Theor Appl Genet (1996) 93: 262–267

R. Bernardo · A. Murigneux · Z. Karaman

Marker-based estimates of identity by descent and alikeness in state among maize inbreds

Received: 20 January 1996 / Accepted: 23 February 1996

Abstract Molecular markers are useful for determining relationships and similarity among inbreds, especially if the proportion of marker loci with alleles common to inbreds *i* and *j* is partitioned into: (1) the probability that marker alleles are identical by descent (${}_{M}f_{ij}$); and (2) the conditional probability that marker alleles are alike in state, given that they are not identical by descent (θ_{ij}). Our objectives were to: develop a method, based on tabular analysis of restriction fragment length polymorphism marker data, for estimating ${}_{M}f_{ij}$, θ_{ij}, and the parental contribution to inbred progeny; validate the accuracy of the method with a simulated data set; and compare the pedigree-based coefficient of coancestry (f_{ij}) and ${}_{M}f_{ij}$ among a set of maize (*Zea mays* L.) inbreds. Banding patterns for 73 probe-enzyme combinations were determined among 13 inbreds. Iterative estimation of ${}_{M}f_{ij}$, θ_{ij}, and the parental contribution to progeny was performed with procedures similar to a tabular analysis of pedigree data. Deviations of ${}_{M}f_{ij}$ from pedigree-based f_{ij} ranged from 0.002 to 0.288, indicating large effects of selection and/or drift during inbreeding for some inbreds. Differences between marker-based estimates and expected values of parental contribution to inbred progeny were as large as 0.205. Results for a simulated set of inbreds indicated that tabular analysis of marker data provides more accurate estimates of ${}_{M}f_{ij}$ and θ_{ij} than other methods described in the literature. Tabular analysis requires the availability of marker data for all the progenitors of each inbred. When marker data are not available for the parents of a given inbred, ${}_{M}f_{ij}$ and θ_{ij} may still be calculated if parental contributions to the inbred are assumed equal to their expectations.

Key words Molecular markers · Maize · Identity by descent · Coefficient of coancestry · Alikeness in state

Introduction

Malécot's (1948) coefficient of coancestry (f_{ij}) is the classic measure of genetic relationship between individuals or inbreds *i* and *j*. The f_{ij} coefficient is the probability that, at a given locus, *i* and *j* have alleles that are identical by descent, i.e., are copies of the same ancestral allele. In plant breeding programs, f_{ij} is useful for: specifying covariances among single crosses in best linear unbiased prediction of maize (*Zea mays* L.) single-cross performance (Bernardo 1996a); selecting parental inbreds to maximize mean performance (Panter and Allen 1995) and genetic variance (Cowen and Frey 1987) in biparental populations; determining effective population sizes (Souza and Sorrells 1989); assigning germ plasm to breeding or heterotic groups (Dudley 1994); and specifying minimum genetic distances for varietal protection (Hunter 1989).

In calculating f_{ij} from pedigree records, the actual parental contributions to inbred progeny are assumed equal to their expected values, e.g., 0.5 for an F_2-derived inbred and one 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. But intense selection for yield and other agronomic traits is practiced during maize inbred development (Hallauer 1990) and deviations from expected parental contributions to inbred progeny may occur. Calculation of f_{ij} is also difficult for inbreds developed from recurrent selection schemes or selfed from commercial hybrids with confidential pedigrees (Bernardo 1993). Lynch (1988), Melchinger et al. (1991), and Bernardo (1993) proposed molecular marker-based estimators (denoted ${}_{M}f_{ij}$) of f_{ij}. Molecular markers allow direct sampling of the genome. Unlike f_{ij} calculated from pedigree records, ${}_{M}f_{ij}$ may account for the effects of deviations from expected parental contributions to progeny due to selection or drift.

