Genetic Diversity of Public Inbreds of Sorghum Determined by Mapped AFLP and SSR Markers

Monica A. Menz, Robert R. Klein,* Natalie C. Unruh, William L. Rooney, Patricia E. Klein, and John E. Mullet

ABSTRACT

The objectives of this study were (i) to assess the level of genetic diversity in elite sterility-maintaining (B) and fertility-restoring (R) sorghum [*Sorghum bicolor* (L.) Moench] lines as compared with a group of exotic and converted germplasm (IS) from the World Collection, (ii) to compare the classification of germplasm on the basis of estimates of genetic similarities obtained by means of AFLP and microsatellite (SSR) markers, and (iii) to compare the classification of germplasm obtained by different classes of molecular markers. A set of 100 SSRs, 1318 *Eco*RI/*Mse*I AFLP, and 496 *Pst*I/*Mse*I AFLP markers with known map positions were utilized to determine the genetic similarity in a group of B, R, and IS public inbreds. Cluster analysis of genetic similarity estimates (GS_{ij}) revealed that the classification of sorghum inbreds is based on the sorghum working groups, Zera-zera, Kafir, Kafir-Milo, Durra, and Feterita. Cluster analyses failed to give a clear differentiation between B- and R-lines, suggesting that R- and B-lines do not represent well-defined heterotic groups in this set of public lines. By comparing the different classes of molecular markers (SSRs, AFLPs, combinations of SSRs and AFLPs), we determined that the distribution of the markers and the coverage of the genome by the markers did affect the classification of genotypes. Dendrograms of genetic similarity (GS) based on *Pst*I/*Mse*I AFLP markers, or a set of markers spaced at 1- to 2-cM intervals across the genome, produced clusters that were in better agreement with pedigree information than the analysis based solely on the *Eco*RI/*Mse*I AFLP or SSR markers used in this study.

Sorghum is native to Sub-Saharan Africa and has been cultivated for centuries as a staple cereal grain in much of Asia and Africa. Sorghum became economically important in the semiarid plains of the central USA in the 1800s (Duncan et al., 1991). Before 1950, sorghum production was based on self-pollinated cultivars, and germplasm enhancement activities were restricted to a limited number of photoperiod insensitive accessions, mainly from Africa and India. New varieties were eventually selected from segregating progenies resulting from the hybridization of two or more cultivars. Commercial exploitation of F_1 hybrids in grain sorghum started in the USA in the mid-1950s, with the development of the cytoplasmic male sterility system (CMS) A1 (Quinby, 1971, 1974; Kramer, 1987). While additional CMS systems have been discovered, almost all sorghum hybrid seed production relies on the A1 CMS system originally described by Stephens and Holland (1954). Because of the uniform use of a single CMS system, the relatively recent introduction of the crop, and the photoperiod sensitive nature of most exotic sorghum germplasm, there was concern that the genetic base of sorghum in the USA was restricting genetic gains. To increase the amount of germplasm available to sorghum improvement programs, the Texas Agricultural Experiment Station (TAES)–USDA Sorghum Conversion Program was initiated in 1960. The goal of this program is to convert tall, photoperiod sensitive sorghum genotypes to short, photoperiod insensitive genotypes that can be used in temperate breeding programs (Stephens et al., 1967). This program has been successful and many of the improvements made in sorghum hybrids in the past 30 yr are due to the availability of germplasm from this program (Rooney and Smith, 2001).

The genetic diversity in the germplasm of a breeding program affects the potential genetic gain through selection. Information about genetic diversity also permits the classification of germplasm into heterotic groups, which is particularly important to hybrid breeding. Even though the genetic mechanisms that explain heterosis are not fully understood, it is well documented that crosses between unrelated, and consequently genetically distant parents, show greater hybrid vigor than crosses between closely related parents (Stuber, 1994; Hallauer, 1999). Estimates of molecular-marker based genetic distance have proven to be a useful way to describe existing heterotic groups, to identify new heterotic groups, and to assign inbreds of unknown genetic origin to established heterotic groups (Dubreuil et al., 1996; Saghai-Maroof et al., 1997; Hongtrakul et al., 1997; Pejic et al., 1998; Casa et al., 2002).

Genetic diversity in sorghum has been estimated utilizing several types of molecular markers (Tao et al., 1993; Vierling et al., 1994; Brown et al., 1996; Taramino et al., 1997; Uptmoor et al., 2003). Unfortunately, these analyses did not thoroughly sample the germplasm pool of many breeding programs nor did they provide extensive marker coverage of the sorghum genome. In contrast, Ahnert et al. (1996) used a set of 104 RFLP probes to evaluate the genetic diversity among a larger set of elite proprietary sorghum inbred lines. Estimates of genetic diversity among male parental restorer (R) and maintainer female (B) lines were in agreement with pedigree and breeding information. However, this last study did not include germplasm from the World Collection; therefore, the elite lines surveyed in this study ignored a significant portion of the germplasm pool utilized by both public and private sorghum breeders.

M.A. Menz, N.C. Unruh, P.E. Klein, and J.E. Mullet, Institute for Plant Genomics and Biotechnology, and W.L. Rooney, Dep. of Soil and Crop Science, Texas A&M Univ., College Station, TX 77843; R.R. Klein, USDA-ARS, Southern Plains Agricultural Research Center, 2765 F&B Road, College Station, TX 77845. Received 6 Aug. 2003. *Corresponding author (bob@sorggenome.tamu.edu).

Published in Crop Sci. 44:1236–1244 (2004).

677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: GS, genetic similarity; SSR, simple sequence repeat; TAES, Texas Agricultural Experiment Station.

With the advent of high-density genetic maps and high-throughput marker systems for sorghum, it is now feasible to estimate genetic diversity with a large number of markers that are well distributed across the sorghum genome. The advantage of using markers with known map positions instead of a random sample is that there is control over the coverage of the genome. It is thus possible to avoid overrepresentation of certain regions of the genetic map that can produce inaccurate estimates of genetic similarities among individuals.

