Theor Appl Genet (1993) 85:1055–1062

THEORETICAL AND APPLIED GENETICS

Estimation of coefficient of coancestry using molecular markers in maize*

R. Bernardo

Limagrain Genetics, 4805 W. Old Church Road, Champaign, IL 61821, USA

Received August 24, 1992; Accepted September 3, 1992
Communicated by A. R. Hallauer

Summary. The coefficient of coancestry (f_{AB}) between individuals A and B is the classical measure of genetic relationship. f_{AB} is determined from pedigree records and is the probability that random alleles at the same locus in A and B are copies of the same ancestral allele or identical by descent (*ibd*). Recently, the proportion of molecular marker variants shared between A and B (S_{AB}) has been used to measure genetic relationship. But S_{AB} is an upwardly-biased estimator of f_{AB}, especially between distantly-related lines. f_{AB}, S_{AB}, and adjusted (to remove bias) estimates of molecular marker similarity (f^M_{AB}) were compared. RFLP banding patterns at 46 probe-restriction enzyme combinations were obtained for 23 maize inbred lines derived from the Iowa Stiff Stalk Synthetic (BSSS) maize (*Zea mays* L.) population, and for 4 non-BSSS lines. f^M_{AB} was estimated as $f^M_{AB} = [S_{AB} - \frac{1}{2}(\delta_{A.} + \delta_{B.})]/[1 - \frac{1}{2}(\delta_{A.} + \delta_{B.})]$, where $\delta_{A.}$ (or $\delta_{B.}$) was the average proportion of RFLP variants shared between inbred A (or inbred B) and the non-BSSS lines. The average f_{AB} among 253 pairwise combinations of BSSS lines was 0.212, whereas the average S_{AB} was 0.397. The average f^M_{AB} was 0.162, indicating that the upward bias in S_{AB} was effectively removed. S_{AB} and f_{AB} were significantly different ($\alpha = 0.05$) in 76.3% of the comparisons, whereas 24.9% of the f^M_{AB} values differed significantly from f_{AB}. The latter result suggests that selection and/or drift were present during inbred line development and that f_{AB} may not be an accurate measure of the true proportion of *ibd* alleles between two lines. Cluster analyses based on S_{AB} and f^M_{AB} grouped lines according to pedigree, although several exceptions were noted. The presence of shared molecular marker variants between unrelated lines must be considered when setting S_{AB}-based minimum distances for varietal protection. Under simplified conditions, more than 250 molecular marker loci are necessary to obtain sufficiently precise estimates of coefficient of coancestry using molecular markers.

Key words: Coefficient of coancestry – Molecular markers – RFLPs – Genetic distance – *Zea mays* L.

Introduction

Knowledge of genetic relationships among individuals or inbred lines greatly enhances the organization of breeding programs. Malécot's (1948) coefficient of coancestry (f_{AB}) between individuals A and B is the classical measure of genetic relationship. f_{AB} is the probability that a random allele from A and a second random allele at the same locus in B are copies of the same ancestral allele [i.e., identical by descent (*ibd*)]. In animal breeding, f_{AB} is used routinely in obtaining best linear unbiased predictions of breeding value. In plant breeding, f_{AB} is useful for determining effective population sizes in allogomous crops (Souza and Sorrells 1989), selecting parental lines to optimize genetic variances in subsequent inbred generations (Cowen and Frey 1987), and assigning germ plasm to different breeding or heterotic groups. f_{AB} could also be used to specify minimum genetic distances for varietal protection (Hunter 1989).

At least three reasons have limited the use of f_{AB} as a measure of relationship among maize (*Zea mays* L.) inbred lines. First, the calculation of f_{AB}

* A contribution from Limagrain Genetics, a Group Limagrain company

assumes the absence of selection or drift during inbred line development. But intense selection for yield and other agronomic traits is practiced during inbreeding (Hallauer 1990) such that unequal parental contributions to the inbred progeny may occur. Second, widely used inbred lines have been developed from recurrent selection programs. Procedures for determining f_{AB} under recurrent selection schemes are not well developed. Third, pedigree information for a particular line may be unreliable or unavailable. For example, default f_{AB} values equal to zero have to be assigned between inbred lines selfed from different commercial hybrids with confidential pedigrees.

Molecular markers (MM) such as isozymes and restriction fragment length polymorphisms (RFLPs) have been used to assign maize inbred lines to different heterotic groups (Godshalk et al. 1990; Smith et al. 1990; Messmer et al. 1991; Melchinger et al. 1991; Dudley et al. 1991). Different types of similarity measures were used in these studies, but the general basis for determining genetic relationship was the proportion of MM variants shared between pairs of lines. However, the proportion of shared MM variants is an upwardly-biased estimator of f_{AB} (Cox et al. 1985; Lynch 1988). The amount of bias depends on the level of true relationship and is greater for distantly related individuals than for closely related individuals. Estimation of this bias is difficult in natural populations (Lynch 1988), but less difficult in breeding populations wherein genetic backgrounds or heterotic groups are clearly defined. Two undesirable consequences of this bias are: (1) the direct interpretation of the proportion of shared MM variants in terms of f_{AB} is not possible, and (2) the direct comparison of genetic relationship measured in different experiments using different sets of isozyme or RFLP markers cannot be carried out. For example, Dudley et al. (1991) found that maize inbred lines B73 and B84 had similar variants at 37 out of 66 (56%) MM loci. In a study using 144 RFLP loci, Messmer et al. (1991) found that the unrelated lines B73 and Mo17 had common variants at 47% of the MM loci. The difference in the proportion of shared MM variants between these two pairs of inbred lines was not significant ($P = 0.11$). Thus, one might conclude that genetic variance between B73 and B84 is expected to be only slightly less than, if not equal to, that between B73 and Mo17. However, maize breeders know from the genetic backgrounds of the lines that much greater variability is expected between B73 and Mo17 than between B73 and B84.