Communicated by P. M. A. Tigerstedt

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

A. Murigneux · Z. Karaman
BIOCEM-Groupe Limagrain, 24 Avenue des Landais, 63170 Aubière Cedex, France

Marker-based estimates of genetic similarity are functions of the proportion of either bands or banding patterns (i.e., marker alleles) common to *i* and *j* (Dudley 1994). The proportion of marker loci with alleles common to *i* and *j* is $S_{ij}={}_{M}f_{ij}+(1-{}_{M}f_{ij})\theta_{ij}$, where θ_{ij} is the conditional probability that marker alleles are alike in state (i.e., with the same banding pattern) in *i* and *j*, given that the marker alleles are not identical by descent (Lynch 1988). Unrelated inbreds may have a non-zero θ_{ij} (Melchinger et al. 1991; Bernardo 1993), and S_{ij} is an unbiased estimator of ${}_{M}f_{ij}$ only if θ_{ij} is equal to zero (Lynch 1988) or if the non-zero θ_{ij} can be estimated. Lynch (1988) and Melchinger et al. (1991) proposed estimating θ_{ij} as the average S_{ij} among unrelated individuals or inbreds in a population. However, θ_{ij} values between specific pairs of inbreds may differ from the average θ_{ij}. Bernardo (1993) suggested estimating θ_{ij} as the average S_{ij} between each inbred and an unrelated set of inbreds in a different heterotic group. But the implicit assumption in Bernardo's method that marker allele frequencies are homogeneous among different heterotic groups may not hold true.

Tabular analysis is useful for calculating f_{ij} from pedigree records of inbreds and their progenitors (Emik and Terrill 1949). Similarly, ${}_{M}f_{ij}$ and θ_{ij} may be systematically calculated by tabular analysis if marker data are available for a set of inbreds and all their progenitors. Values of θ_{ij} calculated from tabular analysis account for variability in θ_{ij} among pairs of inbreds and do not rely on the assumption of a homogeneous θ_{ij} among heterotic groups. Our objectives in the present study were to: (1) develop an iterative method, based on tabular analysis of restriction fragment length polymorphism (RFLP) marker data, for calculating ${}_{M}f_{ij}$, θ_{ij}, and the parental contribution to progeny; (2) validate the accuracy of the method using a simulated data set with known true values of ${}_{M}f_{ij}$, θ_{ij}, and the parental contribution to progeny; and (3) compare f_{ij} calculated from pedigree data and ${}_{M}f_{ij}$ among a set of maize inbreds.

Materials and methods

Inbreds and RFLP analysis

Thirteen private Limagrain Genetics maize inbreds, arbitrarily designated L1–L13, were used in this study (Table 1). L1–L5 were progenitor inbreds unrelated to each other. L6–L12 were second-cycle inbreds developed from F_2 or BC_1 crosses among L1–L5. L6 and L7 were both selfed from the (L1×L2)F_2 population whereas L13 was developed from the (L3×L6)F_2 population. Pedigree-based f_{ij} values were calculated by tabular analysis of pedigree records (Emik and Terrill 1949). Patterns of hybridization fragments (bands) were determined using 73 well-dispersed probes and restriction digests of genomic DNA from each of the 13 inbreds. Either *Eco*RI, *Eco*RV, or *Hin*dIII was used as the restriction enzyme in combination with each probe. Each of the 73 probe-enzyme combinations was considered an RFLP locus and each unique banding pattern an RFLP allele. The 73 RFLP loci had been previously screened for consistency of parent-progeny banding patterns, i.e., an RFLP allele in an inbred could be traced to either or both parental inbreds. Extraction of DNA, restriction enzyme digestion, gel electrophoresis, Southern blotting, and probe hybridization were done as described by Murigneux et al. (1993).

Table 1 Pedigrees of maize inbreds, expected contribution of parents to progeny, and marker-based estimates of contribution of parents to progeny

Inbred	Parent 1	Parent 2	Expected contribution		Marker-based contribution [a]	
			Parent 1	Parent 2	Parent 1	Parent 2
L6	L1 [b]	L2	0.500	0.500	0.533	0.467
L7	L1	L2	0.500	0.500	0.689	0.311
L8	L3	L1	0.750	0.250	0.743	0.257
L9	L1	L5	0.750	0.250	0.955	0.045
L10	L4	L1	0.750	0.250	0.690	0.310
L11	L4	L1	0.500	0.500	0.571	0.429
L12	L1	L4	0.750	0.250	0.619	0.381
L13	L3	L6	0.500	0.500	0.541	0.459

