Theor Appl Genet (1991) 83:97–107

THEORETICAL AND APPLIED GENETICS

Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize population: comparison of allozyme and RFLP data *

M.M. Messmer [1], A.E. Melchinger [1], M. Lee [2,], W.L. Woodman [2], E.A. Lee [2] and K.R. Lamkey [3]**

[1] University of Hohenheim, W-7000 Stuttgart 70, FRG
[2] Department of Agronomy, Iowa State University, Ames, IA 50011, USA
[3] Cereal and Soybean Research Unit, USDA, Agricultural Research Services and Department of Agronomy, Iowa State University, Ames, IA 50011, USA

Received October 25, 1990; Accepted March 7, 1991
Communicated by A.R. Hallauer

Summary. Data for restriction fragment length polymorphisms (RFLPs) of 144 clone-enzyme combinations and for 22 allozyme loci from 21 U.S. Corn Belt maize (*Zea mays* L.) inbreds were analyzed. The genetic materials included 14 progenitors of the Iowa Stiff Stalk Synthetic (BSSS) maize population, both parents of one missing BSSS progenitor, four elite inbreds derived from BSSS, and inbred Mo17. Objectives were to characterize the genetic variation among these 21 inbreds for both allozymes and RFLPs, to compare the results from both types of molecular markers, and to estimate the proportion of unique alleles in the BSSS progenitors. Genetic diversity among the 21 inbreds was substantially greater for RFLPs than for allozymes, but the percentages of unique RFLP variants (27%) and unique allozyme alleles (25%) in the BSSS progenitors were similar. Genetic distances between inbreds, estimated as Rogers' distance (RD), were, on average, twice as large for RFLP (0.51) as for allozyme data (0.24). RDs obtained from allozyme and RFLP data for individual line combinations were only poorly correlated ($r=0.23$); possible reasons for discrepancies are discussed. Principal component analysis of RFLP data, in contrast to allozyme data, resulted in separate groupings of the ten BSSS progenitors derived from the 'Reid Yellow Dent' population, the four BSSS elite lines, and Mo17. The remaining six BSSS progenitors were genetically rather diverse and contributed a large number of rare alleles to BSSS. The results of this study corroborate the fact that RFLPs are superior to allozymes for characterizing the genetic diversity of maize breeding materials, because of (1) the almost unlimited number of markers available and (2) the greater amount of polymorphisms found. In particular, RFLPs allow related lines and inbreds with common genetic background to be identified, but a large number of probe-enzyme combinations is needed to estimate genetic distances with the precision required.

Key words: Allozymes – RFLPs – Genetic diversity – Rare alleles – Genetic distances – Principal component analysis – *Zea mays* L.

Introduction

Assessment of the genetic diversity that exists in the available germ plasm is fundamental in the improvement of agricultural plants. During the past two decades, allozymes have been extensively used to examine the genetic variability in breeding materials and natural populations of more than 30 plant species (for review, see Tanksley and Orton 1983). In maize (*Zea mays* L.), allozymes have been employed to characterize the genetic diversity among elite inbreds (Stuber and Goodman 1983; Smith et al. 1985a, b), commercial hybrids (Smith 1984, 1986), open-pollinated and exotic populations (Kahler et al. 1986; Smith 1986), and germ plasm collections (Goodman and Stuber 1983; Doebley et al. 1983, 1988). Allozyme assays are relatively simple, rapid, and inexpensive. However, their usefulness for certain breeding problems is limited by the insufficient sampling of the genome, owing to (a) the small number of enzyme loci available in most species and (b) the reduced number of polymorphic loci generally found in breeding materials with a narrow genetic base (Melchinger et al. 1990).

* Joint contribution from Cereal and Soybean Research Unit, USDA, Agricultural Research Service, and Journal Paper No. J-14236 of the Iowa Agricultural and Home Economics Experiment Station, Projects 2818 and 2778

** To whom correspondence should be addressed

Restriction fragment length polymorphisms (RFLPs) have been proposed as an alternative approach to characterize the genetic diversity in plant species (Burr et al. 1983; Beckmann and Soller 1983). A major advantage of RFLPs in maize, as well as in other crop species, is the abundance of polymorphic loci found within breeding materials (Clegg 1989). This enabled maize geneticists to develop a well-populated RFLP linkage map (Coe et al. 1988) and allows genetic distances among inbred lines and populations to be estimated with high precision. Results from recent studies (Lee et al. 1989; Melchinger et al. 1990) suggested that RFLP data can be used for assigning maize inbreds to heterotic groups and for quantifying the genetic similarity between related lines, but a large number of probe-enzyme combinations is required to obtain reliable estimates of genetic distances.

A fundamental difference between allozymes and RFLPs concerns the level at which genetic variation is detected. Allozymes are proteins that represent different alleles of an enzyme locus affecting the electrophoretic mobility in the gel. In contrast, RFLPs reveal differences directly at the level of DNA; in principle, all types of mutational events (point mutations, insertions, deletions, etc.), which alter the DNA sequence of the restriction enzyme recognition site, can be detected, irrespective of whether they occurred in coding or non-coding regions of the genome. This raises the question of whether or not the genetic diversity for RFLPs is a good predictor of genetic diversity for allozymes and, more generally, for expressed genes. To our knowledge, only a single study (Landry et al. 1987) with lettuce (*Lactuca sativa* L.) has been published to date, in which genetic distances determined from both types of molecular markers have been compared for the same materials.

Objectives of our study were to (1) characterize the genetic variation among progenitors of and elite lines derived from the Iowa Stiff Stalk Synthetic (BSSS) maize population for both allozymes and RFLPs, (2) compare the results of both types of markers, and (3) estimate the proportion of unique alleles present in the BSSS progenitor lines. The materials investigated have been included in previous allozyme studies (Smith et al. 1985a, b); they are of special interest to maize breeders and geneticists because elite inbred lines developed from BSSS (B14, B37, B73, B84) have been extensively used as parents in commercial hybrid seed production (Darrah and Zuber 1986). In addition, the BSSS maize population has been used for numerous basic genetic studies concerning selection methods, gene effects, and inbreeding depression (Hallauer et al. 1983).

Materials and methods

The 'Iowa Stiff Stalk Synthetic' (BSSS) maize population was developed in the early 1930s by intermating 16 inbreds selected by corn breeders for superior stalk quality (Sprague and Jenkins 1943; Hallauer et al. 1983). Of these progenitors, 10 were derived from various strains of the 'Reid Yellow Dent' open-pollinated population, 4 had miscellaneous origins, and the genetic background of 2 is unknown; the former 10 and latter 6 progenitors will hereinafter be designated as RYD and non-RYD lines, respectively.

