Theor Appl Genet (2005) 111: 1288–1299
DOI 10.1007/s00122-005-0055-7

ORIGINAL PAPER

P. R. Laborda · K. M. Oliveira · A. A. F. Garcia
M. E. A. G. Z. Paterniani · A. P. de Souza

Tropical maize germplasm: what can we say about its genetic diversity in the light of molecular markers?

Received: 15 April 2004 / Accepted: 14 July 2005 / Published online: 25 August 2005

Abstract Knowledge about genetic variability of a crop allows for more efficient and effective use of resources in plant improvement programs. The genetic variation within temperate maize has been studied extensively, but the levels and patterns of diversity in tropical maize are still not well understood. Brazilian maize germplasm represents a very important pool of genetic diversity due to many past introductions of exotic material. To improve our knowledge of the genetic diversity in tropical maize inbred lines, we fingerprinted 85 lines with 569 AFLP bands and 50 microsatellite loci. These markers revealed substantial variability among lines, with high rates of polymorphism. Cluster analysis was used to identify groups of related lines. Well-defined groups were not observed, indicating that the tropical maize studied is not as well organized as temperate maize. Three types of genetic distance measurements were applied (Jaccard's coefficient, Modified Rogers' distance and molecular coefficient of coancestry), and the values obtained with all of them indicated that the genetic similarities were small among the lines. The different coefficients did not substantially affect the results of cluster analysis, but marker types had a large effect on genetic similarity estimates. Regardless of genetic similarity coefficient used, estimates based on AFLPs were poorly correlated with those based on SSRs. Analyses using AFLP and SSR data together do not seem to be the most efficient manner of assessing variability in highly diverse materials because the result was similar to using AFLPs alone. It was seen that molecular markers can help to organize the genetic variability and expose useful diversity for breeding purposes.

Communicated by D. A. Hoisington

P. R. Laborda · K. M. Oliveira · A. P. de Souza (✉)
Centro de Biologia Molecular e Engenharia Genètica (CBMEG),
Universidade Estadual de Campinas (UNICAMP),
Cidade Universitária "Zeferino Vaz", CP 6010,
CEP 13083-970 Campinas, SP, Brasil
E-mail: anete@unicamp.br
Tel.: +55-19-37881132
Fax: +55-19-37881089

A. A. F. Garcia
Departamento de Genética,
Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ),
Universidade de São Paulo (USP), CP 83,
CEP 13400-970 Piracicaba, SP, Brasil

M. E. A. G. Z. Paterniani
Instituto Agronômico de Campinas (IAC),
Av. Theodureto de Almeida Camargo, 1500,
CP 28, CEP 13020-902 Campinas, SP, Brasil

Introduction

Access to and use of genetic diversity is the basis of genetic improvement through breeding. In maize, a large and diverse gene pool permits the manipulation of different genotypes that has led to improved performance of hybrids in terms of yield, resistance to diseases, and other agronomic characteristics.

During the 1990s, restriction fragment length polymorphisms (RFLPs) were used to determine diversity in temperate European maize (Boppenmaier et al. 1993) and U.S. maize (Melchinger et al. 1991), and to assign lines to different heterotic groups. RFLPs, however, proved to be an expensive methodology, and molecular technology improvements made other kinds of markers more attractive. Random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs) and simple sequence repeats (SSRs) were quickly adopted for use in genetic analyses because they could provide data more efficiently. Despite being a simple technique, RAPDs were determined to be unreliable, and did not always correlate with the results obtained with other marker types in maize (Hahn et al. 1995; Pejic et al. 1998; Garcia et al. 2004) and other species (Doldi et al. 1997; Russell et al. 1997).

AFLPs and SSRs have remained very reproducible techniques, while having different features that affect their utility for genetic analyses. AFLPs, which may be applied to any plant species without previous knowledge

of DNA sequences, have already been proven to be an efficient marker, disclosing a great number of bands per single assay (Russell et al. 1997; Pejic et al. 1998; Garcia et al. 2004). SSRs are very abundant and dispersed throughout the genome, usually codominant in inheritance, and can uncover a great number of polymorphisms since multiallelic loci are very common (Chin et al. 1996). As a consequence of these special characteristics, many studies have been carried out using these two markers to access diversity in various species, such as sugarcane (Lima et al. 2002), wheat (Plaschke et al. 1995; Manifesto et al. 2001), cotton (Liu et al. 2000; Abdalla et al. 2001), soybean (Priolli et al. 2002) and sorghum (Smith et al. 2000; Ghebru et al. 2002).