[a] Calculated from tabular analysis of data for 73 RFLP loci

[b] L1 to L5 were progenitor inbreds that were unrelated to each other

Tabular analysis of marker data

Assume we wish to calculate the relationship between inbreds *i* and *j* from marker data. Inbreds *k* and *k′* are the parental inbreds of inbred *j*, and *i* may be the same inbred as *k* or *k′*. The symbols ${}_{M}f_{ij}$, ${}_{M}f_{ik}$, ${}_{M}f_{ik'}$, ${}_{M}f_{kj}$, ${}_{M}f_{k'j}$, and ${}_{M}f_{kk'}$ denote marker-based coefficients of coancestry between the inbreds in subscript, whereas S_{ij}, $S_{ik'}$, and $S_{ik'}$ denote the proportions of marker loci with alleles common to the inbreds in subscript. The parental contributions of inbreds *k* and *k′* to inbred *j* are denoted as p_{kj} and $p_{k'j}$, respectively. The marker-based coefficient of coancestry between *i* and *j* is:

$${}_{M}f_{ij}={}_{M}f_{ik}\,p_{kj}+{}_{M}f_{ik'}\,p_{k'j}.$$

The proportion of marker loci with alleles common to *i* and *j* is:

$$\begin{aligned}S_{ij}&=S_{ik}\,p_{kj}+S_{ik'}\,p_{k'j}\\&={}_{M}f_{ij}+[p_{kj}(1-{}_{M}f_{ik})\theta_{ik}+p_{k'j}(1-{}_{M}f_{ik'})\theta_{ik'}]\\&={}_{M}f_{ij}+(1-{}_{M}f_{ij})\theta_{ij},\end{aligned}\qquad(1)$$

Assume a total of N inbreds. Using a Fortran program written by R. Bernardo, tables of ${}_{M}f_{ij}$ and θ_{ij} values were obtained as follows:

(1) Sort the inbreds from the oldest progenitors to the inbreds developed most recently. Thus, each parent is listed before any of its progeny.

(2) Set up an N×N table for ${}_{M}f_{ij}$ values and a second N×N table for θ_{ij} values. The ith inbred in the list of N inbreds corresponds to the ith row and column in each table.

(3) Put 1.0 (complete inbreeding) in each diagonal element of the ${}_{M}f_{ij}$ table and zero in each diagonal element of the θ_{ij} table. For the progenitor inbreds that are unrelated to each other, enter ${}_{M}f_{ij}$ values equal to zero in the corresponding off-diagonal elements in the ${}_{M}f_{ij}$ table and the observed S_{ij} values in the corresponding off-diagonal elements in the θ_{ij} table.

(4) Begin with the first row in the ${}_{M}f_{ij}$ and θ_{ij} tables. In the first iteration, assume equal parental contributions to progeny, i.e., $p_{kj}=p_{k'j}=0.5$. Starting with the first inbred (e.g., jth column) with known parentage in the first row (e.g., ith row):

(a) if both ${}_{M}f_{ik}$ and ${}_{M}f_{ik'}$ are equal to zero, enter an ${}_{M}f_{ij}$ value equal to zero in the ith row and jth column of the ${}_{M}f_{ij}$ table and the observed S_{ij} value in the ith row and jth column of the θ_{ij} table; or

(b) if either or both ${}_{M}f_{ik}$ and ${}_{M}f_{ik'}$ are greater than zero, calculate ${}_{M}f_{ij}=S_{ij}-[p_{kj}(1-{}_{M}f_{ik})\theta_{ik}+p_{k'j}(1-{}_{M}f_{ik'})\theta_{ik'}]$ and enter this value in the ${}_{M}f_{ij}$ table. Calculate $\theta_{ij}=[p_{kj}(1-{}_{M}f_{ik})\theta_{ik}+p_{k'j}(1-{}_{M}f_{ik'})\theta_{ik'}]/(1-{}_{M}f_{ij})$ and enter this value in the θ_{ij} table.

Write the same ${}_{M}f_{ij}$ and θ_{ij} values down the first column, i.e., ${}_{M}f_{ji}={}_{M}f_{ij}$ and $\theta_{ji}=\theta_{ij}$. Continue calculating ${}_{M}f_{ij}$ and θ_{ij} values across the first row.

(5) Repeat Step 4 until values have been calculated for all the rows in both the ${}_{M}f_{ij}$ and θ_{ij} tables. Always do one row at a time.