Studies comparing f_{AB} and the proportion of shared MM variants as estimates of relatedness are limited. In soybean [*Glycine max* (L.) Merr], Cox et al. (1985) obtained rank correlations between f_{AB} and genetic similarity indices (determined from 20 isozyme and morphological loci) ranging from 0.24 to 0.60 ($P < 0.05$). For 12 pairs of maize inbred lines, Melchinger et al. (1991) obtained a rank correlation of 0.71 ($P < 0.05$) between f_{AB} and the proportion of homomorphic MM loci. Smith et al. (1990) found a correlation of 0.90 ($P < 0.05$) between f_{AB} and MM similarity among 37 maize inbred lines. In these studies, the absolute values of f_{AB} and the proportion of shared MM variants were not compared. The objective of the study reported here was to compare (1) Malécot's coefficient of coancestry, (2) the proportion of shared RFLP variants, and (3) the adjusted (to remove bias) proportion of shared RFLP variants as measures of genetic relationship among a set of maize inbred lines.

Materials and methods

Genetic material

Out of the 27 lines studied 23 were related to the Iowa Stiff Stalk Synthetic (BSSS) population (Table 1). BSSS was developed by Dr. G. F. Sprague in 1933–1934 by intercrossing 16 lines selected for superior stalk quality (Sprague 1946). The BSSS lines can be divided into B14, B37, N28, and B73 families. B14 and B37 were derived from the original cycle of BSSS, whereas B73 was developed from the fifth cycle following half-sib recurrent selection. N28 was developed from the Nebraska version of BSSS (Henderson 1984). For classification purposes, B84 (derived from the seventh cycle of recurrent selection in BSSS) was considered a B73-derived line. The 4 non-BSSS lines and their genetic backgrounds (in parentheses) were C103 (Lancaster Sure Crop), OH43 (Lancaster Sure Crop), T8 (Jarvis Prolific), and WF9 (Wilson Farm Reid). These non-BSSS lines were used for determining the proportion of MM variants shared between unrelated lines.

RFLP analyses

RFLP data were obtained from R. M. Hogan and J. W. Dudley of the University of Illinois. The patterns of the hybridization fragments (bands) were determined using 46 well-dispersed probes (mostly genomic clones) and restriction digests of genomic DNA from each of the 27 inbred lines. One of three restriction enzymes (*Eco*RI, *Hin*dIII, or *Sst*I) was used in combination with each probe. Each probe-enzyme combination was considered to be a RFLP locus, and each unique banding pattern a RFLP variant. All chromosome arms, except the short arms of Chromosomes 6, 7, and 10, were marked. DNA extraction, restriction enzyme digestion, gel electrophoresis, Southern blotting, and probe hybridization were performed by Native Plants Incorporated (Salt Lake City, Utah). Procedures described by Helentjaris et al. (1986) were followed.

Relationship measures

Coefficients of coancestry (f_{AB}) were calculated for each pair of inbred lines as described by Falconer (1981). Equal contributions of each parent and complete homozygosity of lines were assumed. The coancestry coefficient among B14, B37, and N28 was $f_{AB} = 1/16$, because 16 unrelated lines were intercrossed to form the original BSSS popultion. Likewise, $f_{AB} = 1/16$ for an advanced cycle inbred (B73 or B84) and an inbred from the

Table 1. Inbred lines studied

Line	Pedigree
B73-type lines	
B73	Iowa Stiff Stalk Synthetic (BSSS), Cycle 5
FR2352	87.5% B73, 10.94% B14, 1.56% '41.2504B'
FR1141	96.875% B73, 3.125% KY288
FR564	75% B73, 25% H99
B84	BSSS, Cycle 7
FR1128	50% B73, 50% B84
FR618	50% B73, 50% P3183
FR986	50% B73, 18.75% OH43, 6.25% V3, 18.75% WF9, 6.25% A171
B14-type lines	
B14	BSSS, Cycle 0
FR902	68.75% B14, 3.125% M42, 18.75% OH43, 3.125% ND203, 6.25% A171
FR460	43.75% B14, 6.25% '41.2504B', 25% B73, 25% H99
FR2200	93.55% B14, 6.05% M42, 0.39% ND203
FR632	93.75% B14, 6.25% M42
A635	87.5% B14, 12.5% ND203
A665	82.03% B14, 17.97% ND203
CM105	75% B14, 25% V3
A641	50% B14, 50% ND203
FR31	43.75% B14, 6.25% ND203, 37.5% OH43, 12.5% A171
N28-type lines	
N28	Nebraska version of BSSS
FR15A	75% N28, 25% Reid × Krug
FR088	50% N28, 37.5% B14, 12.5% V3
B37-type lines	
B37	BSSS, Cycle 0 selection
FR4A	50% B37, 50% '33–16'
Other lines	
C103	Lancaster Sure Crop
OH43	50% OH40B (Lancaster Sure Crop), 50% W8
T8	Jarvis Prolific
WF9	Wilson Farm Reid

original population. As suggested by Melchinger et al. (1991), f_{AB} between B73 and B84 was derived using Eqs. (3.11) and (3.12) of Falconer (1981).

The proportion of RFLP loci with shared variants (denoted S_{AB}) was determined for each pair of inbred lines. Thus, $S_{AB} = 0$ if none of the 46 RFLP loci had variants common to both inbreds, whereas $S_{AB} = 0.5$ if the inbreds had identical variants at 23 of the 46 RFLP loci. The expectation of S_{AB} is (Cox et al. 1985; Lynch 1988)

$$S_{AB} = f'_{AB} + (1 - f'_{AB})\delta_{AB} \quad (1)$$

where f'_{AB} = the true probability that inbred A and inbred B carry alleles at the same locus that are identical by descent (*ibd*); and δ_{AB} = the average probability that a variant from one parent of inbred A and a variant from one parent of inbred B are alike in state (*ais*), given that they are not *ibd*. Thus, δ_{AB} is a function of the proportion of variants common to unrelated lines and is specific for each pair of lines (Lynch 1988). Two inbred lines may carry specific variants that occur in above-average frequencies in unrelated lines. In this situation, δ_{AB} for this pair of lines is expected to be greater than the average δ_{AB} among all possible pairs of lines. f'_{AB} is equal to $f_{AB} + \Delta_{AB}$, where f_{AB} is the coefficient of coancestry determined from pedigree records and Δ_{AB} is a deviation due to selection, drift, or other factors that change allele frequency.