In 1939, half-sib recurrent selection (HT) for increased grain yield was initiated in BSSS with the double-cross tester 'Iowa 13' (Sprague 1946) and continued over seven cycles. In each cycle, approximately 10 of 100 S_1 lines were selected primarily on the basis of their testcross performance in replicated field experiments across several locations. The 10 selected S_1 lines were recombined by making all 45 possible crosses among them to produce the source material of the next cycle. Details of the experimental procedures have been published (Eberhart et al. 1973).

Maize inbred lines examined

The materials analyzed in this study comprised (Table 1): (a) 14 of the 16 original progenitors of BSSS (F1B1-7-1 and CI.617 had been lost) and both parental inbreds of one missing progenitor, F1B1-7-1, (these 16 inbreds will hereinafter be designated as BSSS progenitors); (b) four elite lines derived from different cycles of half-sib recurrent selection of BSSS(HT), hereinafter designated as BSSS elite lines; and (c) Mo17, an inbred widely used in commercial hybrids, particularly in combination with lines from BSSS, derived from the cross between CI.187-2 (a BSSS progenitor from RYD) and C103 (a line developed from

Table 1. Inbreds used in the analysis of allozyme and RFLP data

Lines	Background[a]
Progenitors of BSSS	
	(a) Reid Yellow Dent
CI.187-2	Krug, Nebraska Reid × Iowa Gold Mine
Fe[b]	Reid Early
I159	Iodent, Reid Yellow Dent
I224A2	Iodent, Reid Yellow Dent
Ind.461-3	Reid Medium (Duddleston No. 461)
Ind.AH83	Funks 176A, Reid Yellow Dent
Ind.B2[b]	Reid Yellow Dent
Ind.TR9-1-1-6	Reid Early Dent (Troyer strain)
Os420	Osterland, Reid Yellow Dent
WD456	Walden Dent, Reid Yellow Dent
	(b) Non-Reid Yellow Dent
A3G-3-1-3	unknown
CI.540	Illinois Two Ear
Ill.12E	unknown
Ill.Hy	Illinois High Yield
LE23	Illinois Low Ear
Oh3167B	Echelberger Clarage
Elite lines	
B14A	Cuzco × B14[8] rust res. sel.[c]
B37	BSSS(HT)C0
B73	BSSS(HT)C5
B84	BSSS(HT)C7
Mo17	CI.187-2 × C103

[a] Hallauer et al. (1983) and Henderson (1984)
[b] Parent of progenitor F1B1-7-1
[c] B14 was developed from BSSS(HT)C0

Lancaster Sure Crop). The lines were highly inbred and been maintained in cold storage at the Iowa Agricultural Experiment Station in Ames/IA, from the original stocks by carrying out one generation of selfing and rogueing for off-type plants about every 10 years. Seeds used for RFLP assays originated from a generation of line maintenance later than those used for allozyme assays. On the basis of phenotypic appearance and results from the molecular marker analyses, there was no evidence of seed-stock contamination or remnant heterozygosity in any of the lines.

Allozyme analyses

Data of allozyme genotypes for 22 enzyme loci (Table 2) from the 21 inbreds were taken from Stuber and Goodman (1983). Details of laboratory techniques and allele designations were described by these authors. For the statistical analyses, allele designations were converted into binary data, i.e., presence or absence of an allele in a line was coded by 1 and 0, respectively.

RFLP analyses

The 21 inbreds were assayed for their respective RFLP patterns. From each inbred, equal quantities of leaf tissue harvested from five seedlings were lyophilized and bulked. The techniques for plant genomic DNA isolation, separate digests with restriction enzymes *Eco* RI, *Hin*dIII, and *Eco*RV, gel electrophoresis, Southern blotting, filter hybridization with ^{32}P-labelled clones, and autoradiography were as described by Lee et al. (1989). At every eighth lane in the gel, a set of molecular markers was loaded, which was composed of lambda fragments of 2.3, 4.3, 6.7, 9.4, 13.3, and 21.3 kb. DNA probes used for hybridization were selected from collections of mapped maize clones provided by B. Burr (Brookhaven National Laboratory, Upton/NY) and D. Hoisington (University of Missouri, Columbia/MO).

Data were recorded only for those clone-enzyme combinations that yielded a clear hybridization signal and a single-copy hybridization pattern (detecting differences for one band per line). Altogether, the RFLP data set included 79 clones (77 random genomic clones and 2 cloned genes) and 144 clone-enzyme combinations (Table 3), because not all clones were hybridized against all restriction digests. A total of 11 clones was assayed with the three restriction enzymes (*Eco*RI, *Hin*dIII, *Eco*RV); 43 clones were assayed with both *Eco*RI and *Hin*dIII; 6, 11, and 8 clones were assayed only with *Eco*RI, *Hin*dIII, and *Eco*RV, respectively. Each chromosome was covered by at least five clones, and the average map distance between adjacent markers was approximately 28 c*M*.

RFLP profiles for inbreds in autoradiographs were scored visually. Each unique RFLP band that could be clearly distinguished from other bands due to difference in migration distances was considered as a distinct variant. Data were binary coded, i.e., presence or absence of a band in a line was coded by 1 or 0, respectively.

Statistical analyses

Multilocus Rogers' distances (RD) were calculated for all 210 possible pairs of inbreds for both allozyme and RFLP data, according to the formula given by Rogers (1972). In calculating RD of RFLP data, it was assumed that, for a given clone-enzyme combination, different variants correspond to different alleles of the same locus. However, the allelic nature of variants has not been proven explicitly in most cases. Because pure-breeding lines were used in this study, the RD is equal to the number of marker loci (allozymes or RFLPs) for which two lines differed, divided by the total number of marker loci considered.

Allozyme and RFLP data were analyzed separately by principal component analysis (PCA) to arrive at a condensed graphical representation of the association among the 21 inbreds. Calculations of principal components were based on the respective covariance matrix of binary allozyme allele and RFLP variant frequencies, by using the PROC PRINCOMP program of SAS (SAS Institute 1988).

Simple correlations were calculated between RDs determined from allozyme data and from the complete set of RFLP data (144 clone-enzyme combinations), hereinafter designated as set 0. To evaluate the effects of the number of clones and restriction enzymes, as well as the distribution of clones over the genome on the precision of RD estimates, correlations were calculated between RDs determined from various subsets of set 0: sets 1 and 2 each consist of 27 different clones (approximately three clones per chromosome) with both restriction enzymes *Eco*RI and *Hin*dIII; sets 3 and 4 refer to the data obtained from 54 clones used in sets 1 and 2 in combination with *Eco*RI and *Hin*dIII, respectively; set 5 comprises data of 85 clone-enzyme combinations with clones that map to chromosome arms (1L, 3S, 3L, 4S, 5S, 5L, 6L, 9, 10L) with at least one polymorphic allozyme locus.