Maize has also been extensively studied, revealing how important is temperate maize germplasm. Lines from the U.S. Corn Belt were analyzed in a large number of investigations (Smith et al. 1997; Senior et al. 1998; Lu and Bernardo 2001; Gethi et al. 2002). These studies showed the moderate diversity of temperate maize and were able to recover the relationships among the inbred lines by means of molecular techniques. Moreover, Enoki et al. (2002) used SSR analysis of 51 Japanese lines to compare them to genotypes introduced from Europe and the U.S.

Despite the extensive knowledge acquired about temperate maize germplasm, relatively few genetic diversity studies involving large samples of lines or populations of tropical maize have been published (Warburton et al. 2002; Reif et al. 2003a, b; Oliveira et al. 2004; Reif et al. 2004). Lanza et al. (1997) used RAPDs to predict maize single-cross performance among 18 inbred lines, and Barbosa et al. (2003), with the same lines, made a comparative study of genetic distances and single-cross performance using RAPDs, RFLPs, AFLPs and SSRs. The Agronomic Institute of Campinas (IAC) houses an important maize genebank that is composed of many genotypes developed from important tropical maize races, such as Cateto and Tuxpeño, as well as many lines obtained from populations developed by the Centro Internacional de Mejoramiento de Maiz y Trigo (CIMMYT) in Mexico. IAC also has one of the most important maize improvement programs in Brazil, the third largest maize producer in the world. However, the pedigrees of many lines are not available and no previous work has assessed the genetic variability in its genebank. Many crosses for the production of hybrids and new breeding populations are being made exclusively on the basis of phenotypic characters (Paterniani et al. 2000), requiring substantial resource investments. Understanding the level and organization of the genetic diversity of tropical materials would improve the efficiency of breeding and conservation.

The objectives of this study were to: (1) assess the genetic diversity in a set of tropical maize inbred lines, (2) compare the capacity of two types of molecular markers and three types of coefficients to determine variability and genetic relationships among lines, and (3) evaluate the power of AFLPs and SSRs to organize the germplasm and separate the lines for breeding purposes.

Materials and methods

Plant material

Eighty-five inbred lines were chosen from the Agronomic Institute of Campinas (IAC) genebank (Table 1). These inbreds represent the diversity of the tropical maize lines used in the Maize Breeding Program of the IAC. Among these 85 inbreds, there are recent lines, which were obtained from populations introduced from CIMMYT (named "L-") as well as Brazilian lines that have been used for a long time in Brazil. Unfortunately, the origins of some Brazilian lines are unknown because records have been lost.

DNA isolation and AFLP/SSR procedures

Young leaves from 15 plants of each genotype were collected, lyophilized and ground to powder. DNA extraction followed the CTAB method described in Hoisington et al. (1994).

AFLP

AFLP analysis profiles were performed as described by Vos et al. (1995), using the "AFLP Analysis Kit" (Life Technologies—GIBCO BRL). Genomic DNA (400 ng) was double-digested with *Eco*RI and *Mse*I. Restriction fragments were linked to adaptors and the ligation product preamplified for 20 cycles (94°C for 30 s, 56°C for 1 min, 72°C for 1 min) using primers carrying one selective nucleotide. *Eco*RI primers carrying three selective nucleotides were end-labeled with $\gamma[^{33}P]$-ATP (4,000 Ci/mmol), and mixed with unlabeled *Mse*I primer for hot selective amplification using the following cycles: 94°C for 30 s, 65°C (−0.7°C/cycle) for 30 s and 72°C for 1 min for 12 cycles, until the final temperature of 56°C was reached. All amplifications were carried out in a PTC-100 thermalcycler (MJ Research, Inc.). Reaction products and formamide dye (3.5 µl, 1:1) were loaded in 6% denaturing polyacrylamide gels and electrophoresed (Sequi-Gen GT- Nucleic Acid- Electrophoresis Cell/BIO RAD Apparatus of Electrophoresis) for 4 h at 75 W. Samples were visualized by autoradiography and manually scored in a conservative manner where all ambiguous bands were discarded. Twenty AFLP primer-enzyme combinations were tested and nine combinations were selected.