Equation (1) indicates that the expected proportion of MM variants shared between two inbred lines is equal to the sum of (1) the probability that the lines carry *ibd* alleles and (2) the probability that the lines carry variants that are not *ibd* but still *ais* (Lynch 1988). The second term in Equation (1) represents the upward bias when S_{AB} is used to estimate f'_{AB}. The contribution of δ_{AB} to S_{AB} increases as f'_{AB} approaches zero. Thus, S_{AB} is a more biased estimate of f'_{AB} between distantly related lines than between closely related lines. Also, the bias in S_{AB} increases as the proportion of MM variants shared between unrelated lines increases.

Rearranging Eq. (1) gives

$$f'_{AB} = (S_{AB} - \delta_{AB})/(1 - \delta_{AB}).$$

In practice, the value of δ_{AB} for each pair of inbred lines is unknown. By definition, alleles shared between unrelated lines are *ais* but not *ibd*. The average proportion of RFLP variants common with C103, OH43, T8, and WF9 was calculated for each BSSS line. δ_{AB} was estimated as

$$\delta_{AB} = (\delta_{A.} + \delta_{B.})/2 \quad (2)$$

where $\delta_{A.}$ (or $\delta_{B.}$) was the average proportion of RFLP variants shared between inbred A (or inbred B) and the unrelated lines C103, OH43, T8, and WF9.

FR31 and FR902 were related to OH43, whereas FR986 was related to OH43 and WF9 (Table 1). Hence, S_{AB} between FR31 and OH43 cannot be used directly in estimating $\delta_{A.}$. $f_{FR31,OH43} = 0.375$, and an adjusted proportion of RFLP variants shared between these two lines was obtained as $(S_{FR31,OH43} - 0.375)/(1 - 0.375)$. This adjusted estimate was then used in obtaining $\delta_{A.}$. Similar adjustments were made for FR902 and FR986.

MM-based estimates of relatedness were obtained as

$$f^{M}_{AB} = [S_{AB} - \tfrac{1}{2}(\delta_{A.} + \delta_{B.})]/(1 - \tfrac{1}{2}(\delta_{A.} + \delta_{B.})]$$

where f^{M}_{AB} = estimate of f'_{AB} based on MM data.

Data analyses

Differences between f^{M}_{AB} and f_{AB} and between S_{AB} and f_{AB} were tested for significance ($\alpha = 0.05$) for each of the 253 pairwise combinations of the 23 BSSS lines. Jackknife estimates of standard errors (Efron 1981) of f^{M}_{AB} and S_{AB} were used in z-tests. The frequencies of significant differences between S_{AB} and f_{AB} and between f^{M}_{AB} and f_{AB} were determined. Rank correlation coefficients among f_{AB}, S_{AB}, and f^{M}_{AB} were calculated.

Cluster analyses were performed using $(1 - f_{AB})$, $(1 - S_{AB})$, and $(1 - f^{M}_{AB})$ as measures of genetic dissimilarity. Negative f^{M}_{AB} values were set to zero. Cluster analyses were done using the PC-compatible statistical package CSS:Statistica (StatSoft, Tulsa, Okla.), and distances between clusters were computed using unweighted pair-group average linkages.

Results and Discussion

RFLP diversity

A total of 247 variants was found across the 46 RFLP loci. The number of variants at each marker locus ranged from 2 to 12, with an average of 5.4. Among the

247 variants, 24 (9.7%) were unique to the non-BSSS lines (C103, OH43, T8, or WF9).

Measures of genetic relationship

Among the 23 BSSS-derived lines, Malécot's coefficient of coancestry (f_{AB}) ranged from 0.014 (FR31, FR4A) to 0.969 (B73, FR1141) (Table 2). The average f_{AB} across all 253 pairwise combinations of BSSS lines was 0.212. The average f_{AB} within BSSS families ranged from 0.481 (for B73-derived lines) to 0.568 (for B14-derived lines) (Table 3). Average f_{AB} between the B73, B14, N28, and B37 families ranged from 0.036 to 0.135.

The proportion of RFLP variants shared between BSSS lines (S_{AB}) ranged from 0.174 (for 5 different pairs of lines) to 0.935 (A635, A665), with an average of 0.397 among all possible pairs of lines (Table 2). The average S_{AB} within BSSS families ranged from 0.471 for N28-derived lines to 0.544 for B37-type lines (Table 3). Average S_{AB} between lines of different families ranged from 0.266 to 0.383.

S_{AB} is affected by the proportion of alleles that are not identical by descent (*ibd*) but alike in state (*ais*), and is a biased estimator of f_{AB}. Average S_{AB} of each BSSS line with the 4 non-BSSS lines (C103, OH43, T8, and WF9; Table 4) was used to obtain adjusted estimates of RFLP similarity (f^{M}_{AB}). The average S_{AB} with non-BSSS lines ranged from 0.217 (FR2352) to 0.364 (B84), with an average of 0.278 and standard error of 0.028. Average S_{AB} with BSSS lines was 0.275 for WF9, 0.276 for OH43, 0.280 for T8, and 0.283 for C103.

The average coefficient of coancestry estimated from RFLP data (f^{M}_{AB}) was 0.162 (Table 4). (A635, A665) had the largest f^{M}_{AB} (0.909). Negative estimates of f^{M}_{AB} may be obtained with the procedure for calculating f^{M}_{AB} and occurred in 17.8% of the 253 pairwise combinations of BSSS lines. f^{M}_{AB} is zero or positive by definition, and negative estimates of f^{M}_{AB} may be interpreted as being zero. When negative f^{M}_{AB} estimates are set to zero, the average f^{M}_{AB} between lines is 0.173. Average f^{M}_{AB} within families was smaller than the average f_{AB} or S_{AB} within families (Table 3). Average f^{M}_{AB} ranged from 0.274 for N28-type lines to 0.360 for B73-type lines.