Uniqueness measures

To estimate the unique gene contribution of each progenitor line to the BSSS population, two different measures of 'uniqueness' were applied. An unweighted uniqueness measure, U_i, for each progenitor line i was obtained by counting the number of allozyme alleles or RFLP variants, which occurred exclusively in this progenitor, but not in the other progenitors. A weighted uniqueness measure, W_i, for line i was calculated as:

$$W_i = \frac{1}{N} \sum_{l=1}^{N} 1/P_{li} ,$$

where P_{li} is the frequency of the allozyme allele or RFLP variant, present in line i at the l^{th} allozyme locus or clone-enzyme combination, respectively (with regard to the entire set of the 16 BSSS progenitors), and N is the total number of allozyme loci ($N=22$) or clone-enzyme combinations ($N=144$) assayed.

U_i can range between 0 and N. W_i can range between 1 and 16, the total number of BSSS progenitors assayed. A W_i value of 1 indicates that all loci were monomorphic over all 16 progenitors, whereas a W_i value of 16 indicates that the respective line has unique allozyme alleles or RFLP variants at all loci or clone-enzyme combinations considered.

Results

Genetic variation for allozymes and RFLPs

Of the 22 allozyme loci, 15 (68%) showed polymorphisms among the 21 inbreds. Altogether, 45 alleles were found across all loci and inbreds (Table 2). The maximum number of alleles at a given locus was three.

Of the 144 clone-enzyme combinations used in this study, 136 (94%) revealed polymorphisms among the 21 inbreds. A total of 472 RFLP variants was detected, corresponding with an average of 3.3 variants per clone-enzyme combination (Table 3). The maximum number of RFLP variants per clone-enzyme combination was eight. Chromosomes 1, 4, and 6 were more polymorphic than chromosomes 7, 8, and 9. *Eco*RI and *Hin*dIII each yield-

ed an average of 3.1 RFLP variants per clone for the 54 clones assayed with both restriction enzymes.

Figure 1 shows histograms for the number of BSSS progenitors that carried the same allozyme allele or RFLP variant, considering the 45 allozyme alleles and the 464 RFLP variants found in this subset of 16 inbreds. Allozyme and RFLP data showed a high percentage of unique allozyme alleles (27%) and RFLP variants (25%), present in only one progenitor. Although 49% of the RFLP variants, occurred in three or fewer progenitors only 35% of the allozyme alleles did. A larger fraction (29%) of the allozyme alleles, compared with 4% of the RFLP variants, occurred in at least 15 progenitors.

Figure 2 shows the unique contribution of each progenitor to the BSSS population, based on the weighted (W_i) and unweighted (U_i) uniqueness measure for allozyme and RFLP data. A3G-3-1-3 had the greatest W_i value for RFLP data, followed by Os420, CI.540, and LE23, whereas Ind.B2, I224A2, and CI.187-2 had the smallest W_i values. Progenitors from RYD averaged smaller W_i values (3.0 and 1.7) than non-RYD lines (3.6 and 2.6) determined from RFLP and allozyme data, respectively. For RFLP data, W_i and U_i statistics yielded a similar ranking ($r_s = 0.90$) for most progenitors. However, there were considerable discrepancies in W_i values between allozyme and RFLP data, except for the extremes A3G-3-1-3 and Ind.B2.

Two (17%) of the 12 unique allozyme alleles and 20 (17%) of the 117 unique RFLP variants found in the BSSS progenitors were also present in at least one of the four BSSS elite lines. Both unique allozyme alleles of CI.540 were found in B14A. Four unique RFLP variants present in the BSSS elite lines came from Ind.TR9-1-1-6 and three were from Ind.461-3. B14A showed three new RFLP variants not present in any of the progenitors, all detected by one clone (UMC84) in combination with the three restriction enzymes. Mo17 had five new RFLP variants (detected by clones UMC60, UMC136, BNL9.44), absent in the other 20 inbreds.

Table 2. Chromosomal location of allozyme loci assayed and number of alleles found in the 21 inbreds

Chromosomal location[a]	Locus[a]	No. of alleles	Chromosomal location	Locus	No. of alleles
1L	*Adh1*	3	5S	*Mdh5*	2
1L	*Mdh4*	1	5S	*Pgm2*	3
1L	*Mmm*	1	5L	*Got2*	2
1L	*Pgm1*	2	6L	*Enp1*	1
1L	*Phi1*	3	6L	*Idh2*	3
3S	*E8*	3	6L	*Mdh2*	3
3L	*Got1*	1	6L	*Pgd1*	2
3L	*Mdh3*	2	8L	*Idh1*	1
3L	*Pgd2*	2	8	*Mdh1*	1
4S	*Cat3*	2	9	*Acp1*	3
5S	*Got3*	1	10L	*Glu1*	3
Total				22	2.1

[a] Chromosomal location and nomenclature according to Stuber and Goodman (1983) and Smith (1989)

Table 3. Chromosomal location of DNA clones assayed and average number of RFLP variants found per clone-enzyme combination in the 21 inbreds

Chromosome	Clone designation[a]	Clone-enzyme combination	Avg. no. of variants per clone-enzyme combination
1	BNL5.62, UMC76, UMC11, BNL5.59, UMC58, UMC119, UMC33, UMC83, UMC107, UMC106, UMC84, BNL6.32	24	4.2
2	UMC5, UMC34, UMC131, UMC4, UMC88, BNL6.20	13	3.1
3	UMC92, UMC10, UMC50, UMC102, UMC60, UMC3, UMC17	13	3.5
4	UMC31, UMC66, UMC19, BNL15.45, UMC15	10	3.7
5	UMC90, UMC27, UMC43, BNL4.36, UMC126, UMC104, UMC35	11	3.3
6	UMC85, BNL6.29, UMC59, UMC65, UMC46, UMC132, UMC134	15	3.7
7	BNL15.40, UMC136, UMC116, BNL15.21, UMC110, UMC56, BNL14.07, BNL16.06, UMC80	14	2.9
8	UMC103, UMC124, BNL9.44, UMC120, UMC12, UMC30, UMC93, UMC7	14	2.3
9	UMC109, BNL-Sh1, UMC113, BNL3.06, UMC81, BNL5.10, UMC20, BNL5.04, UMC114, UMC95, BNL-Css1, BNL5.09	20	2.8
10	BNL3.04, UMC130, UMC64, BNL10.13, UMC57, BNL7.49	10	3.0
Total	79	144	3.3