SSR

The SSR primer sequences were obtained from the Maize Genetics and Genomics Database site (http://

Table 1 Maize inbred lines maintained in the Agronomic Institute of Campinas, the material they were selected from and their origins

Inbred line	Selected from	Origin	Inbred line	Selected from	Origin
AL - 124	Cateto race	IAC	L - 111	Pop. 26	CIMMYT
AL - 218	Cateto race	IAC	L - 112	Pop. 26	CIMMYT
AL - 491	Cateto race	IAC	L - 114	Pop. 26	CIMMYT
AL - 516	Cateto race	IAC	L - 116	Pop. 27	CIMMYT
AL - 535	Cateto race	IAC	L - 117	Pop. 24	CIMMYT
AL - 604	Cateto race	IAC	L - 118	Pop. 27	CIMMYT
AL - 614	Cateto race	IAC	L - 120	Pop. 28	CIMMYT
AL - 673	Cateto race	IAC	L - 121	Pop. 27	CIMMYT
AL - 745	Cateto race	IAC	L - 123	Pop. 27	CIMMYT
AL - 758	Cateto race	IAC	L - 126	Pop. 27	CIMMYT
IA - 278	Cateto race	IAC	L - 128	Pop. 24	CIMMYT
IA - 606	Cateto race	IAC	L - 130	ACROSS 7543	CIMMYT
IA - 2938	Cateto race	IAC	L - 131	ACROSS 7543	CIMMYT
IA - 3040	Cateto race	IAC	L - 132	Pool 23	CIMMYT
IAC - B	Tehua race, Tx 303	IAC	L - 134	Pop. 24	CIMMYT
IP - 48	Cateto race	IAC	L - 137	Pop. 36	CIMMYT
IP - 301	Cateto race	IAC	L - 155	Pop. 25	CIMMYT
IP - 330	Cateto race	IAC	L - 156	Pop. 36	CIMMYT
IP - 365	Cateto race	IAC	L - 157	Pop. 27	CIMMYT
IP - 398	Cateto race	IAC	L - 158	Pop. 27	CIMMYT
IP - 661	Cateto race	IAC	L - 160	Pop. 28	CIMMYT
IP - 701	Tuxpeño race	IAC	L - 161	Pop. 26	CIMMYT
IP - 3644	Cateto race	IAC	L - 162	Pop. 26	CIMMYT
IP - 3668	Cateto race	IAC	L - 163	Pop. 26	CIMMYT
IP - 3854	Cateto race	IAC	L - 164	Pop. 27	CIMMYT
IP - 3855	Cateto race	IAC	L - 165	Pop. 27	CIMMYT
IP - 3999	Cateto race	IAC	L - 166	Pop. 28	CIMMYT
IP - 4022	Cateto race	IAC	L - 167	Pop. 36	CIMMYT
L - 1	MJ 268	CIMMYT	L - 168	Pop. 24	CIMMYT
L - 2	MJ 274	CIMMYT	L - 169	Pop. 26	CIMMYT
L - 3	Pop. 24	CIMMYT	L - 170	Pop. 27	CIMMYT
L - 4	Pop. 24	CIMMYT	L - 171	Pop. 28	CIMMYT
L - 5	Pop. 26	CIMMYT	L - 172	Pop. 28	CIMMYT
L - 6	Pop. 26	CIMMYT	PM - 219	Tuxpeño race	IAC
L - 8	Pop. 28	CIMMYT	PM - 308	Tuxpeño race	IAC
L - 9	Pop. 36	CIMMYT	PM - 518	Tuxpeño race	IAC
L - 10	Pop. 36	CIMMYT	PM - 624	Tuxpeño race	IAC
L - 11	Pop. 27	CIMMYT	PM - 888	Tuxpeño race	IAC
L - 13	Pop. 26	CIMMYT	PM - 2837	Tuxpeño race	IAC
L - 100	Pool 27	CIMMYT	SLP - 103	Tuxpeño race	IAC
L - 101	Pool 27	CIMMYT	SLP - 365	Tuxpeño race	IAC
L - 105	Pop. 26	CIMMYT	VER - 266	Tuxpeño race	IAC
L - 110	Pop. 24	CIMMYT			