Negative estimates of f^{M}_{AB} were not unexpected because f_{AB} was near zero between lines derived from different families. A weakness in the model for estimating f^{M}_{AB} may also have led to negative estimates. In particular, an upward bias in the estimator of δ_{AB} (i.e., average probability that a variant from one parent of inbred A and a variant from one parent of inbred B are *ais* but not *ibd*; see eq. 2) for specific lines would have led to negative f^{M}_{AB} estimates. This situation may have occurred with (1) B37 in combination with B73-type lines and (2) FR15A in combination with

Table 2. Coefficient of coancestry (f_{AB}, above diagonal) and proportion of shared molecular marker variants (S_{AB}, below diagonal) between Iowa Stiff Stalk Synthetic lines

Inbred	B73	FR2352	FR1141	FR564	FR902	B84	FR1128	FR618	FR986	FR460	FR2200	FR632	A635	A665	CM105	A641	FR31	FR088	FR15A	N28	FR4A	B14	B37
B73		0.882	0.969	0.750	0.043	0.265	0.633	0.500	0.500	0.277	0.058	0.059	0.055	0.051	0.047	0.031	0.027	0.055	0.047	0.063	0.031	0.063	0.063
FR2352	0.870		0.854	0.661	0.113	0.239	0.560	0.441	0.441	0.417	0.153	0.154	0.144	0.135	0.123	0.082	0.072	0.092	0.046	0.062	0.031	0.164	0.062
FR1141	0.761*	0.652*		0.727	0.042	0.257	0.613	0.484	0.484	0.269	0.057	0.057	0.053	0.050	0.045	0.030	0.026	0.053	0.045	0.061	0.030	0.061	0.061
FR564	0.717	0.674	0.652		0.032	0.199	0.475	0.375	0.375	0.278	0.044	0.044	0.041	0.038	0.035	0.023	0.021	0.041	0.035	0.047	0.023	0.047	0.047
FR902	0.609*	0.543*	0.587*	0.522*		0.043	0.043	0.021	0.092	0.312	0.702	0.646	0.605	0.570	0.516	0.359	0.705	0.279	0.032	0.043	0.021	0.688	0.043
B84	0.522*	0.543*	0.500*	0.500*	0.457*		0.633	0.133	0.133	0.094	0.058	0.059	0.055	0.051	0.047	0.031	0.027	0.055	0.047	0.063	0.031	0.063	0.063
FR1128	0.674	0.630	0.609	0.565	0.543*	0.652		0.316	0.316	0.186	0.058	0.059	0.055	0.051	0.047	0.031	0.027	0.055	0.047	0.063	0.031	0.063	0.063
FR618	0.630	0.543	0.522	0.587*	0.500*	0.457*	0.478*		0.250	0.139	0.029	0.029	0.027	0.026	0.023	0.016	0.014	0.027	0.023	0.031	0.016	0.031	0.031
FR986	0.370	0.435	0.326*	0.348	0.413*	0.348*	0.304	0.261		0.139	0.029	0.029	0.027	0.026	0.039	0.016	0.139	0.035	0.023	0.031	0.016	0.031	0.031
FR460	0.413	0.413	0.391	0.435*	0.370	0.348*	0.391*	0.370*	0.370*		0.424	0.425	0.396	0.372	0.340	0.227	0.198	0.191	0.032	0.043	0.021	0.453	0.043
FR2200	0.435*	0.500*	0.370*	0.435*	0.413*	0.391*	0.391*	0.370*	0.609*	0.413		0.881	0.826	0.774	0.702	0.470	0.413	0.380	0.044	0.058	0.029	0.936	0.058
FR632	0.391*	0.457*	0.391*	0.391*	0.348*	0.435*	0.391*	0.348*	0.457*	0.435	0.630*		0.820	0.769	0.703	0.469	0.410	0.381	0.014	0.059	0.029	0.938	0.059
A635	0.370*	0.435*	0.348*	0.348*	0.370*	0.348*	0.348*	0.326*	0.609*	0.457	0.630*	0.674*		0.945	0.656	0.500	0.500	0.355	0.041	0.055	0.027	0.875	0.055
A665	0.326*	0.391*	0.326*	0.304*	0.348*	0.348*	0.326*	0.261*	0.565*	0.413	0.565*	0.674	0.935		0.615	0.500	0.473	0.333	0.038	0.051	0.026	0.820	0.051
CM105	0.283*	0.326*	0.239*	0.261*	0.304*	0.326*	0.326*	0.261*	0.457*	0.391	0.500*	0.609	0.717	0.696		0.375	0.328	0.523	0.035	0.047	0.023	0.750	0.047
A641	0.283*	0.326*	0.326*	0.283*	0.326	0.348*	0.304*	0.304*	0.457*	0.348	0.565	0.565	0.696*	0.717*	0.522*		0.250	0.203	0.023	0.031	0.016	0.500	0.031
FR31	0.283*	0.304*	0.261*	0.326*	0.370*	0.304*	0.283*	0.304*	0.565*	0.326	0.370	0.413	0.565	0.543	0.413	0.370		0.178	0.021	0.027	0.014	0.438	0.027
FR088	0.413*	0.457*	0.348*	0.457*	0.304	0.457*	0.457*	0.370*	0.283*	0.413*	0.391	0.457	0.370	0.348	0.478	0.283	0.304		0.393	0.523	0.027	0.406	0.055
FR15A	0.348*	0.326*	0.326*	0.304*	0.326*	0.348*	0.348*	0.304*	0.239*	0.261*	0.326*	0.196*	0.217*	0.174*	0.239*	0.239*	0.261*	0.348		0.750	0.023	0.047	0.047
N28	0.348*	0.326*	0.348*	0.348*	0.326*	0.304*	0.326*	0.326*	0.261*	0.217*	0.348*	0.217*	0.196*	0.174*	0.239*	0.283*	0.174*	0.348*	0.717		0.031	0.063	0.063
FR4A	0.283*	0.304*	0.304*	0.370*	0.326*	0.391*	0.304*	0.239*	0.326*	0.304*	0.370*	0.304*	0.370*	0.370*	0.348*	0.326*	0.326*	0.304*	0.304*	0.174*		0.031	0.500
B14	0.348*	0.413*	0.304*	0.283*	0.391*	0.326*	0.370*	0.283*	0.500*	0.413	0.652*	0.674*	0.783	0.717	0.674	0.609	0.435	0.413	0.261*	0.239*	0.261*		0.063
B37	0.196*	0.217*	0.196*	0.196*	0.261*	0.261*	0.217*	0.174*	0.283*	0.304*	0.326*	0.304*	0.326*	0.348*	0.370*	0.283*	0.283*	0.283*	0.378*	0.283*	0.543	0.283*	