[a] Clone designations according to maize RFLP linkage maps of Coe et al. (1988)

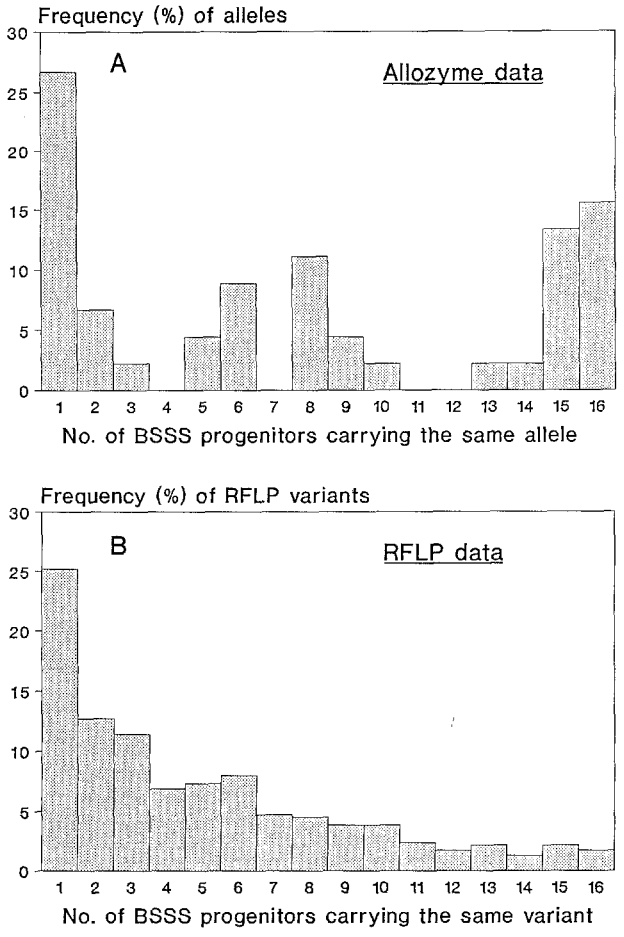

Fig. 1 A, B. Histogram of the number of BSSS progenitors with the same allozyme allele or RFLP variant, detected by analysis of 22 allozyme loci and 144 clone-enzyme combinations, respectively

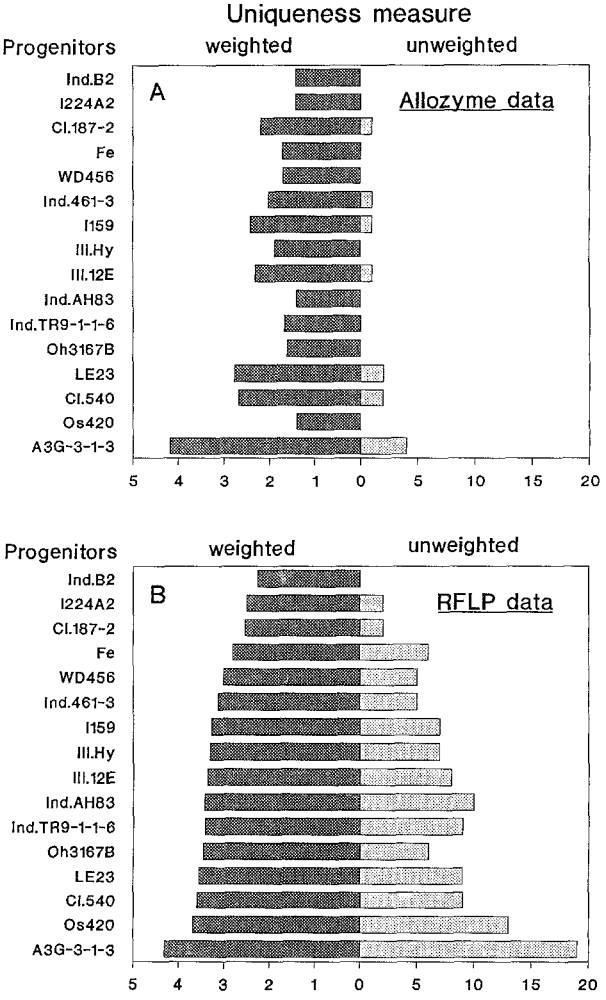

Fig. 2 A, B. Histogram of weighted (W_i) and unweighted (U_i) uniqueness measure for each progenitor of BSSS detected by analysis of 22 allozyme loci and 144 clone-enzyme combinations, respectively

Genetic distances among inbreds

RDs calculated from allozyme data across the 210 possible combinations between the 21 lines ranged from 0.05 (I224A2 × Os420, Ind.B2 × B37) to 0.46 (A3G-3-1-3 × I159, A3G-3-1-3 × LE23, and A3G-3-1-3 × B14A) and had a mean of 0.24 (Table 4). In comparison, RDs calculated from RFLP data had a smaller range (0.30 to 0.66) and a greater mean (0.51).

For RDs based on RFLP data, line combinations of type BSSS progenitors × BSSS progenitors, BSSS elite lines × BSSS progenitors, and Mo17 × BSSS progenitors had similar means and ranges (Table 4). The mean RD among the BSSS elite lines (0.40) was considerably smaller than among the BSSS progenitors (0.51); within the latter group, the ten RYD lines on average had a smaller mean RD (0.48) than the other six (non-RYD) lines (0.56). Mo17 had a slightly smaller mean RD with the RYD lines (0.49) than with the six non-RYD lines (0.55) and with the BSSS elite lines (0.57). The allozyme data corroborated these results, except that Mo17 was equidistant to the BSSS progenitors and BSSS elite lines.

RDs from RFLP data for individual line combinations are presented in Table 5. The greatest RD (0.66) among the 16 BSSS progenitors was found between lines CI.540 and Ill.Hy, and the smallest RD (0.37) was between RYD lines CI.187-2, I.224A2, and Ind.B2. The two pairs of related lines, B73 × B84 and Mo17 × CI.187-2, had the smallest RD (0.30 and 0.31, respectively) among all 21 inbreds.