(6) For the inbreds with known parents, calculate parental contributions to progeny as $p_{kj}=({}_{M}f_{kj}-{}_{M}f_{kk'})/(1-{}_{M}f_{kk'})$ and $p_{k'j}=({}_{M}f_{k'j}-{}_{M}f_{kk'})/(1-{}_{M}f_{kk'})$.

(7) Repeat Steps 4 to 6 using updated values of p_{kj} and $p_{k'j}$ in each iteration. Continue until ${}_{M}f_{ij}$ values do not differ between successive iterations.

Numeric example

Assume the following S_{ij} values among four hypothetical inbreds (A, B, C, and D):

	B	C	D
A	0.3206	0.7290	0.8628
B		0.5916	0.4578
C			0.8662

Inbreds A and B are unrelated to each other. Inbred C was developed from the A×B cross, whereas D was developed from the A×C cross. Calculate values of ${}_{M}f_{ij}$ and θ_{ij} as follows:

(1) Arrange the inbreds so that each parent is listed before any of its progeny. Hence, the inbreds can be arranged as either (A, B, C, D) or (B, A, C, D), but not as (A, C, B, D) or (A, B, D, C).

(2) Enter values of 1.0 in the diagonal elements of the ${}_{M}f_{ij}$ table and zero in the diagonal elements of the θ_{ij} table. Because A and B are unrelated, enter a corresponding value of zero in the ${}_{M}f_{ij}$ table and the observed S_{ij} value (0.3206) in the θ_{ij} table:

${}_{M}f_{ij}$ table:

	A	B	C	D
A	1	0		
B	0	1		
C			1	
D				1

θ_{ij} table:

	A	B	C	D
A	0	0.3206		
B	0.3206	0		
C			0	
D				0

(3) Values of ${}_{M}f_{ij}$ and θ_{ij} for *i*=A and *j*=C are calculated as follows. The parental inbreds of *j* are *k*=A and *k'*=B. Assume equal parental contributions of $p_{kj}=p_{k'j}=0.5$ in the first iteration. For (A, C), calculate: ${}_{M}f_{ij}=\{S_{ij}-[p_{kj}(1-{}_{M}f_{ik})\theta_{ik}+p_{k'j}(1-{}_{M}f_{ik'})\theta_{ik'}]\}=\{0.7290-[0.5(1-1)0+0.5(1-0)0.3206]\}=0.5687$; and $\theta_{ij}=[p_{kj}(1-{}_{M}f_{ik})\theta_{ik}+p_{k'j}(1-{}_{M}f_{ik'})\theta_{ik'}]/(1-{}_{M}f_{ij})=[0.5(1-1)0+0.5(1-0)0.3206]/(1-0.5687)=0.3717$. Copy these ${}_{M}f_{ij}$ and θ_{ij} values in the corresponding elements below the diagonal. Finish calculating the rest of the ${}_{M}f_{ij}$ and θ_{ij} values in the first row (i.e., A, D), then begin calculating ${}_{M}f_{ij}$ and θ_{ij} values in the second [(B, C) and (B, D)] and third (C, D) rows. After one iteration, the ${}_{M}f_{ij}$ and θ_{ij} values are:

${}_{M}f_{ij}$ table:

	A	B	C	D
A	1	0	0.5687	0.7827
B	0	1	0.4313	0.2174
C	0.5687	0.4313	1	0.7861
D	0.7827	0.2174	0.7861	1

θ_{ij} table:

	A	B	C	D
A		0.3206	0.3717	0.3688
B	0.3206		0.2819	0.3072
C	0.3717	0.2819		0.3746
D	0.3688	0.3072	0.3746	

Table 2 Pedigrees of simulated inbreds used for method validation, true values of contribution of parents to progeny, and marker-based estimates of contribution of parents to progeny

Inbred	Parent 1	Parent 2	True contribution		Marker-based contribution [a]	
			Parent 1	Parent 2	Parent 1	Parent 2
V5	V1 [b]	V2	0.804	0.196	0.805	0.195
V6	V2	V3	0.596	0.404	0.595	0.405
V7	V1	V3	0.501	0.499	0.497	0.503
V8	V3	V4	0.305	0.695	0.298	0.701
V9	V5	V6	0.811	0.189	0.808	0.192
V10	V7	V8	0.404	0.596	0.404	0.596
V11	V6	V8	0.698	0.302	0.702	0.298
V12	V7	V11	0.504	0.496	0.500	0.500