* S_{AB} estimate significantly different ($\alpha = 0.05$) from f_{AB}. S_{AB} was determined using 46 probe-enzyme combinations. Average $\sigma_{error} = 0.070$

Table 3. Average measures of relatedness within and between Iowa Stiff Stalk Synthetic families

	f_{AB}	S_{AB}	f_{AB}^{M}
Within families			
B73-type lines	0.481	0.540	0.360
B14-type lines	0.568	0.519	0.337
N28-type lines	0.555	0.471	0.274
B37-type lines	0.500	0.544	0.331
Between families			
B73- vs B14-type lines	0.070	0.383	0.146
B73- vs N28-type lines	0.048	0.349	0.101
B73- vs B37-type lines	0.039	0.266	−0.045
B14- vs N28-type lines	0.135	0.289	0.023
B14- vs B37-type lines	0.036	0.320	0.034
N28- vs B37-type lines	0.041	0.286	−0.014

B14-type lines, wherein f_{AB}^{M} values were mostly negative. Four unrelated lines were used in estimating δ_{AB}, and perhaps a larger number of unrelated lines should have been used. An alternative procedure is to estimate δ_{AB} using several distantly related lines from BSSS with known f_{AB} values. If f_{AB} is fixed (e.g., $f_{AB} = 1/16$ among BSSS lines derived from Cycle 0), Eq. (1) can be rearranged to estimate the proportion of shared variants between unrelated lines: $\delta_{AB} = (S_{AB} - f_{AB})/(1 - f_{AB})$. But this approach may be circuitous because it requires detailed pedigree information that may not be available. Also, δ_{AB} estimated using this procedure will be biased by the effects of selection and/or drift on f_{AB}.

The discrepancy between average f_{AB} (0.212) and S_{AB} (0.397) across all pairs of BSSS lines indicates that S_{AB} is a biased estimator of f_{AB}. In contrast, the average f_{AB}^{M} (0.162) indicates that the upward bias in S_{AB} was effectively removed. These results were supported by the frequency of significant ($\alpha = 0.05$) differences between S_{AB} and f_{AB} and between f_{AB}^{M} and f_{AB}. Among the 253 pairwise combinations of BSSS lines, 76.3% had S_{AB} values that differed significantly from f_{AB}. In contrast, 24.9% of the f_{AB}^{M} estimates differed significantly from f_{AB}. Three inbred lines (FR902, FR986, and B14) accounted for more than half of the significant differences between f_{AB}^{M} and f_{AB}. The frequencies of significant differences between S_{AB} and f_{AB} and between f_{AB}^{M} and f_{AB} may not be directly comparable because of different standard errors of S_{AB} and f_{AB}^{M}. The average jackknife standard error was 0.070 for S_{AB} and 0.095 for f_{AB}^{M}. When the standard error of f_{AB}^{M} was substituted as a conservative estimate of the standard error of S_{AB}, the frequency of significant differences between S_{AB} and f_{AB} was 69.2%. Although the average f_{AB}^{M} (0.162) was close to the average f_{AB} (0.212), within-family f_{AB}^{M} as consistently smaller than within-family f_{AB}. In addition, the average f_{AB}^{M} between B73- and B14-type lines (0.146) was more than twice the corresponding average f_{AB} (0.070).

Selection and drift during inbred line development are possible reasons for the (1) discrepancies in within- and between-family values of f_{AB}^{M} and f_{AB} and (2) substantial proportion (24.9%) of f_{AB}^{M} estimates significantly different from f_{AB}. If selection, drift, and/or other forces that change allele frequency during inbred line development were present, f_{AB} may not be an accurate measure of the true proportion of *ibd* alleles (f'_{AB}) between two lines. Deviation of f'_{AB} from f_{AB} may be especially large for lines developed through recurrent selection because of the compounded effects of successive cycles of selection and the intermating of a limited number of selected lines. The most notable example of divergent f_{AB} and f_{AB}^{M} was with the combination (B73, FR1141). Four backcrosses using B73 as the recurrent parent were used in developing FR1141, and $f_{AB} = 0.969$. In contrast, f_{AB}^{M} between B73 and FR1141 was 0.680 and S_{AB} was 0.761. Although sampling error in f_{AB}^{M} ($\sigma_{error} = 0.085$ for B73, FR1141) may have been involved, the large discrepancy between f_{AB} and f_{AB}^{M} for (B73, FR1141) suggests significant effects of selection and/or drift.

The rank correlation between f_{AB}^{M} and f_{AB} ($r_c = 0.65$; $P < 0.05$) was slightly greater than that between S_{AB} and f_{AB} ($r_c = 0.60$; $P < 0.05$). The rank correlation between f_{AB}^{M} and S_{AB} was $r_c = 0.97$;($P < 0.05$). As expected, cluster analyses based on f_{AB} values grouped the B73-, B14-, N28-, and B37-type lines in separate clusters (Fig. 1). Based on f_{AB}^{M} values, lines of the B73, B14, N28, and B37 families were grouped separately, except for FR902, FR088, and FR986. FR902, a line with 69% B14 germ plasm, was clustered with the B73- instead of the B14-type lines. FR088, which had 50% N28 and 38% B14 germ plasm, was loosely grouped with the B14-type lines. FR986, which had 50% B73 germ plasm, was clustered with the B14- instead of the B73-type lines. Based on S_{AB} values, lines of the different BSSS families were grouped in separate clusters, except for FR902, FR088, FR986, and FR460. FR460, which had 44% B14 and 25% B73 germ plasm, was clustered with the B73-type lines.

