Theor Appl Genet (1997) 94 : 652–656

**R. Bernardo · A. Murigneux · J. P. Maisonneuve**
**C. Johnsson · Z. Karaman**

# RFLP-based estimates of parental contribution to $F_2$- and $BC_1$-derived maize inbreds

Received: 15 August 1996 / Accepted: 11 October 1996

**Abstract** Selection and genetic drift during inbreeding may cause differences between the actual and expected proportions of the genome derived by an inbred from each of its parents. We used 70 RFLP loci to determine the frequency and magnitude of deviations from the expected parental contribution among $F_2$- and $BC_1$-derived maize (*Zea mays* L.) inbreds. Assuming inbreds $i$ and $j$ were the parents of inbred $k$, the parental contribution of $i$ to $k$ was estimated as $p = (S_{ik} - S_{ij})/(1 - S_{ij})$, where $S_{ik}$ and $S_{ij}$ were the average proportions, across the ten linkage groups in maize, of RFLP loci with alleles common to the inbreds in subscript. Bootstrap confidence intervals (CIs) were obtained for $p$ by re-sampling RFLP similarity for each linkage group. Among 62 $F_2$-derived inbreds, 13 had estimates of $p$ that deviated significantly from the expected value of 0.5. One $F_2$-derived inbred obtained $p = 0.801$ of its genome from a parent. Among 34 $BC_1$-derived inbreds, eight had estimates of $p$ that deviated significantly from the expected contribution of 0.75 from the recurrent parent. Two inbreds, both from the same $BC_1$ population, had an estimated $p \geqslant 0.94$. The results suggested that selection during backcrossing generally favored the recurrent parent over the donor parent. Among the inbreds with significant deviations from the expected $p$, the width of 95% CIs with 70 RFLP loci was $> 0.20$. Inbreds selfed from the same $F_2$ or $BC_1$ population varied in $p$, indicating that coefficients of co-ancestry calculated from pedigree records may give erroneous estimates of genetic relationship.

**Key words** Bootstrap · Genetic drift · Parental contribution · RFLP · Selection · *Zea mays* L.

---

Communicated by P. M. A. Tigerstedt

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

A. Murigneux · J. P. Maisonneuve
C. Johnsson · Z. Karaman
BIOCEM-Groupe Limagrain, 24 Avenue des Landais,
63170 Aubière Cedex, France

## Introduction

Pedigree selection is the most common method used to develop new maize (*Zea mays* L.) inbreds (Hallauer 1990). Two elite inbreds are crossed to form a segregating $F_2$ or $BC_1$ population and new inbreds are developed by selfing, visual selection for plant type and disease resistance among and within ear-to-row progenies, and testing for yield performance when crossed with unrelated inbred testers. The goal in pedigree selection is to combine a high proportion, if not all, of the favorable alleles from both parents in a single new inbred (Ho and Comstock 1980).

Expected parental contributions (i.e., the proportion of the genome derived by an inbred from a parent) with Mendelian inheritance are 0.5 for an $F_2$-derived inbred and either of its parents, 0.75 for a $BC_1$-derived inbred and the recurrent parent, and 0.25 for a $BC_1$-derived inbred and the donor parent. However, selection and genetic drift during selfing may cause differences between actual and expected parental contributions to inbred progeny (St. Martin 1982). With intense selection during selfing, the probability of fixation may range from 0.256 to 0.744 for an allele derived from either parent of an $F_2$ population, 0.531 to 0.905 for an allele contributed by the recurrent parent to a $BC_1$-derived inbred, and 0.095 to 0.469 for an allele contributed by the donor parent to a $BC_1$-derived inbred (Bernardo 1996). Estimates of parental contribution to inbred progeny would give useful insights on the effects of selection and genetic drift in pedigree breeding. Such information would also be extremely useful for germ plasm organization, varietal protection, and choosing pairs of inbreds to form new populations for developing inbreds.

Molecular markers provide a useful means for estimating parental contributions to inbred progeny. Lorenzen et al. (1995) used restriction fragment length polymorphism (RFLP) markers to detect significant differences between estimated and expected parental contributions to progeny in 4 out of 26 soybean [*Glycine max* (L.) Merr.] varieties. But in maize, published reports on the frequency and magnitude of differences between the estimated and expected parental contribution to inbred progeny are currently not available. Lorenzen et al. also used the chi-square test to determine the significance of deviations from expected parental contributions. Such a test requires that markers are independent, an assumption that is violated when markers are linked. Lorenzen et al. lessened the effects of linkage by using only RFLP loci that did not seem to exhibit linkage drag. However, large numbers of markers are necessary to obtain precise estimates of marker-based genetic similarity (Bernardo 1993), and the use of many RFLP loci to saturate the genome would cause linkage between markers on a given chromosome. A non-parametric procedure that allows significance tests for parental contributions even with linkage among markers would be useful.