[a] Calculated from tabular analysis of simulated data for 1000 RFLP loci

[b] V1 to V4 were simulated inbreds that were unrelated to each other

(4) The parents of *j*=D are *k*=A and *k'*=C. The marker-based parental contribution of A to D is calculated as $p_{kj}=({}_{M}f_{kj}-{}_{M}f_{kk'})/(1-{}_{M}f_{kk'})=(0.7827-0.5687)/(1-0.5687)=0.4961$. The following p_{kj} and $p_{k'j}$ values are then used in the second iteration:

Inbred *j*	Parent *k*	Parent *k'*	p_{kj}	$p_{k'j}$
C	A	B	0.5687	0.4313
D	A	C	0.4961	0.5039

(5) Continue calculating ${}_{M}f_{ij}$ and θ_{ij} until the difference between ${}_{M}f_{ij}$ values in successive iterations is small (e.g., $<1\times10^{-10}$) for all pairs of inbreds.

Method validation

The accuracy of ${}_{M}f_{ij}$ and θ_{ij} estimates, obtained from tabular analysis of marker data, was evaluated using a set of simulated inbreds with known true values of ${}_{M}f_{ij}$, θ_{ij}, p_{kj}, and $p_{k'j}$. Each of 12 simulated inbreds, designated V1–V12, had 1000 independent marker loci (Table 2). V1–V4 were progenitor inbreds unrelated to each other. These progenitor inbreds were assigned random marker alleles at each locus. V5–V8 were derived from simulated crosses among V1–V4, V9–V11 from crosses among V5–V8, and V12 from V7×V11. The true values of the parental contribution to inbred progeny varied among crosses (Table 2). Values of S_{ij} among the 12 simulated inbreds were obtained. True values of ${}_{M}f_{ij}$, θ_{ij}, p_{kj}, and $p_{k'j}$ among simulated inbreds were compared to the corresponding estimates obtained from tabular analysis of S_{ij}.

Results and discussion

Values of S_{ij} among the 13 maize inbreds ranged from 0.370 (L2, L8) to 0.973 (L1, L9) (Table 3). Across a wide range of inbred relationships (i.e., f_{ij} ranging from 0 to 0.750, Table 4), the correlation between S_{ij} and pedigree-based f_{ij} was 0.90 ($P<0.01$) whereas the correlation between S_{ij} and ${}_{M}f_{ij}$ was 0.95 ($P<0.01$). But, as indicated in Equation 1, S_{ij} was an upwardly-biased estimator of ${}_{M}f_{ij}$ and the bias increased as ${}_{M}f_{ij}$ approached zero. Despite this bias, S_{ij} will always be higher among related inbreds than among unre-

Table 3 RFLP-based estimates of similarity [a] (above diagonal) and conditional probability that alleles are alike in state, given that they are not identical by descent (below diagonal), among 13 maize inbreds

Inbred	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13
L1		0.384	0.521	0.425	0.397	0.712	0.808	0.644	0.973	0.603	0.671	0.781	0.712
L2	0.384		0.411	0.452	0.397	0.671	0.575	0.370	0.384	0.411	0.466	0.479	0.534
L3	0.521	0.411		0.438	0.411	0.493	0.507	0.877	0.507	0.521	0.466	0.479	0.767
L4	0.425	0.452	0.438		0.534	0.384	0.438	0.384	0.411	0.822	0.753	0.644	0.452
L5	0.397	0.397	0.411	0.534		0.425	0.384	0.411	0.425	0.548	0.479	0.425	0.425
L6	0.384	0.384	0.493	0.384	0.425		0.712	0.534	0.685	0.493	0.548	0.589	0.726
L7	0.384	0.384	0.507	0.438	0.384	0.394		0.603	0.795	0.493	0.644	0.644	0.685
L8	0.521	0.370	0.521	0.384	0.411	0.470	0.490		0.616	0.479	0.548	0.575	0.753
L9	0.397	0.384	0.507	0.411	0.397	0.376	0.391	0.503		0.575	0.658	0.753	0.685
L10	0.425	0.411	0.521	0.425	0.548	0.387	0.394	0.425	0.405		0.657	0.685	0.562
L11	0.425	0.466	0.466	0.425	0.479	0.396	0.467	0.460	0.415	0.370		0.671	0.562
L12	0.425	0.479	0.479	0.425	0.425	0.385	0.423	0.476	0.405	0.424	0.400		0.644
L13	0.558	0.404	0.493	0.452	0.425	0.493	0.525	0.515	0.534	0.494	0.469	0.514	