Results from other studies

Results from other studies reveal an upward bias when S_{AB} is used to estimate f_{AB} (Table 5). From the data of Godshalk et al. (1990), average f_{AB} among ten pairs of B73-related lines was 0.536 whereas average S_{AB} was 0.702. When the average proportion of RFLP variants shared between the BSSS lines and seven Lancaster Sure Crop-type lines was used to remove bias, the average f_{AB}^{M} was 0.513. Reduction in upward bias in the resulting f_{AB}^{M} estimates was also demonstrated using data of Melchinger et al. (1991). Among six pairs

Table 4. Adjusted (to remove bias) estimates of molecular marker similarity (f^{M}_{AB}) and proportion of marker variants shared with non-Iowa Stiff Stalk Synthetic lines

Inbred	B73	FR2352	FR1141	FR564	FR902	B84	FR1128	FR618	FR986	FR460	FR2200	FR632	A635	A665	CM105	A641	FR31	FR088	FR15A	N28	FR4A	B14	B37	C103	OH43	T8	WF9	AIS[a]
B73	1.000	0.829	0.680*	0.609	0.460*	0.307	0.547	0.489	0.160*	0.209	0.235	0.191	0.131	0.088	0.019	0.040	−0.011	0.203	0.108	0.121	−0.035	0.134	−0.096*	0.174	0.326	0.304	0.217	0.255
FR2352		1.000	0.546*	0.560	0.386*	0.356	0.500	0.385	0.265	0.229	0.341	0.296	0.241	0.197	0.101	0.121	0.045	0.281	0.101	0.114	0.023	0.239	−0.040	0.152	0.261	0.261	0.196	0.217
FR1141			1.000	0.521*	0.432*	0.278	0.459	0.341	0.105*	0.182	0.150	0.194	0.104	0.092	−0.037	0.101	−0.037	0.118	0.081	0.124	0.000	0.079	−0.092	0.196	0.283	0.261	0.261	0.250
FR564				1.000	0.319*	0.252	0.377	0.411	0.105*	0.215	0.212	0.167	0.073	0.030	−0.042	0.011	0.020	0.240	0.019	0.094	0.061	0.019	−0.130*	0.217	0.391	0.326	0.261	0.299
FR902					1.000	0.189	0.348*	0.288*	0.196	0.126	0.183*	0.109*	0.106*	0.093*	0.021*	0.073*	0.085*	0.028*	0.052	0.066	−0.002	0.169*	−0.036	0.283	0.251	0.304	0.348	0.296
B84						1.000	0.478	0.187	0.063	0.051	0.111	0.191	0.061	0.048	0.004	0.059	−0.061	0.203	0.036	−0.012	0.047	0.035	−0.088	0.391	0.326	0.391	0.348	0.364
FR1128							1.000	0.253	0.041*	0.152	0.148	0.164	0.070	0.057	0.046	0.038	−0.047	0.237	0.077	0.061	−0.041	0.134	−0.103*	0.304	0.326	0.326	0.261	0.304
FR618								1.000	−0.015*	0.125	0.121	0.108	0.042	−0.030	−0.042	0.041	−0.011	0.118	0.019	0.064	−0.134	0.019	−0.160*	0.239	0.348	0.261	0.348	0.299
FR986									1.000	0.157	0.475*	0.283*	0.465*	0.416*	0.262*	0.278*	0.392*	0.033	−0.033	0.011	0.035	0.341*	0.030	0.326	0.251	0.283	0.118	0.244
FR460										1.000	0.203*	0.246	0.248	0.203	0.164	0.124	0.047	0.200	−0.015	−0.059	−0.008	0.217*	0.048	0.239	0.261	0.283	0.261	0.261
FR2200											1.000	0.505*	0.487*	0.407*	0.311*	0.414	0.105*	0.167*	0.071	0.114	0.083	0.535*	0.075	0.304	0.217	0.261	0.283	0.266
FR632												1.000	0.556*	0.564*	0.471*	0.424	0.182*	0.270	−0.088	−0.043	0.008	0.571*	0.062	0.304	0.196	0.239	0.217	0.239
A635													1.000	0.909	0.603	0.582	0.371	0.121*	−0.099	−0.113	0.065	0.704*	0.057	0.326	0.283	0.261	0.304	0.293
A665														1.000	0.581	0.619	0.352	0.108*	−0.139*	−0.122*	0.083	0.622*	0.104	0.304	0.239	0.261	0.261	0.266
CM105															1.000	0.348	0.157	0.278*	−0.061	−0.045	0.040	0.559*	0.125	0.348	0.196	0.261	0.326	0.283
A641																1.000	0.115	0.029	−0.037	0.036	0.031	0.482	0.026	0.304	0.217	0.217	0.261	0.250
FR31																	1.000	0.008	−0.062	−0.168*	−0.024	0.213*	−0.027	0.370	0.235	0.348	0.348	0.325
FR088																		1.000	0.098*	0.111*	−0.016	0.212*	0.011	0.261	0.348	0.239	0.239	0.272
FR15A																			1.000	0.612	−0.024	0.000	0.125	0.261	0.304	0.283	0.283	0.283
N28																				1000	−0.197*	−0.014	0.019	0.239	0.283	0.239	0.283	0.261
FR4A																					1.000	−0.054	0.331	0.348	0.348	0.370	0.370	0.359
B14																						1.000	0.033	0.304	0.217	0.196	0.239	0.239
B37																							1.000	0.326	0.239	0.261	0.283	0.277
C103																								1.000	0.391	0.304	0.413	0.283[b]
OH43																									1.000	0.370	0.304	0.276[b]
T8																										1.000	0.391	0.280[b]
WF9																											1.000	0.275[b]