Because the AFLP assay generates multiple polymorphic bands in a single gel, it has been used increasingly for genetic diversity studies (Hongtrakul et al., 1997; Zhu et al., 1998; Fregene et al., 2000; Lima et al., 2002; Ubi et al., 2003). AFLP markers can be generated by CNG methylation-sensitive (*Pst*I/*Mse*I) or by CNG methylation-insensitive (*Eco*RI/*M*seI) enzyme combinations (ECs). The *Eco*RI/*Mse*I EC will produce AFLP markers uniformly distributed across the physical map of sorghum, but in a genetic map they will look clustered along the hypermethylated regions of the genome that show low recombination rates, such as around the centromeres (Menz et al., 2002). Conversely, the *Pst*I/*Mse*I EC will produce markers that will be uniformly distributed along the genetic map except for the methylated regions of the chromosome. Comparisons of the efficiency of different types of markers to discriminate among genotypes have been performed in the past (Djé et al., 1999; Uptmoor et al., 2003; Pejic et al., 1998). However, with the exception of one study in corn (*Zea mays* L., Vuylsteke et al., 2000), previous comparisons utilized unmapped molecular markers and, therefore, regions of the genome may have been over- or underrepresented in each of these diversity studies.

The objectives of this study were (i) to estimate the genetic diversity present among elite public sorghum inbreds and compare this with the diversity present among a set of converted lines from the World Collection, (ii) to classify elite sorghum inbred lines based on genetic similarity estimated using different sets of molecular markers, and (iii) to compare the classification of germplasm obtained by means of different classes of molecular markers.

MATERIALS AND METHODS

Plant Material

The 50 sorghum inbreds included in this study were selected on the basis of their contribution to the sorghum hybrid industry during the past 50 yr (Table 1). All the lines included in this study have been publicly released and used by breeders in both the public and private sectors to create commercial hybrids and/or proprietary inbred lines. This collection of germplasm includes older lines developed directly from original pure-line cultivars in the 1950s to more recent public germplasm releases that can be considered second generation lines (e.g., derived from combinations of old lines, Table 1). The 50 lines include 31 elite temperate adapted inbreds (13 B-lines and 18 R-lines) and 19 exotic inbreds converted to photoperiod insensitivity (16 IS lines and 3 pure-line cultivars) that have been widely used as progenitors of elite inbreds. Two different seed sources of BTx623 and IS3620C were included as controls, and one version of each of these two lines was loaded on each gel in triplicate for use as reference standards.

AFLP and SSR Analysis

For each of the sorghum lines analyzed, DNA was extracted from a bulk of leaf tips from 20 different seedlings harvested 1 wk after germination with the Fast-Prep DNA kit (QBioGene Inc., La Jolla, CA).

For AFLP analysis, DNA samples were digested with *Eco*RI plus *Mse*I, or *Pst*I plus *Mse*I restriction endonucleases. AFLP template preparation and PCR reaction conditions were as described by Klein et al. (2000) and Menz et al. (2002). One hundred ninety-two (+3/+3) AFLP primer combinations were examined in all lines with a dual-dye LI-COR 4200 IR^2 gel detection system (LI-COR Inc., Lincoln, NE). For the 100 SSR markers analyzed, forward primers were labeled either with one of the IR fluorescent dyes (LI-COR Inc.) or with one of the three phosphoramidite fluorescent dyes (6-FAM, NED, or HEX) for use with the ABI Prism 3700 DNA sequencing system (Applied Biosystems, Foster City, CA). Primer information, PCR conditions, and map positions of all the markers included in this study have been published elsewhere (Menz et al., 2002) and are available from the Sorghum Genome website at http://SorghumGenome.tamu.edu; verified 18 March 2004. AFLP data were analyzed by Bionumerics software version 2.5 (Applied Maths, Belgium), while SSR data were analyzed either by the software package Genotyper version 3.6 (Applied Biosystems) (SSRs run on the ABI 3700), or scored visually (SSRs run on the LI-COR system).

Data Analysis

To estimate the genetic similarity between each pair of lines, 1814 AFLP and 100 SSR marker loci were scored. All 1914 markers used in this study were previously mapped in a population produced by crossing two inbred lines included here, BTx623 and IS3620C (Menz et al., 2002). Therefore, for all AFLP bands identified in each gel, only those bands corresponding to markers previously mapped by Menz et al. (2002) were included in the analyses (see supplementary documentation at http://SorghumGenome.tamu.edu). Dominant AFLP markers were scored as present (1) or absent (0). For SSRs, differences in band size were scored as different allelic forms of that locus. With the exception of linkage group D (LG-D), LG-F, and LG-G, SSRs were well distributed along the sorghum genetic map (for mapping position and information about these markers see http://SorghumGenome.tamu.edu). In the case of LG-D, LG-F, and LG-G, each was analyzed with 7, 5, and 6 SSRs, respectively, but the markers were clustered into a small region of the map.

Samples of BTx623 and IS3620C were run in triplicate to provide an estimate of the reproducibility of band size determination. For each SSR examined, this variability was always less than one base pair. All data were transformed to binary code producing a matrix of presence (1) versus absence (0) of each allele. The resulting matrix was used to estimate genetic similarity among all pairs of lines by Dice coefficient of similarity (Nei and Li, 1979) as follows:

$$GS_{ij} = 2\,N_{ij}/(N_i + N_j),$$

where N_{ij} is the number of alleles (scored bands) shared by lines i and j, and N_i and N_j are the total number of scored bands in lines i and j, respectively. Negative matches (0–0) were not included. Values of GS may range from 1 (identical profiles for all markers) to 0 (no bands in common).

Five similarity matrices were generated for all pair-wise

Table 1. Pedigree of the 50 sorghum inbred lines used for SSR and AFLP data collection.