Our objective in this study was to determine, with 70 RFLP markers, the frequency and magnitude of deviations from the expected parental contribution among 62 $F_2$- and 34 $BC_1$-derived maize inbreds.

## Materials and methods

### Inbreds and RFLP analysis

Sixty two $F_2$- and 34 $BC_1$-derived inbreds and their parents were used in this study. The $F_2$- and $BC_1$-derived inbreds were proprietary lines used by Limagrain Genetics in its breeding programs. Inbreds designated 'S', as well as B73, B37 and A632, were related to the Iowa Stiff Stalk Synthetic (BSSS) population (Hallauer and Miranda 1981), whereas those designated 'N' were non-BSSS inbreds. Patterns of hybridization fragments (bands) were determined for 70 well-dispersed probes and restriction digests of genomic DNA from the inbreds and their parents. Either *Eco*RI, *Eco*RV, or *Hin*dIII was used as the restriction enzyme in combination with each probe. In previous screening and mapping experiments (data not shown), each of the 70 RFLP probes mapped to a single locus. Therefore, each of the 70 probe-enzyme combinations was considered an RFLP locus and each unique banding pattern an RFLP allele. Extraction of DNA, restriction enzyme digestion, gel electrophoresis, Southern blotting, and probe hybridization were as described by Murigneux et al. (1993).

### Estimation of parental contribution to inbred progeny

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 *n*th linkage group ($n = 1$ to 10 in maize) with alleles common to the inbreds in subscript. The parental contribution of *i* to *k* was denoted as *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 $(1 - p)$. When *k* was a $BC_1$-derived inbred, *i* was the recurrent parent whereas *j* was the donor parent. When *k* was an $F_2$-derived inbred, *i* was the first parent listed in the pedigree record of *k*. The RFLP-based estimate of genomic contribution of parent *i* to inbred progeny *k* was:

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

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

The bootstrap re-sampling procedure (Efron 1981) was used to obtain 95% confidence intervals (CIs) on estimates of *p* for a given inbred and its parents. 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 calculated and subsequently used to calculate *p*. This re-sampling procedure was repeated 10 000 times. The 10 000 *p* values were sorted in ascending order, and the 250th *p* value represented the lower limit whereas the 9750th *p* value represented the upper limit of a 95% CI. The significance of the deviation between the estimated and expected parental contribution to progeny was declared when the expected parental contribution (0.5 for an $F_2$-derived inbred and 0.75 for a $BC_1$-derived inbred) was not within the bounds of the CI. All calculations were done with a FORTRAN computer program written by Rex Bernardo.

## Results and discussion

### Variability and distribution of RFLP marker loci

The 70 RFLP loci were distributed across all ten maize chromosomes (Table 1). The number of RFLP loci on each chromosome ranged from four (Chromosomes 8 and 10) to ten (Chromosome 2). All 70 RFLP loci exhibited polymorphism among the inbreds studied. The average number of alleles (i.e., banding patterns) detected per RFLP locus ranged from 4.6 in Chromosome 9 to 7.5 in Chromosome 10.

The bootstrap procedure used for constructing 95% CIs on estimates of *p* required two assumptions: the

**Table 1** Number and distribution across ten chromosomes of RFLP loci used to determine parental contributions among $F_2$- and $BC_1$-derived maize inbreds

| Chromosome | Number of RFLP loci | Average number of alleles per marker |
|---|---|---|
| 1 | 8 | 5.0 |
| 2 | 10 | 5.0 |
| 3 | 7 | 5.3 |
| 4 | 9 | 5.2 |
| 5 | 9 | 4.9 |
| 6 | 7 | 4.9 |
| 7 | 7 | 6.6 |
| 8 | 4 | 5.3 |
| 9 | 5 | 4.6 |
| 10 | 4 | 7.5 |
| Total | 70 | |
| Average | | 5.4 |

data points were (1) independently and (2) identically distributed (Efron 1981). The data points used in this study for bootstrapping were estimates of $p$ for each linkage group (i.e., each chromosome) rather than for each individual RFLP locus. The first assumption was, therefore, satisfied because the linkage groups were by definition independent. The second assumption (that the estimates of $p$ for each chromosome have the same underlying mean and variance) may be met if the markers are dispersed evenly (Press et al. 1992) throughout the genome. The RFLP loci used in this study had been previously selected to provide genome coverage that was as uniform as possible. Consequently, more RFLP loci generally were found on the larger than on the smaller chromosomes (Table 1). The use of well-dispersed RFLP loci in this study was assumed to have adequately satisfied the requirement of identically distributed data points.