The mean RD of each BSSS progenitor line in combination with all other BSSS progenitors varied from 0.46 (I224A2, Ind.B2) to 0.55 (A3G-3-1-3, Ill.Hy, Oh3167B). Non-RYD lines had a higher mean RD than RYD lines, except for Ill.12E. The mean RD of each BSSS progenitor in combination with the four BSSS elite lines varied

Table 4. Mean, minimum, and maximum of Rogers' distance (RD × 100) of different groups of inbreds based on allozyme and RFLP data

Group	N^a	RD × 100					
		Allozyme data			RFLP data		
		Mean	Min.	Max.	Mean	Min.	Max.
Among all inbreds	210	24.2	5	46	50.5	30	66
Among BSSS progenitors	120	24.9	5	46	51.3	37	66
Within RYD lines	45	21.5	5	41	47.6	37	57
Within non-RYD lines	15	30.9	14	46	55.8	41	66
RYD lines × non-RYD lines	60	26.0	14	46	52.8	43	63
BSSS elite lines × BSSS progenitors	64	23.2	5	46	49.9	41	63
BSSS elite lines × RYD lines	40	21.0	5	36	49.1	41	60
BSSS elite lines × non-RYD lines	24	26.7	9	46	51.1	42	63
Mo17 × BSSS progenitors[b]	15	25.8	18	41	51.3	38	60
Mo17 × RYD lines[b]	9	23.7	18	32	48.7	38	60
Mo17 × non-RYD lines	6	28.8	18	41	55.1	51	59
Among BSSS elite lines[c]	5	20.0	14	27	40.4	37	47
Mo17 × BSSS elite lines	4	25.0	18	32	57.0	48	64

[a] N = number of line combinations
[b] Mean and minimum without RD between related lines Mo17 × CI.187-2
[c] Mean and minimum without RD between related lines B73 × B84

Table 5. Rogers' distance (RD × 100) calculated from RFLP data of 144 clone-enzyme combinations for the 21 inbreds

Line	CI.187-2	Fe	I159	I224A2	Ind.461-3	Ind.AH83	Ind.B2	Ind.TR9-1-1-6	Os420	WD456	A3G-3-1-3	CI.540	Ill.12E	Ill.Hy	Le23	Oh3167B
	Progenitors × progenitors										*Elite lines × elite lines*					
CI.187-2											Line		Line			
Fe	46												B14A	B37	B73	B84
I159	46	49														
I224A2	37	42	46													
Ind.461-3	42	46	50	44												
Ind.AH83	42	50	52	47	49											
Ind.B2	37	45	43	37	49	47					B37		47			
Ind.TR9-1-1-6	52	42	54	49	47	54	57				B73		40	37		
Os420	50	51	55	45	52	49	50	45			B84		40	39	30	
WD456	45	51	51	46	57	52	40	51	49		Mo17		48	67	53	62
A3G-3-1-3	51	52	53	52	54	57	52	53	54	57						
CI.540	54	51	63	49	57	58	47	57	51	48	60					
Ill.12E	50	45	54	46	50	51	45	57	55	60	58	41				
Ill.Hy	56	47	49	51	54	58	48	55	56	53	60	66	56			
Le23	50	53	56	43	46	48	48	59	57	58	52	53	50	61		
Oh3167B	53	53	59	60	59	56	47	55	56	46	59	53	54	48	65	
Mean[a]	48	48	52	46	50	51	46	52	52	51	55	54	52	55	53	55
	Elite lines × progenitors															
B14A	47	44	54	40	48	57	44	52	44	44	52	42	47	52	53	47
B37	56	53	51	49	56	60	55	47	50	43	58	51	63	49	56	56
B73	48	43	47	42	41	54	50	51	46	52	51	53	47	50	47	57
B84	51	46	47	49	48	60	48	50	46	51	53	49	53	44	47	51
Mean[b]	51	47	50	45	48	58	49	50	46	47	54	49	52	49	51	53
Mo17	31	48	51	43	38	50	41	60	54	54	59	51	51	57	59	54

[a] Mean of RDs for each progenitor with the other 15 progenitors
[b] Mean of RDs for each progenitor with 4 BSSS-derived elite lines

Table 6. Correlations (r_{xy}) between Rogers' distances calculated from various sets[a] of allozyme and RFLP data for the 21 inbreds

Variable		r_{xy}
X	Y	
Rogers' distance based on		
Allozyme data	RFLP data (Set 0)	0.23**
Allozyme data	RFLP data (Set 5)	0.17*
RFLP data (Set 1)	RFLP data (Set 2)	0.17*
RFLP data (Set 3)	RFLP data (Set 4)	0.61**

*,** Significant at the 0.05 and 0.01 levels of probability, respectively

[a] For definition of sets, see "Materials and methods"

from 0.45 (I224A2) to 0.58 (Ind.AH83), and differed considerably for some lines (Ind.AH83, WD456, Ill.Hy) from the corresponding mean RD to the other BSSS progenitors.

The correlation across all 210 line combinations between RDs calculated from allozyme data and the complete set of RFLP data (set 0) was 0.23 (Table 6). The correlation of the RDs calculated from two subsets (sets 3 and 4) of RFLP data with the same 54 clones but with two different restriction enzymes (*Eco*RI and *Hin*dIII, respectively) was 0.61.

Principal component analyses of allozyme and RFLP data

Associations among the 21 inbreds revealed by principal component analyses (PCA) of allozyme and RFLP data are presented in Fig. 3. The first, second, and third principal component explained 17.4, 15.4, and 13.4%, respectively, of the total variation for allozyme data, and 10.6, 8.5, and 8.0%, respectively, of the total variation for RFLP data. In the PCA of RFLP data (Fig. 3B), the six BSSS progenitors with non-RYD origin were mostly located outside the spread of the ten BSSS progenitors from RYD and widely spaced from each other. Within the ten BSSS progenitors from RYD, lines CI.187-2, Ind.AH83, Ind.461-3, I224A2, and Ind.B2 were grouped closely together. The BSSS elite lines formed a loose group that was separated from the BSSS progenitors. Mo17 occupied a position distant from the BSSS elite lines but close to the cluster of RYD lines, especially its parent CI.187-2.

PCA of allozyme data (Fig. 3a) also separated RYD from non-RYD lines. The RYD lines were more widely spread, however the position of progenitors I224A2, Ind.AH83, LE23, and WD456 relative to the other lines showed considerable discrepancies in the allozyme PCA as compared with RFLP PCA. Also, the BSSS elite lines were not grouped together and not as clearly separated from Mo17 as in Fig. 3B.