Line	Pedigree†, variety name, race–working group	Fertility reaction‡
1. BTx3197	SA5765, Combine Kafir-60, Kafir	B
2. BTx398	SA5330, Martin, Wheatland derivative	B
3. BTx399	SA6697, Wheatland-288, Kafir-Milo	B
4. BTx378	Redlan, Kafir	B
5. BTx3042	Redbine Selection, Kafir-Milo	B
6. BTx406	S808, 4-dwarf Martin-Backcross derived	B
7. BOK11	Dwarf Hydro × Rice, Kafir	B
8. BTx623	BTx3197 × SC170-6-4, Kafir × Zera-zera	B
9. BTx626	BTx378 × SC110-6, Kafir × Zera-zera	B
10. BTx631	BTx615 × (BTx378 × SC110-9), Kafir × Zera-zera	B
11. BTx635	RS/R $(C_2)S_1$–102-1, Zera-zera-derivative	B
12. BTxArg-1	MR807 × (BTx3197 × SC170-6-4)	B
13. BTx642	B35, BC1 of IS12555, Durra	B
14. Tx7000	SA7000, Caprock, Kafir-Milo	R
15. Tx7078	SA7078, Combine7078, Kafir-Milo	R
16. RTx414	(SA396 × RTx7078) × RTx7078, Kafir-Milo and Feterita	R
17. Tx2536	SA7529, Short Kaura × Feterita, Caudatum Kaura and Feterita	R
18. TAM2566	BC_2 of IS12666 (SC175-9), Zera-zera	R
19. TAM428	BC_2 of IS12610 (SC110-9), Zera-zera	R
20. Tx2737	TAM2554 × [(SA7536-1 × Tx7000) × Tx2536]	R
21. RTx430	Tx2536 × SC170-6SC110-14E, Feterita × Zera-zera	R
22. Tx2783	IS12610C × [{[(ROKY8 × Tx2536) × SC110-9] × SC599-6} × SC110-14E],Zera-zera	R
23. RTx432	SC599-6 × SC110-9, Caudatum-Nigricans	R
24. RTx433	Tx414 × SC108-6Kafir-Milo × Zera-zera	R
25. Tx2817	BC_1 of IS12661 (SC170-6), Zera-zera	R
26. RTx2858	[(Tx412 × SC599-6) × SC326-6] × QL-1	R
27. Tx2862	Tx2783 × {(Tx2746 × RTx430) × RTx430},Zera-zera derivative	R
28. Tx2880	[RTx430 × {[(RY8 × Tx2536) × (SC110-9 × SC599-6)] × SC175-9}] × Tx2766	R
29. RTx436	(SC120-6 × Tx7000) × Tx7000, Zera-zera and Kafir-milo	R
30. Tx2903	[(SC120-6 × Tx7000) × Tx7000] × Tx433, Zera-zera and Kafir-milo	R
31. Tx2908	SC599-6 × Tx430, Caudatum/Feterita × Zera-zera	R
32. IS12608C	SC108, Caudatum/Zera-zera	R
33. IS12610C	SC110, Caudatum/Zera-zera	R
34. IS12666C	SC175, Caudatum/Zera-zera	R
35. IS2816C	SC120, Caudatum/Zera-zera	R
36. IS2508C	SC414, Caudatum/Caudatum-Kafir	R
37. IS3620C	SC303, Guinea/Margaritiferum	B
38. IS12555C	SC35-14E, Durra/Durra	R
39. IS12661C	SC170, Caudatum/Zera-zera	R
40. IS17459C	SC599, Rio, Caudatum/Caudatum-Nigricans	R
41. SC326-6	BC derivative of IS3758C, Caudatum/Nigricans	R
42. IS6705C	SC265, Guinea/Conspicuum	B
43. IS7452C	SC372, Caudatum/Caudatum-Kaura	PR
44. IS12646C	SC155, Durra-Bicolor/Durra-Dochna	R
45. IS3552C	SC748, Caudatum/Caudatum-Guinea	R
46. IS2856C	SC650, Kafir/Caffrorum-Birdproof	
47. IS9290C	SC1079, Caudatum/Caudatum	
48. ICSV400	ICSV112 × (IS12611C × SC108-3), Zera-zera derivative	
49. MACIA	Zera-zera derivative	
50. MP531	Zera-zera derivative	

† For complete pedigrees please review the germplasm release notice in Crop Sciences and TAES.
‡ *R* = Restorer of male fertility, B = Maintainer of female lines (A), PR = Partial restorer.

comparisons among the 50 inbred lines: Matrix I, utilizing only data generated by 100 SSRs; Matrix II, using data generated by 1318 *Eco*RI/*Mse*I AFLPs (E/M AFLPs); Matrix III, using 496 *Pst*I/*Mse*I AFLPs (P/M AFLPs); Matrix IV, using 1313 markers (SSRs, P/M AFLPs, and E/M AFLPs) mapped every 1 or 2 cM on the genetic map; and Matrix V, using all 1914 markers included in this study. Dendrograms were created from each similarity matrix by the unweighted pair group method with arithmetic average (UPGMA) (Sneath and Sokal, 1973). To determine how accurately the dendrograms represent the estimates of genetic similarity among the genotypes, a cophenetic matrix was generated for each of the dendrograms and compared with the corresponding similarity matrix by the Mantel matrix correspondence test (Mantel, 1967). To compare the dendrograms obtained with different data sets, their corresponding cophenetic matrices were used to estimate cophenetic correlations among them by the same Mantel statistic. Significance of *Z* from the Mantel test was determined by comparing the observed *Z* value with a critical *Z* value obtained after 1000 permutations. All procedures were performed by appropriate routines in NTSYSpc version 2.11a (Exeter Software, Setauket, NY). An estimate of the confidence limits for the grouping produced by each dendrogram was obtained by performing 2000 bootstrap resamplings in SAS (SAS Institute Inc.) and WinBoot (Yap and Nelson, 1996).

RESULTS

Allelic Diversity at SSR Loci in Sorghum Inbreds

Of the 777 total SSR alleles identified, 675, 450, and 355 were detected among the 19 exotic IS-lines, 18 elite R-lines, and 13 elite B-lines examined, respectively. On the basis of SSR markers, the number of alleles per locus (a/l) ranged from 2 (2% of all SSRs examined) to 19 (SSR marker *Xtxp304*). The average number of a/l for the whole set of inbred lines was 7.8. Among the elite B-lines, the maximum number of a/l was 7.0 with an average value of 3.6. This average value decreased to 3.0 when the line BTx642 was excluded from the

Table 2. Mean, minimum, and maximum values of Dice coefficient of genetic similarity among 50 lines of sorghum determined by different sets of molecular markers.