## Parental contribution to inbred progeny

Among $F_2$-derived inbreds, RFLP-based estimates of $p$ ranged from 0.248 to 0.801 (Fig. 1), indicating wide variability in the proportion of the genome derived by an inbred from one of the parents in a biparental cross. Estimates of $p$ among $BC_1$-derived inbreds ranged from 0.540 to 0.980 and were generally higher than estimates of $p$ among $F_2$-derived inbreds (Fig. 1). The average parental contribution of a $BC_1$ recurrent parent across the 34 $BC_1$-derived inbreds was 0.765, which was close to the expected value of 0.75. No assumption was made regarding the homogeneity of the distribution of the estimates of $p$ for each $F_2$- and $BC_1$-derived inbred. Hence, tests of skewness and kurtosis of $p$ were not conducted although the distributions of $p$ for the $F_2$- and $BC_1$-derived inbreds seemed symmetric.

The RFLP-based estimates of $p$ for specific inbreds differed significantly ($P = 0.05$) from their expected values (Table 2). Among $F_2$-derived inbreds, significant deviations were found for 13 out of 62 (22%) inbreds. The largest significant deviation was found for the inbred N17a, which derived $p = 0.801$ of its genome from N6b and $(1 - p) = 0.199$ of its genome from N6a. Among $BC_1$-derived inbreds, significant deviations were found for 8 out of 34 (24%) inbreds. Two $BC_1$-derived inbreds had a $p$ less than the expected value of 0.75, whereas six had a $p$ greater than 0.75. This result suggested that any selection during backcrossing generally favored the recurrent parent rather than the donor parent. The largest significant deviation in favor of the recurrent parent was for the inbred N27a, which derived $p = 0.980$ of its genome from N4 (Table 2). N27b, which was selfed from the same $BC_1$ population as N27a, also derived a large proportion ($p = 0.940$) of its genome from N4.

The distributions of $p$ for $F_2$- and for $BC_1$-derived inbreds overlapped (Fig. 1) and, among the 13 $F_2$-derived inbreds with $p$ significantly different from 0.5, seven had estimated $p$ values that did not differ significantly from the expected $p$ value (0.75) for $BC_1$-derived inbreds (Table 2). Also, variability in $p$ was observed among inbreds derived from a given $F_2$ or $BC_1$ population (Table 2). For example, N20a and N20b were both derived from the N18 × N19 cross, and N20a derived most of its genome from N18 ($p = 0.763$). In contrast, N20b derived a much smaller proportion of its genome from N18 ($p = 0.351$) than from N19. Among $BC_1$-derived inbreds, the recurrent parent S1 contributed a smaller proportion of its genome to S10a ($p = 0.569$) than to S10c ($p = 0.860$).

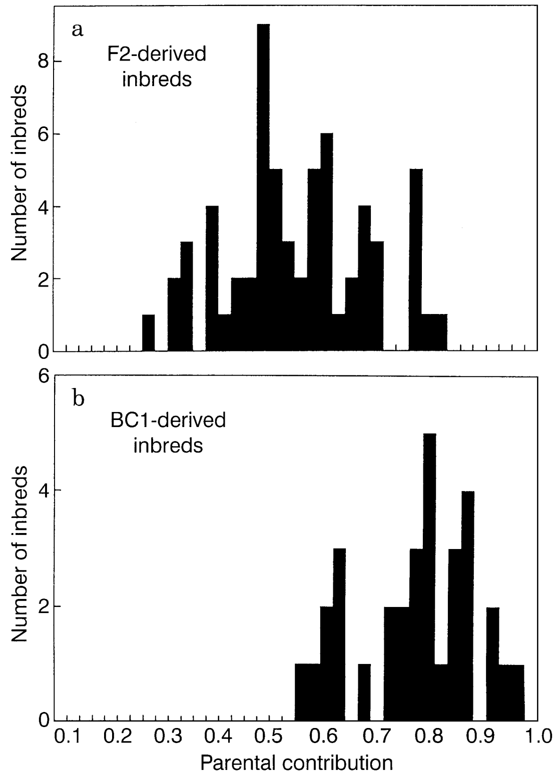

**Fig. 1** Contributions of (a) the first parent listed in the pedigree record to $F_2$-derived inbreds and (b) the recurrent parent to $BC_1$-derived inbreds