[a] Proportion (out of 73) of RFLP loci with alleles common to two inbreds

Table 4 Pedigree-based (above diagonal) and marker-based [a] (below diagonal) coefficients of coancestry among 13 maize inbreds

Inbred	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13
L1		0	0	0	0	0.500	0.500	0.250	0.750	0.250	0.500	0.750	0.250
L2	0		0	0	0	0.500	0.500	0	0	0	0	0	0.250
L3	0	0		0	0	0	0	0.750	0	0	0	0	0.500
L4	0	0	0		0	0	0	0	0	0.750	0.500	0.250	0
L5	0	0	0	0		0	0	0	0.250	0	0	0	0
L6	0.533	0.467	0	0	0		0.500	0.125	0.375	0.125	0.250	0.375	0.500
L7	0.689	0.311	0	0	0	0.525		0.125	0.375	0.125	0.250	0.375	0.250
L8	0.257	0	0.743	0	0	0.122	0.221		0.188	0.063	0.125	0.188	0.438
L9	0.955	0	0	0	0.045	0.495	0.663	0.229		0.188	0.375	0.563	0.188
L10	0.310	0	0	0.690	0	0.173	0.163	0.095	0.286		0.500	0.375	0.063
L11	0.429	0	0	0.571	0	0.252	0.332	0.163	0.415	0.457		0.500	0.125
L12	0.619	0	0	0.381	0	0.332	0.383	0.190	0.586	0.453	0.452		0.188
L13	0.349	0.218	0.541	0	0	0.459	0.336	0.492	0.324	0.133	0.174	0.267	

[a] Calculated from tabular analysis of data for 73 RFLP loci. Convergence at 1×10^{-10} was attained after 33 iterations

lated inbreds if θ_{ij} is constant. However, θ_{ij} varied among the progenitor inbreds (Table 3) and S_{ij} did not always reflect the relative degree of relatedness especially among distantly related inbreds. For example, S_{ij} was higher (0.534) between the unrelated progenitor inbreds L4 and L5 than between the related inbreds L6 and L10 (0.493), L7 and L10 (0.493), and L8 and L10 (0.479).

Because the five progenitor inbreds (L1–L5) were unrelated to each other, the S_{ij} values among these inbreds were direct measures of θ_{ij}. Values of θ_{ij} among the progenitor inbreds ranged from 0.384 (L1, L2) to 0.534 (L4, L5) (Table 3). Among all inbreds, θ_{ij} ranged from 0.370 (L2, L8 and L10, L11) to 0.558 (L1, L13). The variability in θ_{ij} indicated that ${}_{M}f_{ij}$ may be inaccurate if θ_{ij} is estimated as the average S_{ij} among unrelated inbreds (Lynch 1988; Melchinger et al. 1991).

Pedigree-based estimates of f_{ij} among related inbreds ranged from 0.063 between L8 and L10 to 0.750 between BC_1-derived inbreds and their recurrent parents (Table 4). In contrast, ${}_{M}f_{ij}$ ranged from 0.045 (L5, L9) to 0.955 (L1, L9) (Table 4). In this study, convergence of ${}_{M}f_{ij}$ estimates was attained at 1×10^{-10} after 33 iterations of the tabular analysis of marker data. By definition, ${}_{M}f_{ij}$ may range from 0 to 1, but the procedure used previously by Bernardo (1993) for estimating θ_{ij} led to ${}_{M}f_{ij}$ values that were negative 18% of the time in that study. In contrast, negative values of ${}_{M}f_{ij}$ were not obtained in the present study, suggesting that θ_{ij} estimated by tabular analysis provided better estimates of ${}_{M}f_{ij}$ than the procedure suggested previously by Bernardo (1993). The correlation between ${}_{M}f_{ij}$ and pedigree-based f_{ij} among related inbreds was 0.89 (P<0.01). This high correlation indicated a strong degree of association across a wide range of f_{ij} and ${}_{M}f_{ij}$ values, but large deviations between f_{ij} and ${}_{M}f_{ij}$ were found for specific pairs of inbreds. Deviations of ${}_{M}f_{ij}$ from f_{ij} ranged from 0.002 (L6, L11) to 0.288 (L7, L9) with an average deviation of 0.073. Deviations of ${}_{M}f_{ij}$ from f_{ij} were also large for L1 and L7 (0.189), L2 and L7 (0.189), L1 and L9 (0.205), and L5 and L9 (0.205). The pair of inbreds with the lowest f_{ij} did not have the lowest ${}_{M}f_{ij}$ and vice-versa. Unfortunately, significance tests for the deviations between f_{ij} and ${}_{M}f_{ij}$ have not been developed. Bernardo (1993) used the jackknife procedure to test the significance of deviations of ${}_{M}f_{ij}$ from f_{ij}. Re-sampling procedures, such as the jackknife and