Marker set	Number of alleles	Genetic similarity		
		Maximum	Minimum	Mean
Matrix I–SSR	777	0.810	0.00	0.302
Matrix II–*Eco*RI/*Mse*I AFLP	1318	0.911	0.00	0.606
Matrix III–*Pst*I/*Mse*I AFLP	496	0.896	0.00	0.628
Matrix IV–Select map positions	1313	0.890	0.00	0.578
Matrix V–All markers	2591	0.886	0.00	0.584

analysis. The inbred BTx642 is the only elite B-line derived exclusively from Durra germplasm. Of the 100 SSRs tested, 8% were monomorphic among the B-lines examined. The corresponding average number of a/l among the elite R-lines was 4.5 with a maximum of 10.0 and 3% of the SSRs being monomorphic. The greatest allelic diversity was detected among the 19 exotic lines with an average number of a/l of 6.8 and no SSRs being monomorphic. Of the three groups, the maintainer elite B-line parents possess the least allelic diversity and neither the B- nor R-lines capture all the genetic diversity present in the exotic lines. Furthermore, for a given SSR locus, nearly all of the alleles present among the elite germplasm (B- and R-lines) were present among the alleles observed in the exotic IS lines.

Cluster Analyses

The GS data calculated for each set of markers is summarized in Table 2. Maximum values of GS ranged from 0.911 within two Kafir lines determined by Matrix II (E/M AFLP markers) to 0.81 for the same Kafir lines determined by Matrix I (SSR markers). The minimum value of GS for the overall set of markers was 0.0 since all the markers evaluated were polymorphic between two lines included in this study. On average, SSRs produced lower estimates of GS than AFLP markers. Estimates of cophenetic correlations for the dendrograms obtained with each set of markers are illustrated in Table 3. Overall, the cophenetic values ranged from very high (0.98 for Matrix IV, markers every 1–2 cM across the genome) to high for SSRs (0.83 for Matrix I). When comparing similarity matrices or cophenetic matrices between the five different markers sets, in general, matrices generated by SSRs showed the lowest correlation.

Associations among the 50 sorghum lines based on cluster analysis of GS_{ij} indices are illustrated in Fig. 1. All five molecular marker sets were able to uniquely classify each of the 50 inbred lines included in this study. In all five dendrograms, the major clustering of lines corresponded to sorghum race–working groups rather than the R- or B-line classification used to define hybrid combinations. The groups identified by the cluster analysis are Durra, Zera-zera, Feterita, Kafir, and Kafir-Milo, and all of them are genetically distant from the only Margaritiferum line, IS3620C. There were discrepancies in the classification of those lines that have more than one working group in their genetic background. For example, RTx436, a line derived from SC120-6/Tx7000 with a backcross to Tx7000, has Zera-zera and Kafir-Milo in its genetic background. This line was grouped with Feterita lines by the analyses based on Matrix I and Matrix II, and it was grouped with other Kafir-Milo lines by the other sets of markers. Lines derived from crosses between Zera-zera and Kafir or Caudatum were grouped together in a single cluster close to the Kafir cluster by all the marker sets except Matrix I. Analysis with SSR markers placed two of these lines (Tx626 and Tx635) close to the Kafir cluster and the two others (Tx623 and Tx432) within the Zera-zera cluster. The exception was the group that includes the Feterita lines. All of these lines have complex pedigrees resulting from crossing lines from different working groups but all include a Feterita line as a parent. Despite these complex pedigrees, all marker sets grouped the Feterita-related lines in a single cluster. Lines of Caudatum origin were distributed across several clusters and never grouped together with any set of markers examined. Tx3042 and Tx7000 that are designated as Kafir-Milo lines, grouped with the Kafir cluster rather than with other Kafir-Milo lines suggesting that the proportion of their genomes derived from Kafir origin is substantial.

The tree best supported by the bootstrap analysis was obtained with the set of markers selected every 1- to 2-cM across the sorghum genome (Matrix IV). On the basis of this data set, the average GS within elite R-lines (0.61) is similar to the average GS among elite B- and R-lines (0.59), and this estimate is lower than the average GS within any of the working groups examined (Table 4). Overall, the most homogeneous group is Kafir (average GS of 0.82), and this group, on average, is more closely related to Kafir-Milo (average GS of 0.61) than to Zera-zera, Feterita, or Durra (average GS of 0.51 to 0.54). The dendrogram produced by Matrix III (P/M AFLPs) gave similar results. The weakest cluster

Table 3. Correlation between cophenetic matrices (above diagonal) and similarity matrices (below diagonal) obtained with different marker sets. Cophenetic correlation coefficients for the dendrograms are underscored on the diagonal.

Matrix	SSR I	E/M AFLP II	P/M AFLP III	Sel. map pos. IV	All markers V
I	0.830	0.713	0.700	0.710	0.730
II	0.758	0.885	0.958	0.970	0.982
III	0.747	0.954	0.903	0.968	0.970
IV	0.780	0.980	0.984	0.980	0.984
V	0.800	0.990	0.975	0.998	0.894