The significant deviations between RFLP-based estimates and expected values of $p$ have several important implications. First, they demonstrate that variation in parental contribution to progeny among $F_2$- and among $BC_1$-derived inbreds does exist. Second, the use of Malécot's coefficient of co-ancestry (Falconer 1981), which assumes the actual parental contributions to progeny are equal to their expected values, may lead to erroneous estimates of genetic relationship (Bernardo 1996). Third, $F_2$-derived inbreds may be obtained which, in terms of parental contribution to progeny, are similar to $BC_1$-derived inbreds. Fourth, recovering > 95% of the recurrent parent genome by one generation of backcrossing and selection during selfing seems possible.

**Table 2** Maize inbreds derived from $F_2$ and $BC_1$ populations with significant differences between RFLP-based estimates and expected values of parental contribution to inbred progeny (*p*)

| Parent *i* | Parent *j* | $S_{ij}$[a] | Expected *p* | Inbred *k* | RFLP-based *p* | 95% CI[b] |
|---|---|---|---|---|---|---|
| $F_2$-*derived inbreds* | | | | | | |
| N1 | N2 | 0.266 | 0.5 | N3a[c] | 0.301* | (0.166, 0.426) |
| | | | | N3b | 0.450 | (0.248, 0.661) |
| N4 | N5 | 0.308 | 0.5 | N6a | 0.697* | (0.556, 0.822) |
| | | | | N6b | 0.544 | (0.412, 0.681) |
| N7 | N8 | 0.363 | 0.5 | N9 | 0.776* | (0.567, 0.933) |
| S1 | N10 | 0.375 | 0.5 | N11 | 0.658* | (0.514, 0.807) |
| B73 | B37 | 0.421 | 0.5 | S2 | 0.760* | (0.578, 0.922) |
| N12 | N13 | 0.423 | 0.5 | N14 | 0.302* | (0.145, 0.476) |
| S1 | S3 | 0.428 | 0.5 | S4a | 0.694* | (0.538, 0.808) |
| | | | | S4b | 0.596 | (0.363, 0.820) |
| | | | | S4c | 0.435 | (0.200, 0.683) |
| N7 | N6b | 0.496 | 0.5 | N15a | 0.672* | (0.519, 0.833) |
| | | | | N15b | 0.595 | (0.425, 0.801) |
| B73 | A632 | 0.499 | 0.5 | S6 | 0.773* | (0.666, 0.891) |
| N10 | N6b | 0.543 | 0.5 | N16a | 0.774* | (0.590, 0.908) |
| | | | | N16b | 0.511 | (0.357, 0.675) |
| N6b | N6a | 0.630 | 0.5 | N17a | 0.801* | (0.583, 0.985) |
| | | | | N17b | 0.590 | (0.309, 0.809) |
| | | | | N17c | 0.595 | (0.349, 0.792) |
| N18 | N19 | 0.693 | 0.5 | N20a | 0.763* | (0.589, 0.935) |
| | | | | N20b | 0.351 | (0.115, 0.676) |
| S2 | S7 | 0.694 | 0.5 | S8 | 0.767* | (0.573, 0.943) |
| $BC_1$-*derived inbreds* | | | | | | |
| N4 | N21 | 0.392 | 0.75 | N22 | 0.901* | (0.791, 0.996) |
| N23 | N24 | 0.431 | 0.75 | N25 | 0.843* | (0.758, 0.932) |
| N4 | N26 | 0.443 | 0.75 | N27a | 0.980* | (0.937, 1.000) |
| | | | | N27b | 0.940* | (0.808, 1.000) |
| N10 | N28 | 0.542 | 0.75 | N29 | 0.540* | (0.400, 0.707) |
| N6b | N10 | 0.543 | 0.75 | N30a | 0.844* | (0.757, 0.954) |
| | | | | N30b | 0.916* | (0.788, 1.000) |
| S1 | S9 | 0.613 | 0.75 | S10a | 0.569* | (0.440, 0.703) |
| | | | | S10b | 0.736 | (0.542, 0.932) |
| | | | | S10c | 0.860 | (0.741, 0.989) |