Table 5 Coefficients of coancestry, calculated from tabular analysis of marker data (above diagonal) and using average similarity among unrelated inbreds as the estimate of alikeness in state (below diagonal), among 12 simulated inbreds

Inbred	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12
V1		0	0	0	0.805 [a]	0	0.497	0	0.654	0.207	0	0.246
V2	0		0	0	0.195	0.595	0	0	0.269	0	0.415	0.211
V3	0	0		0	0	0.405	0.503	0.298	0.083	0.378	0.384	0.444
V4	0	0	0		0	0	0	0.702	0	0.419	0.210	0.101
V5	0.772 [b]	0.060	0	0		0.120	0.402	0	0.831	0.169	0.088	0.240
V6	0	0.535	0.316	0	0.074		0.197	0.114	0.290	0.150	0.737	0.467
V7	0.562	0	0.567	0	0.426	0.125		0.143	0.365	0.490	0.183	0.592
V8	0	0	0.390	0.741	0	0.059	0.248		0.022	0.654	0.378	0.258
V9	0.621	0.146	0.133	0	0.822	0.252	0.371	0.066		0.166	0.215	0.286
V10	0.313	0	0.457	0.497	0.224	0.090	0.552	0.696	0.199		0.299	0.401
V11	0	0.326	0.356	0.197	0.089	0.720	0.170	0.339	0.214	0.270		0.591
V12	0.285	0.092	0.463	0.159	0.248	0.423	0.585	0.290	0.286	0.422	0.584	

[a] Convergence at 1×10^{-10} was attained after 31 iterations

[b] Calculated as $(S_{ij}-\bar{\theta})/(1-\bar{\theta})$, where S_{ij}=proportion (out of 1000) of marker loci with common alleles between two inbreds and $\bar{\theta}$=average S_{ij} among unrelated inbreds=0.419

bootstrap, assume the data points are identically and independently distributed (Efron 1981). However, linkage among RFLP markers causes non-independence of the data points and significance tests based on re-sampling procedures were not used in the present study.

The large deviations of $_{M}f_{ij}$ from f_{ij} for some pairs of inbreds resulted from differences between marker-based estimates and expected values of parental contribution to inbred progeny (Table 1). Deviations of marker-based estimates of parental contribution from their expected values were less than 0.05 for the parents of L6, L8, and L13. In contrast, L9 was developed from the (L1×L5)×L1 BC_1 population but the marker-based estimate of parental contribution to L9 was 0.955 for L1 and 0.045 for L5. L6 and L7 were both developed from the (L1×L2)F_2 population but, based on marker data, L1 contributed a larger proportion of its genome to L7 (0.689) than to L6 (0.533). Bernardo (1996b) considered the probability of fixation of an allele with selection, and found that the following ranges of parental contribution to inbred progeny are likely: (1) 0.256 to 0.744 for an F_2-derived inbred and either of its parents; (2) 0.531 to 0.905 for a BC_1-derived inbred and its recurrent parent; and (3) 0.095 to 0.469 for a BC_1-derived inbred and its donor parent. Except for the parental contributions of L1 and L5 to L9, the marker-based estimates of parental contribution were within the expected ranges for F_2- and BC_1-derived inbred progeny.