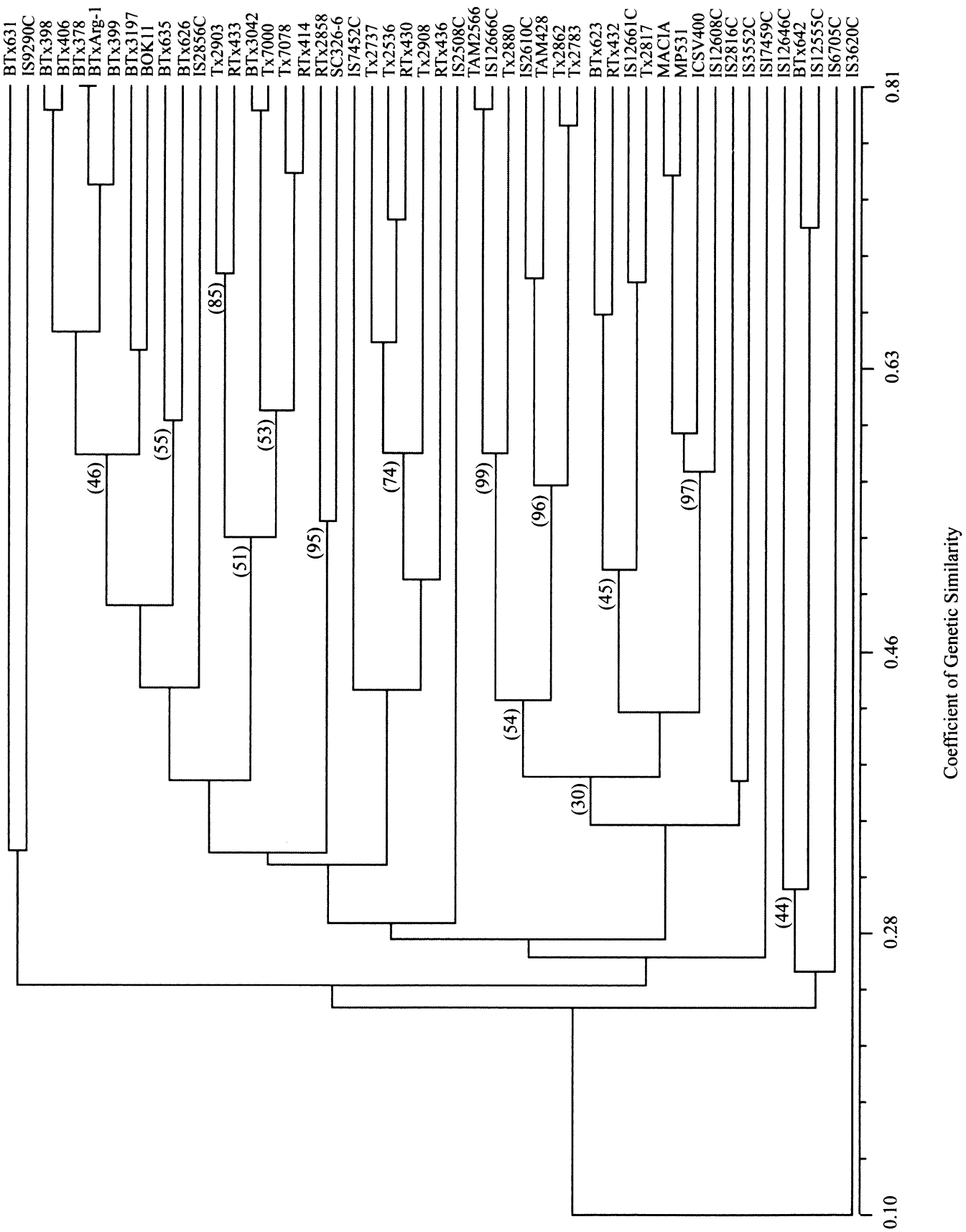

Fig. 1. Dendrograms of 50 sorghum inbreds revealed by cluster analysis of genetic similarity estimates for different marker sets. a. Determined by Matrix I generated by 100 SSRs.

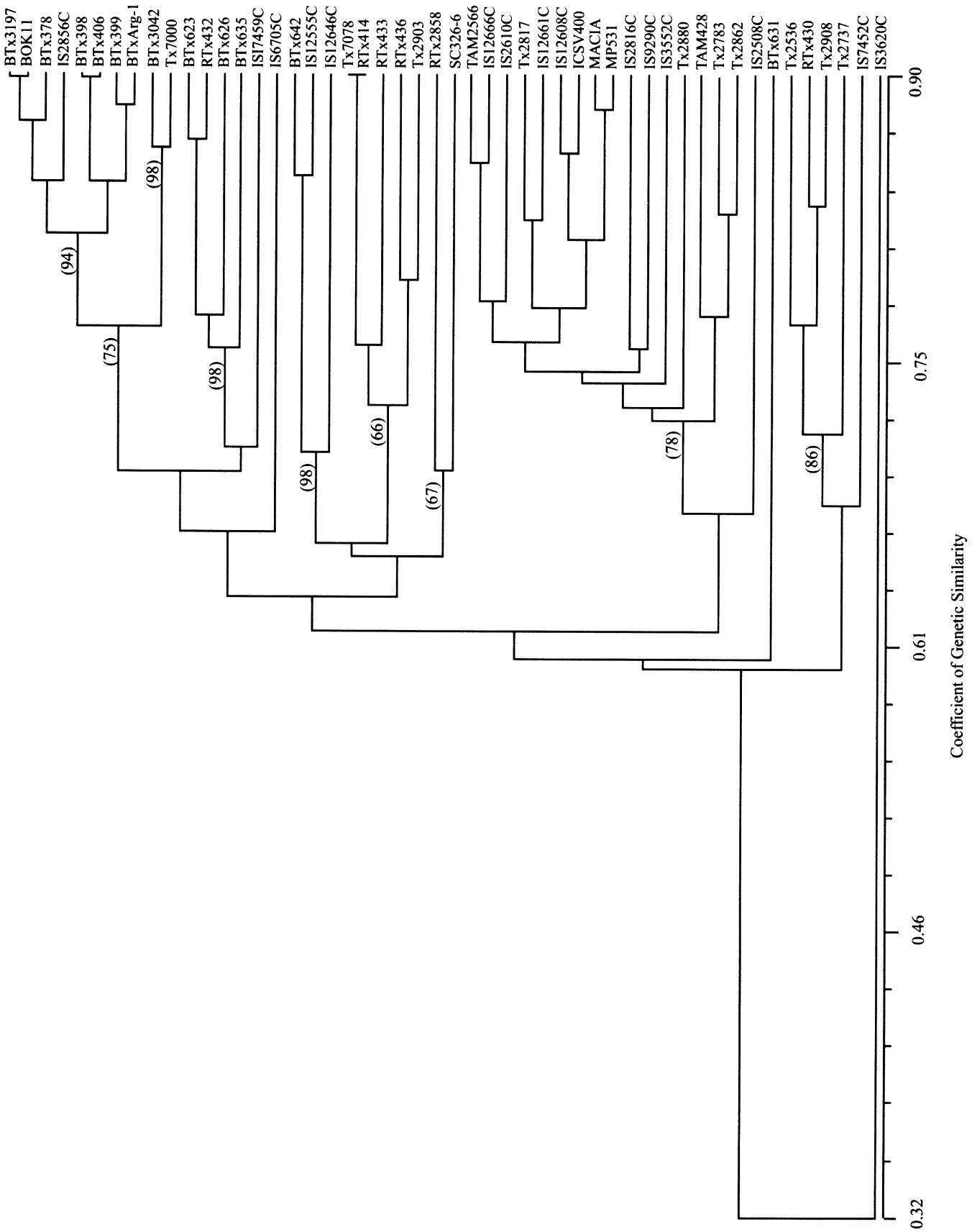

Fig. 1. Dendrograms of 50 sorghum inbreds revealed by cluster analysis of genetic similarity estimates for different marker sets. b. Determined by Matrix III generated by 496 P/M AFLPs. Bootstrap confidence intervals are included in parenthesis at the junctions of each cluster.