* The RFLP-based estimate of *p* was significantly different from expected *p* at $P = 0.05$

[a] Average proportion (across linkage groups) of RFLP loci with alleles common to Parent *i* and Parent *j*

[b] Lower and upper limits of a 95% bootstrap confidence interval on an RFLP-based estimate of *p*

[c] Inbreds with the same number but different letters were derived from the same $F_2$ or $BC_1$ population

Based on the probability of fixation of an allele, Bernardo (1996) hypothesized that obtaining $BC_1$-derived inbreds with $p > 0.905$ is highly unlikely. But, based on a 95% CI, the inbred N27a had an estimated $p = 0.980$ which was significantly different from 0.905. The reasons for this discrepancy between empirical results and theoretical expectations are unclear. N27a was developed with deliberate selection during selfing against the ear type of the N26 donor parent. Bernardo (1996) assumed additive allelic effects, but strong dominance effects, which are present for maize grain yield (Hallauer and Miranda 1981), increase the probability of fixation of the dominant allele (Ho and Comstock 1980). Also, any form of gametic selection against the donor parent would increase the probability of fixation of recurrent parent alleles.

The proportion of RFLP loci with alleles common to both parents ($S_{ij}$) ranged from 0.248 to 0.706 among $F_2$ populations and from 0.234 to 0.619 among $BC_1$ populations. The concept of parental contribution to progeny loses much of its meaning as the proportion of RFLP loci with alleles common to both parents approaches 1, and any inbreds with $S_{ij} \geqslant 0.75$ were not included in this study. The method used to estimate *p* did not require information on which RFLP loci are polymorphic between the two parents of an inbred, but the precision of estimates of *p* may decrease as $S_{ij}$ approaches 1. On average, the 95% bootstrap CIs were wider among $F_2$-derived inbreds (CI width = 0.412) than among $BC_1$-derived inbreds (CI width = 0.303). This result indicated that the estimates of *p* were more precise for $BC_1$- than for $F_2$-derived inbreds. The correlation between $S_{ij}$ and the width of the 95% CI on the estimate of *p* was significant for $F_2$-derived inbreds ($r = 0.44^*$) but not for $BC_1$-derived inbreds ($r = -0.11$). This result indicated a loss of precision in estimates of *p* as the RFLP-based similarity increased between the parents of an $F_2$-derived inbred.

The bootstrap procedure, applied to linkage groups instead of individual RFLP loci, was useful for obtaining confidence intervals on the estimates of $p$. Narrow confidence intervals are desired for varietal protection purposes, yet the width of 95% CIs with 70 RFLP loci was generally $> 0.20$ (Table 2). With the method for estimating $p$ proposed in this study (i.e., measuring RFLP similarity as the proportion of shared RFLP alleles for each linkage group), using a larger number of RFLP loci should lead to increased precision of the estimates of $p$. The precision of estimates of $p$ obtained with varying numbers of RFLP loci needs further study.

## References

Bernardo R (1993) Estimation of coefficient of coancestry using molecular markers in maize. Theor Appl Genet 85: 1055–1062

Bernardo R (1996) Best linear unbiased prediction of maize single-cross performance given erroneous inbred relationships. Crop Sci 36: 862–866

Efron B (1981) The jackknife, the bootstrap and other re-sampling plans. Soc Ind Appl Math, Philadelphia, Pennsylvania

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

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

Hallauer AR, Miranda JB (1981) Quantitative genetics in maize breeding. Iowa State Univ Press, Ames, Iowa

Ho YT, Comstock RE (1980) Combining superior alleles from two homozygous populations in a cross-fertilizing species. Genet Res 36: 277–287

Lorenzen LL, Boutin S, Young N, Specht JE, Shoemaker RC (1995) Soybean pedigree analysis using map-based molecular markers. I. Tracking RFLP markers in cultivars. Crop Sci 35: 1326–1336

Murigneux A, Barloy D, Leroy P, Beckert M (1993) Molecular and morphological evaluation of doubled-haploid lines in maize. 1. Homogeneity within DH lines. Theor Appl Genet 86: 837–842

Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in FORTRAN. 2nd edn. Cambridge University Press, Cambridge, UK

St. Martin SK (1982) Effective population size for the soybean improvement program in maturity groups 00 to IV. Crop Sci 22: 151–152