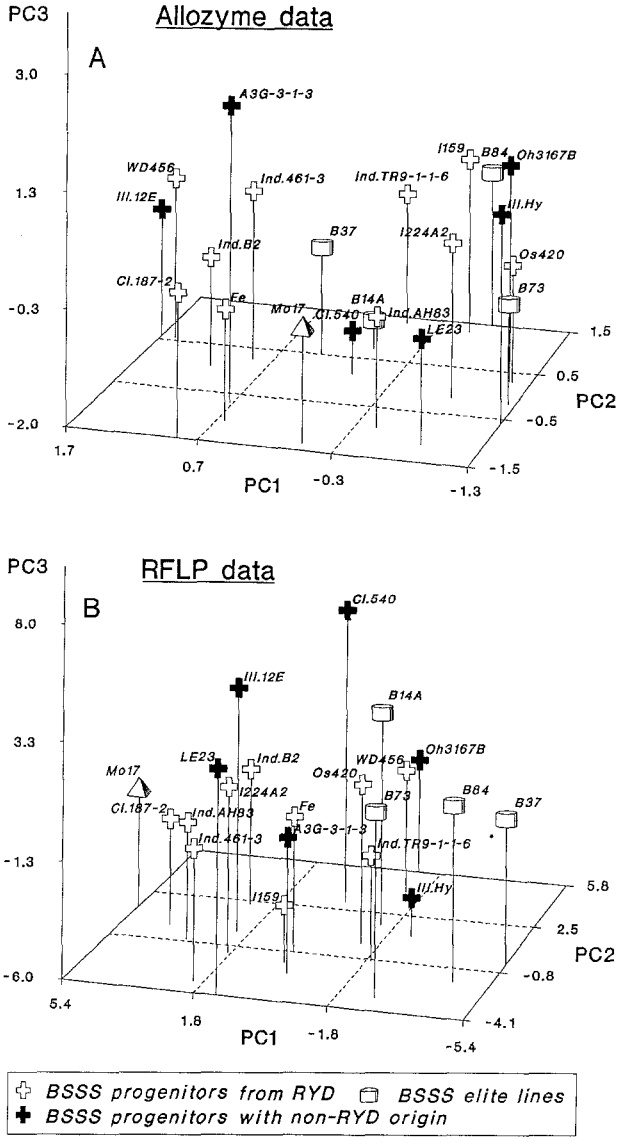

Fig. 3A, B. Associations between the 21 inbreds revealed by principal component analysis based on covariance matrix of allozyme data (22 loci) and RFLP data (144 clone-enzyme combinations), respectively

Discussion and conclusions

Comparison of allozyme and RFLP data

Two major advantages of RFLPs over allozymes with respect to most applications of molecular markers in plant breeding are: (1) the greater amount of polymorphisms found in breeding materials, and (2) the better coverage of the genome (Tanksley 1983). The present study of 21 maize inbreds corroborated the fact that the genetic variation was substantially greater for RFLPs than for allozymes, both in terms of the percentage of polymorphic loci (94% versus 68%) and the average number of variants or alleles found per polymorphic lo-

cus (3.4 versus 2.5). The 76 clones that yielded polymorphisms with at least one restriction enzyme provided a significantly better coverage of the maize genome (approx. 71% within a 20 c*M* distance of a marker) than the 15 polymorphic allozyme loci (approx. 23% of the genome within a 20 c*M* distance). Other RFLP studies with U.S. Corn Belt inbreds (Lee et al. 1989; Melchinger et al. 1990) also found ca. twice the amount of genetic variation per locus as compared to allozyme studies (Smith et al. 1985a, b).

Although each of the 21 inbreds had a distinct allozyme and RFLP genotype, the two types of markers differed considerably in their discriminatory power. Three pairs of unrelated lines differed for only one allozyme locus, and the most distant lines differed for 10 of the 22 allozyme loci. By comparison, the number of different RFLP variants found in two lines for the 144 clone-enzyme combiantions varied form 43 to 95. Thus, RFLPs allowed a more refined measurement of genetic distances than allozymes.

Allozyme and RFLP data agreed well in the rankings of the mean RD for the various sets of line combinations shown in Table 4. RDs for individual line combinations however, were poorly correlated between the two types of molecular markers ($r=0.23$), and this lack of agreement is reflected in the results from PCA (Fig. 3). In inbred lines of lettuce (*Lactuca sativa* L.), Landry et al. (1987) also compared genetic distances for allozymes with those determined for RFLPs and found discrepancies. Similar findings were reported for different *Drosophila* species, for which the level of allozyme diversity did not predict the level of diversity for RFLPs (for review, see Clegg 1989). In the present study, several causes may have contributed to the observed discrepancies.

First, the two types of molecular markers differ in the nature of genetic polymorphisms detected. With allozymes, variation can be detected only for protein-coding genes, and only those mutational events changing the protein mobility in a gel can be resolved. In contrast, any variation in DNA sequence can, in principle, be detected with RFLPs, including non-coding regions and those types of mutational events that are silent in their effect on gene expression because of the redundancy in the genetic code. Comparisons of complete DNA sequences of genes from population samples suggest that the coding region of genes is generally strongly conserved, whereas the DNA in non-coding regions is subject to more complex mutational changes and displays a higher level of diversity (for review, see Clegg 1989). This hypothesis could explain the discrepancies between genetic distances estimated from RFLP and allozyme data, although it raises the question of whether or not RFLP data provide an accurate portrayal of the genetic diversity within breeding materials for expressed genes. Possibly, cDNA clones are more suitable in this respect than random genomic clones. However, the genomic clones used in the present study originated from libraries developed with *Pst*I, which selects for undermethylated (likely transcribed) regions (Bird 1987).

Second, the two types of molecular markers differed in their coverage of the maize genome (Tables 2 and 3). When RDs were calculated from the 85 clone-enzyme combinations (RFLP data set 5) that mapped to the same ten chromosome arms as the 15 polymorphic allozyme loci, they were nonetheless not better correlated ($r=0.17$) with allozyme-based RDs than RDs calculated from the complete set (set 0) of 144 clone-enzyme combinations (Table 6). This suggests that differences in the distribution of the two sets of markers over the genome contributed little to the poor association between allozyme- and RFLP-based genetic distances.

Third, both sets of markers sample only a small proportion of the entire maize genome and, consequently, the RD estimates calculated from them are subject to sampling errors. In order to assess the importance of sampling effects on the reliability of RD estimates, (1) correlations were calculated between RDs based on different subsets of markers within each type, and (2) standard errors of RDs were estimated by the jackknife method (Miller 1974). RDs based on two sets of 11 allozyme loci were not associated ($r=-0.05$), and RDs based on RFLP data sets 1 and 2 (each based on a set of 27 different clones with both enzymes *Eco*RI and *Hin*dIII) were only poorly correlated ($r=0.17$) (Table 6). Standard errors of RDs obtained by the jackknife method (not shown) were approximately two to three times larger for the allozyme data than for the complete set of RFLP data (set 0; $N=144$). This leads to the conclusion that more than 100 markers are required for estimating genetic distances with sufficient precision, and the number of allozyme loci used in this study ($N=22$) was too small for obtaining reliable estimates. Additional support for this conclusion comes from the fact that allozyme data, in contrast to RFLP data, failed to identify the two pairs of related inbreds in our materials on the basis of their RD estimate.