Table 4. Average genetic similarity within and between working groups based on molecular markers selected for their map position (Matrix IV).

Working Group	Number of lines	Zera-zera	Feterita	Kafir	Kafir-Milo	Durra	Z-z × K,C†
Zera-zera	13	0.728					
Feterita	4	0.562	0.722				
Kafir	8	0.513	0.538	0.822			
Kafir-Milo	7	0.568	0.573	0.608	0.694		
Durra	3	0.537	0.538	0.530	0.574	0.680	
Z-z × K,C†	4	0.651	0.566	0.684	0.610	0.523	0.788

† Lines derived from crosses between Zera-zera and Kafir or Zera-zera and Caudatum.

based on the bootstrap analysis was observed by only SSRs (Matrix I).

Comparison between Marker Sets

Comparisons between dendrograms obtained with different sets of markers, through cophenetic correlations of cophenetic or similarity matrices, were all highly significant. However, marker sets with poor and/or non-uniform coverage of the genome [e.g., sets based strictly on SSRs (Matrix I) or E/M AFLPs (Matrix II)], showed a slightly different classification of germplasm although Mantel tests did not detect these differences. The 100 SSRs included in this study covered 92 unique map positions vs. approximately 355 for P/M AFLPs and 825 for E/M AFLPs. The set of markers with the best overall coverage of the sorghum genome represented 1313 unique genetic map positions (Matrix IV). Estimates of GS and classification of germplasm obtained with data derived from this last set of markers, Matrix IV, were very similar to the one obtained with the P/M AFLP markers alone (Matrix III).

DISCUSSION

In general, the level of allelic diversity revealed among the germplasm included in this study is high, but is due mainly to the allelic diversity present among the exotic lines. The average number of alleles per locus identified in this study is similar to the average reported for maize (Romero-Severson et al., 2001; Matsuoka et al., 2002), higher than the average (5.9 a/l) previously reported in elite sorghum lines (Smith et al., 2000), and lower than the average (8.7 a/l) reported in landraces from Southern Africa (Uptmoor et al., 2003) and accessions from the world germplasm collection (Djé et al., 2000; Grenier et al., 2000).

In this study, the elite B-lines displayed only 52% of the allelic diversity present in the selected group of exotic lines from the World Collection while elite R-lines retained 67% of the allelic diversity of the exotic lines. The lower diversity among elite B-lines vs. R-lines is not surprising since B-lines must produce good male-sterile A-lines and the development of new A/B-lines is more difficult and time consuming compared to R-line development. Therefore, B-line development is more restrictive and slower to incorporate new germplasm. In addition, most of the female lines are Kafir derivatives. Despite the effort that would be required, the present results indicate that efforts to increase the genetic diversity of sorghum breeding lines should emphasize the introgression and development of new, diverse B-line germplasm.

Cluster analysis based on GS among the 50 lines examined showed a clear demarcation of the germplasm according to their working group and not to their B- or R-line classification. Sorghum working groups represented in this study included Kafir, Feterita, Zera-zera, Kafir-Milo, Durra, Caudatum, Margaritiferum, and Guinea, and lines developed from inter-group crosses, like Zera-zera × Kafir. All groups but Caudatum were identified as independent clusters. There was only one representative from the Margaritiferum group and one from the Guinea working group. These two lines were clearly differentiated from the other groups. Most of the B-lines included in this study were derived from Kafir germplasm, consequently they were grouped together in the Kafir cluster. This can give the impression that the classification was following the B- or R-line status; however, the Kafir cluster also includes the R-lines with Kafir background and does not include the B-lines derived from Durra or Zera-zera germplasm. Prior efforts to identify heterotic groups in sorghum, whether based on working groups, B- or R-line status, or any other factor, have been inconclusive (Gilbert, 1994). Therefore, for logistical reasons, sorghum breeders have traditionally kept B- and R-line development programs separate and then tested for combining ability between the two groups. The results of this study suggest that elite sorghum germplasm should be grouped by genetic background and not by existing B- or R-line classification. In fact, the traditional inter-group crossing approach utilized by public breeders may actually dilute potential heterotic patterns. The results reported by Ahnert et al. (1996) also support this conclusion. They studied genetic diversity in a group of elite proprietary inbred lines from a single seed company. The B-lines were grouped in a single cluster separate from the R-lines. However, most of the B-lines in that study came from Kafir germplasm. The R-lines clustered in two main groups; one that included Feterita lines, and the other Zera-zera lines (Ahnert et al., 1996).

Temperate maize is probably the most successful story of hybrid breeding in a domesticated crop. The heterotic groups Reid Yellow Dent and Lancaster have been maintained as separate groups while new lines are developed by selfing a cross between inbreds from within the same heterotic group. In this system, the genetic diversity within a heterotic group has been reduced while diversity estimates between groups have markedly increased through modern breeding practices

(Labate et al., 1997, 1999). If the classification of the sorghum lines suggested by this study corresponds to heterotic groups, a more efficient way to produce new inbred lines with good combining ability may involve crossing lines from within the same working group. This also implies that new female lines could be created by crossing R-lines with B-lines from within a working group, thereby increasing the genetic diversity of the female side of the hybrid. Additional research is necessary to determine the relationship between the groups detected in this study based on genetic similarity and heterotic groups in sorghum. In tropical corn where heterotic patterns are not clearly defined, high correlations have been found between genetic diversity and specific combining ability and heterosis (Betran et al., 2003). A similar study using lines representing the groups identified in this study can serve as a basis for describing heterotic groups in sorghum. Associations between single cross hybrid performance and genetic similarities between and within groups can help to identify heterotic groups in sorghum.