* f^{M}_{AB} estimate significantly different ($\alpha = 0.05$) from f_{AB}. f^{M}_{AB} was determined using 46 probe-enzyme combinations. Average $\sigma_{error} = 0.095$

[a] AIS, Average proportion of marker variants shared with C103, OH43, T8, and WF9. $\sigma_{error} = 0.028$

[b] Average proportion of marker variants shared with BSSS lines

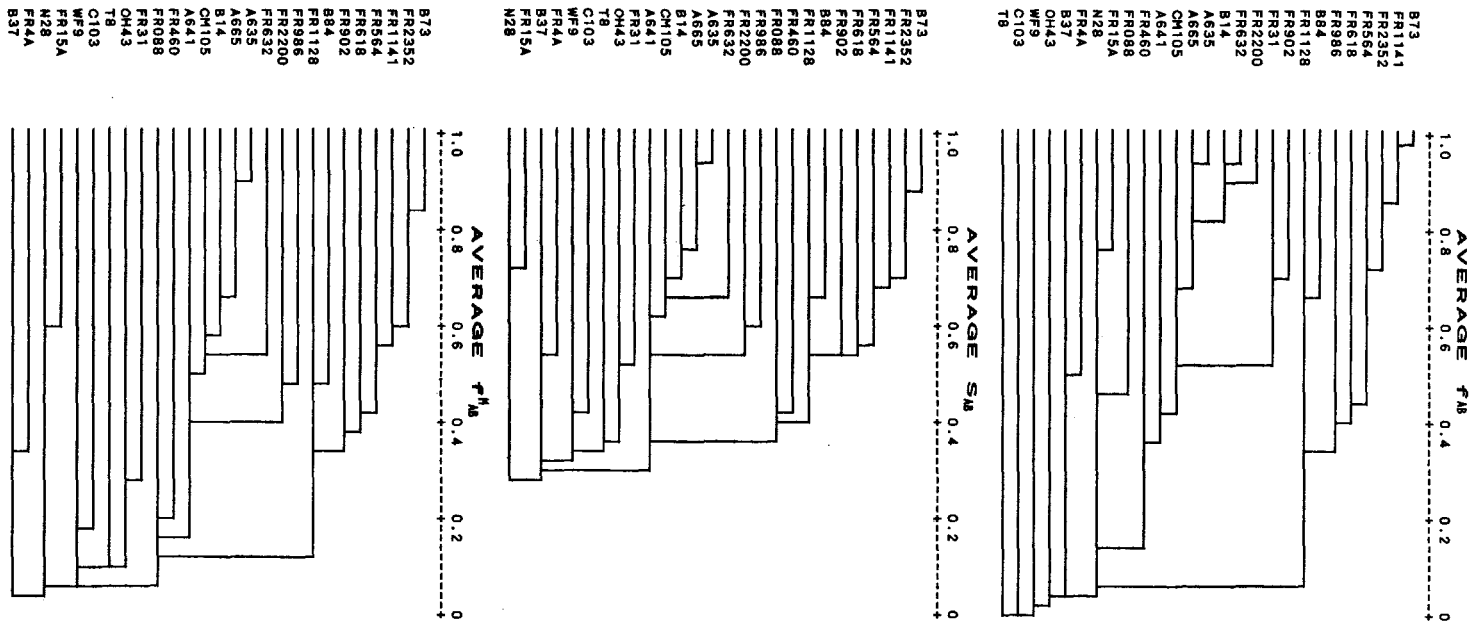

Fig. 1. Average f_{AB}, S_{AB}, and f^{M}_{AB} between clusters of inbred lines

of Lancaster Sure Crop and two pairs of BSSS lines, the average f_{AB} was 0.437, whereas the average S_{AB} was 0.665. The proportion of MM variants shared between Lancaster Sure Corp and nine BSSS lines and between

BSSS and eight Lancaster Sure Crop lines was used to remove the bias in S_{AB}. The resulting average f^M_{AB} was 0.437. In soybean (Cox et al. 1985), the average f_{AB} and S_{AB} (determined using 20 isozyme and morphological trait loci; values in parentheses) was 0.13 (0.64) for the 1950's cultivars, 0.10 (0.62) for the 1960's cultivars, and 0.19 (0.63) for the 1970's cultivars. Average f^M_{AB}, calculated using average S_{AB} among unrelated ($f_{AB} = 0$) cultivars to remove bias, was 0.14 for the 1950's cultivars, 0.10 for the 1960's cultivars, and 0.12 for the 1970's cultivars.

f^M_{AB} may also be used to obtain RFLP- and isozyme-based estimates of relatedness that are more consistent with each other. Among five pairs of BSSS lines studied by Messmer et al. (1991), average S_{AB} was 0.80 based on isozyme data compared with 0.60 based on RFLP data. Using the proportion of isozyme and RFLP variants shared between the BSSS lines and the unrelated line Mo17 to correct for bias, the resulting f^M_{AB} was 0.20 using isozymes and 0.29 using RFLPs.

Independent estimates of S_{AB} and f^M_{AB} among BSSS lines varied substantially (Table 6). For example, the combination (B73, B84) had $S_{AB} = 0.522$ and $f^M_{AB} = 0.307$ in the present study. In contrast, $S_{AB} = 0.720$ and $f^M_{AB} = 0.533$ on the basis of the data of Melchinger et al. (1991). Varying numbers of MM loci have been used by different researchers. For the estimation of f^M_{AB}, perhaps not enough MM loci have been used in these studies. The minimum number of MM loci necessary for obtaining a specified precision of f^M_{AB} values can be determined under simplified conditions. Assume that the total number of shared MM variants follows a binomial distribution, and that no variants are shared between unrelated lines. To obtain f^M_{AB} values with confidence intervals of $f^M_{AB} \pm 0.05$ at $\alpha = 0.05$, the following numbers of MM loci (n) are necessary: (1) $n = 138$ if $f'_{AB} = 0.1$ or 0.9; (2) $n = 246$ if $f'_{AB} = 0.2$ or 0.8; (3) $n = 323$ if $f'_{AB} = 0.3$ or 0.7; (4) $n = 369$ if $f'_{AB} = 0.4$ or 0.6; and (5) $n = 384$ if $f'_{AB} = 0.5$. These calculations ignore the added variation in f^M_{AB} due to the proportion of MM variants shared between unrelated individuals, and n will be larger when unrelated individuals share MM variants. Using MM loci with large numbers of variants at each locus may reduce the likelihood that unrelated individuals share MM variants. More than 250 MM loci seem necessary when f'_{AB} ranges from 0.2 to 0.8, indicating that large numbers of loci are needed to obtain sufficiently precise estimates of genetic relatedness using MM.