Among the simulated set of inbreds, the deviations of estimated $_{M}f_{ij}$ (Table 5) from known true values of f_{ij} among related inbreds ranged from 0.0005 to 0.0102 with an average of 0.0038. Differences between marker-based estimates and true values of parental contribution to progeny were also small (Table 2). Differences between estimates and true values of θ_{ij} among related inbreds ranged from 0.0004 to 0.0114 with an average of 0.0036 (data not shown). The results for the simulated set of inbreds indicated that tabular analysis of marker data provides accurate estimates of $_{M}f_{ij}$, θ_{ij}, and parental contribution to progeny. For comparison, f_{ij} among the simulated inbreds was also estimated as $_{M}f_{ij}{}^{*}=(S_{ij}-\bar{\theta})/(1-\bar{\theta})$, where $\bar{\theta}$ was the average S_{ij} among unrelated inbreds (Lynch 1988; Melchinger et al. 1991). Differences between $_{M}f_{ij}{}^{*}$ (Table 5) and true values of f_{ij} in the simulated data set ranged from 0.0020 to 0.1364 with an average of 0.0468. This average deviation was 13% of the average true f_{ij} among the simulated inbreds. The large differences between $_{M}f_{ij}{}^{*}$ and true f_{ij} for some pairs of inbreds indicated that estimating θ_{ij} as the average S_{ij} among unrelated inbreds may not be desirable. Differences between $_{M}f_{ij}$ and $_{M}f_{ij}{}^{*}$ would depend on the variability of θ_{ij} among pairs of inbreds. In the simulated set of inbreds, true θ_{ij} among unrelated inbreds ranged from 0.321 to 0.503 with an average of 0.419. The accuracy of $_{M}f_{ij}{}^{*}$ as a measure of true f_{ij} is expected to increase as variability of θ_{ij} among inbreds decreases.

The direct use of S_{kj} (or $S_{k'j}$) as an estimate of f_{kj} (or $f_{k'j}$), using only markers that are polymorphic between the parents (*k* and *k'*) of inbred *j*, has been suggested (Lorenzen et al. 1995). This procedure is valid only if *k* and *k'* are unrelated and may be of limited use among inbreds developed by second-cycle breeding, wherein related elite inbreds are crossed to form new base populations for inbred development. Also, such a procedure may be useful with parent-offspring relationships but not with more complex types of relationship among inbreds. In contrast, tabular analysis of marker data is a general procedure for estimating coefficient of coancestry and accounts for the genetic relationship between the parents of an inbred. Knowledge of the parents of inbreds is necessary but no assumptions are made regarding parental contributions to progeny. Marker data are needed for all the progenitors of each inbred. If marker data are available only for *k*, *k'*, and *j*, but not for more distant progenitors of *j*, an estimate of $\theta_{kk'}$ is needed for calculating $_{M}f_{kj}$ and $_{M}f_{k'j}$. Tabular analysis of marker data may also be possible if, whenever marker data are unavailable, parental contributions to inbred progeny are assumed equal to their expectations.

References

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

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 maize single-cross performance given erroneous inbred relationships. Crop Sci 36: (Jul–Aug issue)

Cowen NM, Frey KJ (1987) Relationship between genealogical distance and breeding behavior in oats (*Avena sativa* L.) Euphytica 36:413–424

Dudley JW (1994) Comparison of genetic distance estimators using molecular marker data. In: Analysis of molecular marker data. Proc Joint Symp Am Soc Hort Sci/Crop Sci Soc Am, 5–6 Aug 1994, Corvallis, Oregon, USA, pp 3–7

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

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

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

Hunter RB (1989) ASTA approach on minimum distance. Corn Sorghum Industry Res Conf 44:193–195

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

Lynch M (1988) Estimation of relatedness by DNA fingerprinting. Mol Biol Evol 5:584–599

Malécot G (1948) Les mathématiques de l'hérédité. Masson et Cie, Paris

Melchinger AE, Messmer MM, Lee M, Woodman WL, Lamkey KR (1991) Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci 31:669–678

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

Panter DM, Allen FL (1995) Using best linear unbiased predictions to enhance breeding for yield in soybean: I. Choosing parents. Crop Sci 35:397–405

Souza E, Sorrells ME (1989) Pedigree analysis of North American oat cultivars released from 1951 to 1985. Crop Sci 29:595–601