The present results also emphasize the importance attributable to the nature and genome distribution of molecular markers used in genetic diversity studies. The lower estimates of GS obtained with SSRs compared with AFLP markers can be explained by the higher mutation rate of this type of marker. However, this does not explain the differences in the classification of the germplasm. It is expected that a higher number of markers will provide a more precise estimate of genetic relationship, but the distribution of these markers over the genome is equally as important. Diversity studies based on a set of markers with poor coverage of the genome can give a different classification of the germplasm. Over-representation of certain genomic regions, such as heterochromatic centromeres, also affects the estimates of GS and therefore, the classification of the germplasm. One fourth of the E/M AFLP markers included in this study were mapped on the centromeric regions of each chromosome (Menz et al., 2002), consequently these regions have a higher "weight" in the classification of the germplasm obtained with strictly E/M AFLP markers. With the creation of saturated genetic maps for many crop species, the ability to select molecular markers that adequately cover the entire genome without under- or over-representation of any region is now feasible. The development of a genetic map for sorghum that contains nearly 3000 markers (Menz et al., 2002) was critical for a detailed molecular assessment of the genetic diversity in sorghum. For those crops where saturated genetic maps do not yet exist, *Pst*I/*Mse*I AFLP genetic markers provide a potential alternative based on the distribution of this marker type in largely unmethylated, gene rich regions of genomes (Young et al., 1999; Vuylsteke et al., 1999). Studies in maize have reached similar conclusions concerning the use of AFLP markers as a tool for genetic diversity assessment (Vuylsteke et al., 2000).

It is our intent that the present molecular diversity analysis will assist sorghum breeders in inbred line development and facilitate the development of well-defined heterotic groups for this crop. It is also our intent to continue this molecular classification for additional sorghum lines (both exotic and domesticated materials) to identify new sources of alleles for sorghum improvement.

ACKNOWLEDGMENTS

The authors thank Julie McCollum for her technical assistance and Dr. Javier Betran (Texas A&M University) for his guidance in the statistical analysis. Some of the material presented in this manuscript is based upon work supported by the National Science Foundation under Grant no. 0077713 (JEM, PEK). Additional support for research by the authors (RRK, MAM) was derived from the US Department of Agriculture, Agricultural Research Service (USDA-ARS), from the Texas Agricultural Experiment Station, the Perry Adkisson Chair in Plant Biology (JEM) and the Advanced Technology Program [000517-0242 (JEM, PEK)].

REFERENCES

Ahnert, D., M. Lee, D.F. Austin, C. Livini, S.J. Openshaw, J.S.C. Smith, K. Porter, and G. Dalton. 1996. Genetic diversity among elite sorghum inbred lines assessed with DNA markers and pedigree information. Crop Sci. 36:1385–1392.

Betran, F.J., J.M. Ribaut, D. Beck, and D. Gonzalez de Leon. 2003. Genetic diversity, specific combining ability, and heterosis in tropical maize under stress and nonstress environments. Crop Sci. 43: 797–806.

Brown, S.M., M.S. Hopkins, S.E. Mitchell, M.L. Senior, T.Y. Wang, R.R. Duncan, F. Gonzalez-Candelas, S. Kresovich. 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theor. Appl. Genet. 93:190–198.

Casa, A.M., S.E. Mitchell, O.S. Smith, J.C. Register, III, S.R. Wessler, and S. Kresovich. 2002. Evaluation of *Hbr* (MITE) markers for assessment of genetic relationships among maize (*Zea mays* L.) inbred lines. Theor. Appl. Genet. 104:104–110.

Djé, Y., D. Forcioli, M. Ater, C. Lefébvre, and X. Vekemans. 1999. Assessing population genetic structure of sorghum landraces from North-western Morocco using allozyme and microsatellite markers. Theor. Appl. Genet. 99:157–163.

Djé, Y., M. Heurtz, C. Lefébre, and X. Vekemans. 2000. Assesment of genetic diversity within and among germplasm accessions in cultivated sorghum using microsatellite markers. Theor. Appl. Genet. 100:918–925.

Dubreuil, P., P. Dufour, E. Krejci, M. Causse, D. de Vienne, A. Gallais, and A. Charcosset. 1996. Organization of RFLP diversity among inbred lines of maize representing the most significant heterotic groups. Crop Sci. 36:790–799.

Duncan, R.R., P.J. Bramel-Cox, and F.R. Miller. 1991. Contributions of introduced sorghum germplasm to hybrid development in the USA. p. 69–102. *In* H.L. Shands and L.E. Weisner (ed.) Use of plant introduction in cultivar development. Part 1. CSSA Spec. Publ. 17. CSSA, Madison, WI.

Fregene, M., A. Bernal, M. Duque, A. Dixon, and J. Tohme. 2000. AFLP analysis of African cassava (*Manihot esculenta* Crantz) germplasm resistant to the cassava mosaic disease (CMD). Theor. Appl. Genet. 100:678–685.

Gilbert, M.L. 1994. Identification and search for heterotic patterns in sorghum. p. 117–126. *In* D. Wilkinson (ed.) Proc. Annu. Corn and Sorghum Ind. Res. Conf., 49th, Chicago, IL. 9–10 Dec. 1994. Am. Seed Trade Assn., Washington, DC.

Grenier, C., M. Deu, S. Kresovich, P.J. Bramel-Cox, and P. Hamon. 2000. Assesment of genetic diversity in three subsets constituted from the ICRISAT sorghum collection using random vs non-random sampling procedures B. Using molecular markers. Theor. Appl. Genet. 101:197–202.