An important question in this context is, to what extent are the results from two different restriction enzymes obtained with a particular set of clones correlated. For practical reasons (especially the number of hybridizations with radioactive DNA probes), it would be more efficient to use two or more enzymes with a reduced number of clones rather than a larger number of clones with a single enzyme. The highly significant ($P<0.01$) correlation ($r=0.61$) between RDs calculated from set 3 (54 clones with *Eco*RI) and RDs calculated from set 4 (same 54 clones with *Hin*dIII) suggests that using two instead of one enzyme per clone is less efficient for increasing the precison of RD estimates than doubling the

number of clones, which would also provide a better coverage of the genome.

In the PCAs of the 21 inbreds, the percentage of the total variation explained by the first three principal components was almost twice as large for allozyme data (46%) as for RFLP data (27%). This is consistent with other allozyme (Smith and Smith 1987 1988) and RFLP studies (Godshalk et al. 1990, Melchinger et al. 1990) of maize inbreds from the U.S. Corn Belt. The total variation among the 21 inbreds (calculated as the trace of the covariance matrix of binary data) was 13 times greater for RFLP than for allozyme data as a result of the larger number of RFLP variants. Therefore, the result mentioned does not necessarily imply that PCA of RFLP data yields less reliable results than PCA of allozyme data; rather, only a smaller proportion of the more extensive information on genetic distances among lines can be represented by the first three principal components. Actually, grouping of the inbreds revealed by PCAs (Fig. 3) was in better agreement with information on genetic background and pedigree of the lines for RFLP data than for allozyme data.

Our results support the proposal of Burr et al. (1983) that RFLPs may be superior to allozymes for applications in plant variety protection (e.g., fingerprinting of inbreds), because of the greater amount of polymorphisms present in breeding materials. In addition, the abundance of RFLP markers available in maize allows the genetic diversity of the maize germ plasm to be measured and characterized with excellent precision. In view of the considerable costs and expenditures for RFLP assays, however, further research is needed to determine the minimum number of markers required under a given setting.

Comparison of different groups of inbreds based on RFLP data

Although one original progenitor (CI.617) of BSSS was missing and one (F1B1-7-1) was only represented by its two parents, it seems safe to assume that the results obtained for the 16 inbreds designated as BSSS progenitors in this study provide a fairly accurate portrayal of the genetic composition of the 16 original progenitors of BSSS. The relatively large values for the RFLP-based RDs (Table 4) among the 16 BSSS progenitors indicate that the inbreds used for the synthesis of the BSSS population were genetically diverse and not closely related to each other. In accordance with their origin from the same source population, the ten BSSS progenitors from RYD were less diverse than the remaining six BSSS progenitors designated as non-RYD lines. Two of the latter six lines, A3G-3-1-3 and Ill-12E, were of unknown origin; however, on the basis of their RDs with the other 14 BSSS progenitors (Table 5) and the results from PCA (Fig. 3B), it seems unlikely that they originated from RYD.

The BSSS elite lines had a substantially lower mean RD (0.40) compared with the BSSS progenitors (0.53) (Table 4), suggesting that genetic variation had been lost. This may have occurred during the synthesis of BSSS but, more likely, was the result of selection and drift during recurrent selection and line development.

The increase in the mean RD of Mo17 × BSSS elite lines (0.57) compared with Mo17 × BSSS progenitors (0.51) suggests that gene frequencies in the BSSS elite lines had been shifted in the opposite direction of the genes present in Mo17. This trend is also illustrated by the results from PCA (Fig. 3B). Mo17 had only a moderate RD (0.53) to B73 (Table 5), although their cross is known for superior hybrid performance and heterosis (Lee et al. 1989). The small RD (0.30) found between B73 and B84 is consistent with the expected coancestry coefficient (f) (Malecot 1948) of 0.265 calculated for these two lines (Melchinger et al. 1990). Mo17 had a relatively large RD (0.31) to its parent CI.187-2, considering their coancestry coefficient ($f=0.5$); in accordance with the breeding behavior of Mo17, this suggests that Mo17 inherited a larger proportion of its genome from its other parent C103.

PCA of RFLP data (Fig. 3B) resulted in separate groupings of the ten BSSS progenitors from RYD, the four BSSS elite lines, and Mo17 in agreement with the genetic background and known breeding behavior of these lines. Furthermore, the graph illustrates that the six BSSS progenitors with non-RYD origin represent a genetically diverse set of lines that contributed a large number of rare variants to the BSSS population (Fig. 2B). This result questions the widely accepted notion of considering the BSSS population merely as a subgroup of RYD because it presumably contains a substantial proportion of genes not present in RYD germ plasm.

Our findings corroborate the conclusions of a recent study by Melchinger et al. (1990) that RFLP data are useful for (a) revealing pedigree relationships between lines and (b) grouping inbreds according to their genetic background and origin. In practical breeding programs for hybrid maize, this information could be used to: (1) assign inbreds with unknown genetic background to well-established heterotic groups, (2) identify new heterotic groups, and (3) choose parents for making crosses in recycling breeding programs (Lee et al. 1989; Melchinger et al. 1990).

Relevance of uniqueness measures and rare alleles

The weighted (W_i) and unweighted (U_i) uniqueness measures provide additional information about the genetic diversity potentially contributed by an inbred with respect to a given set of lines. In accordance with genetic expecta-

tions, lines originating from diverse (non-RYD) genetic backgrounds generally had greater values for W_i and U_i than lines derived from the same (RYD) population (Fig. 2B). Information concerning the uniqueness of inbreds, evaluated on the basis of their RFLP genotype, could be used as a criterion for: (1) choice of genetically diverse parents in creating source populations for recurrent selection programs, (2) selection of diverse components for establishing broad-based synthetic varieties, and (3) identification of unique inbred lines for conservation of maize germ plasm in gene banks.

Although RFLPs and allozymes differed in the amount of polymorphisms detected, there was striking agreement in the high percentage of unique RFLP variants (27%) and unique allozyme alleles (25%) found among the 16 BSSS progenitors (Fig. 2). Neuhausen (1989) reported a similar percentage (23%) of unique RFLP variants among the BSSS progenitors with a different set of clones. Provided these findings and especially those for allozymes are representative for expressed genes, they suggest that approximately one-fourth of the alleles present in the original BSSS population traced back to single progenitors.