Table 5. Measures of relatedness for two sets of published data

Pair of inbreds	f_{AB}	S_{AB}	f^M_{AB}
1. From Godshalk et al. (1990), 47 molecular marker loci:			
B73, VA95	0.76	0.79	0.66
B73, VA96	0.76	0.74	0.58
B73, VA97	0.52	0.66	0.46
B73, VA98	0.52	0.64	0.41
VA95, VA96	0.61	0.83	0.73
VA95, VA97	0.45	0.77	0.63
VA95, VA98	0.45	0.69	0.48
VA96, VA97	0.45	0.64	0.42
VA96, VA98	0.45	0.64	0.41
VA97, VA98	0.39	0.62	0.36
Average	0.536	0.702	0.513
2. From Melchinger et al. (1991), 83 molecular marker loci:			
OH43, A619	0.75	0.78	0.64
OH43, B55	0.25	0.59	0.30
OH43, B86	0.50	0.76	0.58
A619, B55	0.38	0.59	0.32
A619, B86	0.38	0.63	0.38
B55, B86	0.13	0.55	0.21
B38, B46	0.38	0.56	0.28
B37, B76	0.75	0.86	0.78
Average	0.437	0.665	0.437

Table 6. Independent estimates of relatedness between Iowa Stiff Stalk Synthetic lines[a]

Pair of lines	S_{AB}	f^M_{AB}	Reference
B73, B84	0.561	0.344	Dudley et al. (1991)
($f_{AB} = 0.265$)	0.720	0.533	Melchinger et al. (1991)
	0.700	0.478	Messmer et al. (1991)
	0.698	0.517	Godshalk et al. (1990)
	0.522	0.307	Present study
Average	0.640	0.436	
B14, B37	0.530	0.183	Messmer et al. (1991)
($f_{AB} = 0.0625$)	0.467	0.152	Godshalk et al. (1990)
	0.283	0.033	Present study
Average	0.427	0.123	
B14, B73	0.600	0.208	Messmer et al. (1991)
($f_{AB} = 0.0625$)	0.551	0.267	Godshalk et al. (1990)
	0.348	0.134	Present study
Average	0.500	0.203	
B14, B37	0.600	0.273	Messmer et al. (1991)
($f_{AB} = 0.0625$)	0.407	0.026	Godshalk et al. (1990)
	0.326	0.035	Present study
Average	0.444	0.111	
B73, B37	0.630	0.383	Messmer et al. (1991)
($f_{AB} = 0.0625$)	0.510	0.241	Godshalk et al. (1990)
	0.196	−0.096	Present study
Average	0.445	0.176	
B84, B37	0.610	0.395	Messmer et al. (1991)
($f_{AB} = 0.0625$)	0.467	0.169	Godshalk et al. (1990)
	0.261	−0.088	Present study
Average	0.446	0.159	

[a] The following numbers of molecular marker loci were used by different researchers: Dudley et al. (1991), 66; Melchinger et al. (1991), 83; Messmer et al. (1991), 144; Godshalk et al. (1990), 47; and present study, 46

Implications for specifying minimum distance between inbred lines

The use of MM for varietal protection has been proposed (Hunter 1989). A minimum distance may be specified, below which two inbred lines are declared essentially similar. The presence of shared MM variants between unrelated individuals (δ_{AB}) must be considered when using S_{AB} to estimate genetic relatedness. For example, suppose a minimum distance of 0.125 is specified, i.e., maximum $f_{AB} = 0.875$. If S_{AB} is used to measure genetic distance and $\delta_{AB} = 0.30$, then the minimum distance actually corresponds to maximum $S_{AB} = [0.875 + (1 - 0.875)0.30] = 0.9125$. Hence, if δ_{AB} is not considered when using S_{AB} to measure relatedness, two inbred lines that meet the minimum distance requirement could be erroneously declared as essentially similar.

Acknowledgements. The author would like to express his grateful appreciation to R. M. Hogan and J. W. Dudley for access to the RFLP data set used in this study.

References

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

Cox TS, Kiang YT, Gorman MB, Rodgers DM (1985) Relationship between coefficient of parentage and genetic similarity indices in the soybean. Crop Sci 25:529–532

Dudley JW, Saghai-Maroof MA, Rufener GK (1991) Molecular markers and grouping of parents in maize breeding programs. Crop Sci 31:718–723

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

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

Godshalk EB, Lee M, Lamkey KR (1990) Relationship between restriction fragment length polymorphisms to single-cross hybrid performance of maize. Theor Appl Genet 80:273–280

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

Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor Appl Genet 72:761–769

Henderson CB (1984) Maize research and breeders manual. No. 10. Illinois Foundation Seeds, Champaign, Ill

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

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

Messmer MM, Melchinger AE, Lee M, Woodman WL, Lee EA, Lamkey KR (1991) Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data. Theor Appl Genet 83:97–107

Smith OS, Smith JSC, Bowen SL, Tenborg RA, Wall SJ (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

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

Sprague GF (1946) Early testing of inbred lines of corn. J Am Soc Agron 38:108–117

Note added in proof

The discussion on minimum distances does not in any way preclude the use of S_{AB} to determine whether two varieties are essentially similar. Rather, if merely indicates that $(1 - S_{AB})$ and pedigree-based minimum distances may not be equivalent if unrelated lines have MM variants in common.