Hallauer, A.R. 1999. Temperate maize and heterosis. p. 353–361. *In* J. Coors and S. Pandey (ed.) Genetics and Exploitation of Heterosis

in Crops, CIMMYT, Mexico City.17–22 Aug. 1997. ASA, Madison, WI.
Hongtrakul, V., G.M. Huestis, and S.J. Knapp. 1997. Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: Genetic diversity among oilseed inbred lines. Theor. Appl. Genet. 95:400–407.
Klein, P.E., R.R. Klein, S.W. Cartinhour, P.E. Ulanch, J. Dong, J.A. Obert, D.T. Morishige, S.D. Schlueter, K.L. Childs, M. Ale, and J.E. Mullet. 2000. A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. Genome Res. 10:789–807.
Kramer, N.W. 1987. Grain sorghum production and breeding. Historical perspectives to future prospects. p. 1–9. *In* D. Wilkinson (ed.) Proc. 42nd Annual Corn and Sorghum Res. Conf., Chicago, IL. Am. Seed Trade Assoc., Washington, DC.
Labate, J., K.R. Lamkey, M. Lee, and W.L. Woodman. 1997. Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. 37:416–423.
Labate, J., K.R. Lamkey, M. Lee, and W.L. Woodman. 1999. Population genetics of increased hybrid performance between two maize populations under reciprocal recurrent selection. p. 127–137. *In* J. Coors and S. Pandey (ed.) Genetics and Exploitation of Heterosis in Crops, CIMMYT, Mexico City. 17–22 Aug. 1997. ASA, Madison, WI.
Lima, M.L.A., A.A.F. Garcia, K.M. Oliveira, S. Matsuoka, H. Arizono, C.L. de Souza, Jr., and A.P. de Souza. 2002. Analysis of genetic similarity detected by AFLP and coefficient of parentage among genotypes of sugar cane (*Saccharum* spp.). Theor. Appl. Genet. 104:30–38.
Mantel, M. 1967. The detection of disease clustering and generalized regression approach. Cancer Res. 27:209–220.
Matsuoka, Y., S.E. Mitchell, S. Kresovich, M. Goodman, and J. Doebley. 2002. Microsatellites in *Zea*- variability, patterns of mutations, and use for evolutionary studies. Theor. Appl. Genet. 104:436–450.
Menz, M.A., R.R. Klein, J.E. Mullet, J.A. Obert, N.C. Unruh, and P.E. Klein. 2002. A high-density genetic map of *Sorghum bicolor* (L.) Moench based on 2926 AFLP®, RFLP and SSR markers. Plant Mol. Biol. 48:483–499.
Nei, M., and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76:5269–5273.
Pejic, I., P. Ajmone-Marsan, M. Morgante, V. Kozumplick, P. Castiglioni, G. Taramino, and M. Motto. 1998. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theor. Appl. Genet. 97:1248–1255.
Quinby, J.R. 1971. A triumph of research….sorghum in Texas. Texas A&M Univ. Press, College Station, TX.
Quinby, J.R. 1974. Sorghum Improvement and the Genetics of Growth. Texas A&M Univ. Press, College Station, TX.
Romero-Severson, J., J.S.C. Smith, J. Ziegle, J. Hauser, L. Joe, and G. Hookstra. 2001. Pedigree analysis and haplotype sharing within diverse groups of *Zea mays* L. inbreds. Theor. Appl. Genet. 103: 567–574.
Rooney, W.L., and C.W. Smith. 2001. Techniques for developing new cultivars. p. 329–347. *In* C.W. Smith and R.A. Frederiksen (ed.) Sorghum: Origin, history, technology, and production. John Wiley & Sons New York.
Saghai-Maroof, M.A., G.P. Yang, Q. Zhang, and K.A. Gravois. 1997. Correlation between molecular marker distance and hybrid performance in U.S. Southern long grain rice. Crop Sci. 37:145–150.
Smith, J.S.C., S. Kresovich, M.S. Hopkins, S.E. Mitchell, R.E. Dean, W.L. Woodman, M. Lee, and K. Porter. 2000. Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Sci. 40:226–232.
Sneath, P.H.A., and R.R. Sokal. 1973. Numerical taxonomy. Freeman. San Francisco.
Stephens, J.C., and R.F. Holland. 1954. Cytoplasmic male-sterility for hybrid sorghum seed production. Agron. J. 46:20–23.
Stephens, J.C., F.R. Miller, and D.T. Rosenow. 1967. Conversion of alien sorghums to early combine genotypes. Crop Sci. 7:396.
Stuber, C.W. 1994. Heterosis in plant breeding. Plant Breed. Rev. 12:227–251.
Tao, Y., J.M. Manners, M.M. Ludlow, and R.G. Henzell. 1993. DNA polymorphisms in grain sorghum (*Sorghum bicolor* (L.) Moench). Theor. Appl. Genet. 86:679–688.
Taramino, G., R. Tarchini, S. Ferrario, M. Lee, and M.E. Pe. 1997. Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor.* Theor. Appl. Genet. 95:66–72.
Ubi, B.E., R. Kölliker, M. Fujimori, and T. Komatsu. 2003. Genetic diversity in diploid cultivars of Rhodesgrass determined on the basis of amplified fragment length polymorphism markers. Crop Sci. 43:1516–1522.
Uptmoor, R., W. Wenzel, W. Friedt, G. Donaldson, K. Ayisi, and F. Ordon. 2003. Comparative analysis on the genetic relatedness of *Sorghum bicolor* accessions from Southern Africa by RAPDs, AFLPs, and SSRs. Theor. Appl. Genet. 106:1316–1325.
Vierling, R.A., Z. Xiang, C.P. Joschi, M.L. Gilbert, and H.T. Nguyen. 1994. Genetic diversity among elite sorghum lines revealed by restriction fragment length polymorphisms and random amplified polymorphic DNAs. Theor. Appl. Genet. 87:816–820.
Vuylsteke, M., R.R. Antonise, E. Bastiaans, M.L. Senior, S.W. Stuber, A.E. Melchinger, T. Lübberstedt, X.C. Xia, P. Stam, M. Zabeau, and M. Kuiper. 1999. Two high-density AFLP linkage maps of *Zea mays* L.: Analysis of distribution of AFLP markers. Theor. Appl. Genet. 99:921–935.
Vuylsteke, M., R. Mank, B. Brugmans, P. Stam, and M. Kuiper. 2000. Further characterization of AFLP data as a tool in genetic diversity assessments among maize inbred lines. Mol. Breed. 6:265–276.
Yap, I.V., and R.J. Nelson. 1996. WinBoot: A program for performing bootstrap analysis of binary data to determine the confidence limits of UPGMA-based dendograms. International Rice Research Institute. Manila, Philippines.
Young, W.P., J.M. Schupp, and P. Keim. 1999. DNA methylation and AFLP marker distribution in the soybean genome. Theor. Appl. Genet. 99:785–790.
Zhu, J., M.D. Gale, S. Quarrie, M.T. Jackson, and G.J. Bryan. 1998. AFLP markers for the study of rice biodiversity. Theor. Appl. Genet. 96:602–611.