An important conclusion from this result concerns the choice of adequate population size in recurrent selection programs. According to population genetic theory (Crow and Kimura 1970, Chaps. 7.5 and 8.6), the probability of loss of a favorable allele during recurrent selection due to genetic drift depends on its selective advantage (s), its initial gene frequency (q_0), and the effective population size (N_e). For $N_e = 10.5$ (corresponding to recombination of ten S_1 lines in the BSSS(HT) recurrent selection program), the probability of losing a rare ($q_0 = 1/16$), selectively neutral ($s = 0$) allele during the first five selection cycles is 0.57 (determined from Eq. 7.4.41, Crow and Kimura 1970); this probability is only slightly reduced for favorable alleles ($s > 0$), unless they have a large gene effect. Consequently, the chances that many favorable, rare alleles have been lost during the initial cycles of recurrent selection programs with the BSSS population are very high, thereby limiting the medium- and long-term selection progress. This is confirmed by empirical results of Smith (1983) and Helms et al. (1989), who found significant effects of genetic drift when selection progress for yield improvement in the BSSS(HT) and BSSS(R) populations was analyzed according to the model of Smith (1983). In agreement with Helms et al. (1989), we therefore conclude that the population size ($N_e = 10.5$) in the original recurrent selection programs with BSSS was too small and should be reconsidered in long-term selection programs.

Acknowledgements. We thank Ursula Klug for typing the manuscript and Iris E. Melchinger for preparing the figures and tables. This work was supported by grants from the Iowa Biotechnology Council and Pioneer Hi-Bred International.

References

Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping, and costs. Theor Appl Genet 67:35–43

Bird A (1987) CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342–347

Burr B, Evola SV, Burr FA, Beckmann JS (1983) The application of restriction fragment length polymorphism to plant breeding. In: Setlow JK, Hollaender A (eds) Genetic engineering principles and methods, vol 5. Plenum, New York, pp 45–59

Clegg MT (1989) Molecular diversity in plant populations. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and, genetic resources. Sinauer, Sunderland/MA, pp 98–115

Coe EH, Neuffer MG, Hoisington DA (1988) The genetics of corn. In: Sprague GF, Dudley JW (eds) Corn and corn improvement, 3rd edn. Agron Monogr 18. American Society of Agronomy, Madison/WI, pp 81–258

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper & Row, New York

Darrah LL, Zuber MS (1986) 1985 United States farm maize germplasm base and commerical breeding strategies. Crop Sci 26:1109–1113

Doebley JF, Goodman MM, Stuber CW (1983) Isozyme variation in maize from the southwestern United States: taxonomic and anthropological implications. Maydica 28:97–120

Doebley JF, Wendel JD, Smith JSC, Stuber CW, Goodman MM (1988) The origin of cornbelt maize: the isozyme evidence. Econ Bot 42:120–131

Eberhart SA, Debela S, Hallauer AR (1973) Reciprocal recurrent selection in the BSSS and BSCB1 maize populations and half-sib selection in BSSS. Crop Sci 13:451–456

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

Goodman MM, Stuber CW (1983) Races of maize. VI. Isozyme variation among races of maize in Bolivia. Maydica 28:169–187

Hallauer AR, Russell WA, Smith OS (1983) Quantitative analysis of Iowa Stiff Stalk Synthetic. Stadler Symp 15:83–104

Helms TC, Hallauer AR, Smith OS (1989) Genetic drift and selection evaluated from recurrent selection programs in maize. Crop Sci 29:602–607

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

Kahler AL, Hallauer AR, Gardner CO (1986) Allozyme polymorphisms within and among open-pollinated and adapted exotic populations of maize. Theor Appl Genet 72:592–601

Landry BS, Kesseli R, Leung H, Michelmore RW (1987) Comparison of restriction endonucleases and sources of probes for their efficiency in detecting restriction fragment length polymorphisms in lettuce (*Lactuca sativa* L.) Theor Appl Genet 74:646–653

Lee M, Godshalk EB, Lamkey KR, Woodman WW (1989) Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. Crop Sci 29:1067–1071

Malecot G (1948) Les mathématiques de l'hérédite. 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

Miller RG (1974) The jackknife: a review. Biometrika 61:1–15

Neuhausen SL (1989) A survey of Iowa Stiff Stalk Synthetic parents, derived inbreds and BSSS (HT)C5 using RFLP analysis. Maize Genet Coop Newslett 63:110–111

Rogers JS (1972) IV. Measures of genetic similarity and genetic distance. Studies in genetics VII. University of Texas Publishers Austin/TX 7213:145–153

SAS Institute (1988) SAS/STAT user's guide for personal computers. Release 6.03 edn. SAS Institute Inc. Cary/NC

Smith JSC (1984) Genetic variability within U.S. hybrid maize: multivariate analysis of isozyme data. Crop Sci 24:1041–1046

Smith JSC (1986) Genetic diversity within the corn belt dent racial complex of maize (*Zea Mays* L.): Maydica 31:349–367

Smith JSC (1989) The characterization and assessment of genetic diversity among maize (*Zea Mays* L.) hybrids that are widely grown in France: chromatographic data and isozymic data. Euphytica 43:73–85

Smith JSC, Smith OS (1987) Associations maong inbred lines of maize using electrophoretic, chromatographic, and pedigree data. 1. Multivariate and cluster analysis of data from 'Lancaster Sure Crop' derived lines. Theor Appl Genet 73:654–664

Smith JSC, Smith OS (1988) Associations among inbred lines of maize using electrophoretic, chromatographic, and pedigree data. 2. Multivariate and cluster analysis of data from Iowa Stiff Stalk Synthetic drived lines. Theor Appl Genet 76:39–44

Smith JSC, Goodman MM, Stuber CW (1985a) Genetic variability within U.S. maize germplasm. I. Historically important lines. Crop Sci 25:550–555

Smith JSC, Goodman MM, Stuber CW (1985b) Genetic variability within U.S. maize germplasm. II. Widely used inbred lines 1970 to 1979. Crop Sci 25:681–685

Smith OS (1983) Evaluation of recurrent selection in BSSS, BSCB1, and BS13 maize populations. Crop Sci 23:35–40

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

Sprague GF, Jenkins MT (1943) A comparison of synthetic varieties, multiple crosses, and double crosses in corn. J Am Soc Agron 35:137–147

Stuber CW, Goodman MM (1983) Allozyme genotypes for popular and historically important inbred lines of corn, *Zea Mays* L. US Dept Agric, Agric Res Serv, ARR-S-16,29 pp

Tanksley SD (1983) Molecular markers in plant breeding. Plant Mol Biol Rep 1:3–8

Tanksley SD, Orton TJ (1983) Isozymes in plant genetics and breeding. Parts 1A and 1B. Elsevier, Amsterdam