performance of maize sister inbreds

Set of sister inbreds	n_T	Predictor hybrids[a]	Correlation between predicted and observed performance		
			Yield	Moisture	Stalk lodging
C1, C2, C3	14	282	0.54*	– 0.64*	0.21
E1, E2, E3	11	526	– 0.31	0.18	0.16
G1, G2, G3	8	584	– 0.24	0.12	– 0.25
H1, H2, H3	5	392	0.21	0.76*	0.14
I1, I2, I3	14	481	– 0.28	0.81*	– 0.20

* Significantly different from zero at $P = 0.05$

[a] Number of predictor hybrids used for BLUP of testcross performance of sister inbreds

their expected values, may lead to erroneous estimates of genetic relationship (Bernardo et al. 1997). E2, G2, H3, and I1 had estimates of p or $(1-p)$ that were greater than the expected value of $p=0.75$ for BC_1-derived inbreds. This result indicated the possibility of obtaining F_2-derived inbreds that, in terms of parental contribution to progeny, are similar to BC_1-derived inbreds.

Predicted versus observed performance of sister inbreds

The pooled (across inbred testers) correlations between predicted and observed performance of sister inbreds were erratic and mostly low (Table 2). The correlations ranged from $r=-0.64^*$ for moisture in C1, C2, and C3 to $r=0.81^*$ for moisture in I1, I2, and I3. Non-negative correlations between predicted and observed performance are expected, and the reasons for the significant negative correlation for moisture in C1, C2, and C3 are unclear.

For yield, the only significant correlation between predicted and observed testcross performance was found for the C1, C2, and C3 set of sister inbreds ($r=0.54^*$, Table 2). For the four other sets of sister inbreds, the correlations between predicted and observed testcross yield were low, non-significant, and often negative. For moisture, the correlations were high for the H1, H2, and H3 ($r=0.76^*$) and the I1, I2, and I3 ($r=0.81^*$) sets of sister inbreds. Such correlations, however, were low and non-significant for the E1, E2, and E3 and the G1, G2, and G3 sets of sister inbreds. For stalk lodging, the correlations between predicted and observed testcross performance were low and non-significant across all five sets of sister inbreds.

These results indicated that RFLP-based estimates of f are not useful for predicting the best sister inbred to cross to a specific inbred tester. Often, breeders are not only interested in predicting the performance of a specific cross but also in the average performance of a sister inbred when crossed to several inbred testers, i.e., general combining ability or GCA. For the C1, C2, and C3 set of sister inbreds, correspondence was observed between ranks for predicted and observed GCA for yield (Table 3). This result was consistent with the significant pooled correlation between predicted and observed testcross yield for the C1, C2, and C3 sister inbreds (Table 2). But for the other sets of sister inbreds, correspondence was poor between predicted and observed GCA for yield (Table 3). For example, the rank for predicted GCA for yield was highest for the E2 sister inbred and lowest for E3. In contrast, the rank for observed GCA for yield was lowest for E2 and highest for E3. For moisture, only H3 had the highest rank for both predicted and observed GCA. For stalk lodging, only C2 and I3 had the highest ranks for both predicted and observed GCA. The inconsistent results across sets of sister inbreds and traits indicated that BLUP with RFLP-based estimates of f is not useful for predicting which sister inbred has the best GCA when crossed with inbreds from an opposite heterotic group.

Table 3 Ranks of maize sister inbreds for predicted and observed (in parentheses) general combining ability performance[a]

Sister inbred	Rank of predicted (observed) performance averaged across n_T testers		
	Yield	Moisture	Stalk lodging
C1	2 (2)	2 (1)	2 (3)
C2	1 (1)	3 (2)	1 (1)
C3	3 (3)	1 (3)	3 (2)
E1	2 (2)	2 (1)	2 (3)
E2	1 (3)	1 (2)	1 (2)
E3	3 (1)	3 (3)	3 (1)
G1	2 (3)	2 (1)	2 (3)
G2	3 (2)	1 (2)	1 (2)
G3	1 (1)	3 (3)	3 (1)
H1	2 (1)	2 (3)	2 (1)
H2	1 (2)	3 (2)	1 (2)
H3	3 (3)	1 (1)	3 (3)
I1	1 (2)	1 (2)	2 (3)
I2	3 (1)	3 (3)	3 (2)
I3	2 (3)	2 (1)	1 (1)

[a] C1, C2, and C3 were crossed with $n_T=14$ testers; E1, E2, and E3 with $n_T=11$ testers; G1, G2, and G3 with $n_T=8$ testers; H1, H2, and H3 with $n_T=5$ testers; and I1, I2, and I3 with $n_T=14$ testers

Bernardo (1996b) found that when untested maize inbreds from one heterotic group were crossed with tested inbreds from an opposite heterotic group, the correlations between predicted and observed performance ranged from -0.03 to 0.69 for yield, 0.62 to 0.90 for moisture, and 0.03 to 0.70 for stalk lodging. The effectiveness of BLUP for untested inbreds increased when the number of tested single crosses in the heterotic pattern was large. The discrepancy between the results in the previous (Bernardo 1996b) and present studies was attributed to the difference in the nature of the inbreds that were considered. In the previous study, the inbreds that were considered untested were not limited to sister inbreds but were derived from a diverse set of F_2 and backcross biparental crosses. In the present study, BLUP was used to assess the performance of sets of sister inbreds, each set being derived from the same F_2 population. The narrower genetic variability that is expected among sister inbreds than among non-sister inbreds may have caused a loss of effectiveness of BLUP and RFLP-based f for predicting testcross performance in the present study.

The low and erratic correlations between predicted and observed testcross performance of sister inbreds suggested that the proportion of the genome derived by a sister inbred from a given parental inbred does not solely determine its testcross performance. Rather, the specific combination of alleles received by a sister inbred from a parental inbred may be an important factor that determines testcross performance. For example, two sister inbreds may both inherit $p = 0.5$ of their genomes from a given parental inbred. Estimates of f determined from either RFLP markers or pedigree records would be identical for both sister inbreds. Yet these sister inbreds may differ in performance if one sister inbred inherits a more favorable set of alleles from one parent than the other sister inbred. Identifying associations with quantitative traits may be necessary for RFLP markers to be an effective tool for predicting the performance of sister inbreds.

In conclusion, RFLP-based f was not useful in BLUP of the performance of untested sister inbreds. Actual field testing will continue to be necessary for preliminary screening of sister inbreds for their testcross performance and GCA.

References

Bernardo R (1996a) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

Bernardo R (1996b) Best linear unbiased prediction of the performance of crosses between untested maize inbreds. Crop Sci 36:872–876

Bernardo R, Murigneux A, Maisonneuve JP, Johnsson C, Karaman Z (1997) RFLP-based estimates of parental contribution to F_2- and BC_1-derived maize inbreds. Theor Appl Genet 94:652–656

Efron B (1981) The jackknife, the bootstrap and other resampling plans. Soc Ind Appl Math, Philadelphia, Pa.

Emik LO, Terrill CE (1949) Systematic procedures for calculating inbreeding coefficients. J Hered 40:51–55

Falconer DS (1981) Introduction to quantitative genetics, 2nd edn. Longman, London

Hallauer AR (1990) Methods used in developing maize inbreds. Maydica 35:1–16

Henderson CR (1985) Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. J Anim Sci 60:111–117

Li CC (1976) First course in population genetics. The Boxwood Press, Pacific Grove, Calif.

St. Martin SK (1982) Effective population size for the soybean improvement program in maturity groups OO to IV. Crop Sci 22:151–152