* *Pool 23* germplasm from Mexico, Colombia, the Caribbean, India, Thailand and Philippines; *Pool 27* germplasm from US, China, Lebanon, and Europe; *Pop. 24* Antigua Veracruz, Tuxpeño race; *Pop. 25* Blanco Cristalino-3; Pool 23 (already mentioned); *Pop. 26* Mezcla Amarilla; Tuxpeño race, Cuban flints, Antigua group, ETO Amarillo, Corn Belt x Tuxpeño crosses and Pool 21 (includes different American germplasm sources); *Pop. 27* Amarillo Cristalino-1;Tuxpeño race, Cuban flints, ETO Amarillo, Pool 25 (includes germplasm from Mexico, Central America, the Caribbean, Ecuador, Colombia and Argentina); *Pop 28* Amarillo Dentado, Tuxpeño race, ETO Amarillo, Caribbean and Brazilian germplasms, Pool 26 (includes germplasm from Central America, Mexico, Asia, Colombia, the Caribbean, the US Corn Belt); *Pop. 36* Cogollero, Caribbean composite including material from Pool 26 (already mentioned) and Pool 22 (includes germplasm from Mexico, Cuba, Dominican Republic, Antigua, Brazil, Peru, Ecuador, Argentina, Colombia, India, Puerto Rico, and Central America)

www.maizegdb.org). To select the best SSRs, primers were screened across three sets of 3, 8 or 28 lines. A total of 215 microsatellites were tested, and classifications according to amplification quality and genotyping difficulty were made. From the 215 microsatellites tested, 50 were chosen for the diversity analysis. The amplification protocol included 50 ng of DNA, 1×reaction buffer (20 mM Tris–HCl, 50 mM KCl pH 8.4), 2 mM $MgCl_2$, 100 μM of each dNTP, 0.5 U of *Taq* DNA polymerase (Invitrogen) and 0.2 μM of each primer. All amplifications were carried out using a PTC-100 thermalcycler (MJ Research, Inc.) and the "touchdown" program described by Senior et al. (1998). Samples were electrophoresed in either 0.5×TBE (Sambrook et al. 1989), 4% agarose/Metaphor or 1×TBE, 6% polyacrylamide gels, depending on the genotyping difficulty of each microsatellite. Horizontal electrophoresis was conducted at 170 V for 1.5 h in HORIZON 20:25 gel system (GIBCO BRL) and products were visualized by ethidium bromide staining. Vertical electrophoresis was conducted at 90 W for 2 h using a Model S2001 Sequencing Gel Electrophoresis Apparatus (Life Technologies—GIBCO BRL),

and samples were detected by silver staining according to Creste et al. (2001).

Data analysis

AFLP and SSR gels were scored for the presence/absence of bands, generating two binary matrices. The SSR matrix was also converted to a table of allelic frequencies. Polymorphic index content (PIC), also named gene diversity or expected heterozigosity (Nei 1987), was calculated for SSRs according to the following formula:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{n} f_i^2 \quad (1)$$

where f_i is the frequency of the ith allele. PIC values show how powerful a locus can be to discriminate samples, considering not only the number of alleles but also their frequencies.

Preliminary diversity analyses among all the lines were based on Jaccard's similarity (Jaccard 1908) and modified Rogers' distance (MRD) (Wright 1978). Jaccard's coefficient is one of the most commonly used coefficients for dominant marker data (Doldi et al. 1997; Lanza et al. 1997; Lima et al. 2002; Barbosa et al. 2003; Meyer et al. 2004), although it can also be used with codominant marker data (Li et al. 2001). MRD has Euclidian metric properties and is largely used for SSRs where allelic frequencies and not only the presence/absence of bands are available.

Jaccard's similarities were calculated using the binary matrices of AFLP, SSR, and combined AFLP and SSR data, the latter to study the possibility of using the data of different marker types simultaneously. MRD was calculated using only SSR data, since the use of this coefficient with AFLPs would not be as informative as it is with SSRs, and a comparison between these two markers would, then, not be enlightening.

In order to compare the ability of different ways in organizing genetic diversity in tropical maize, the molecular coefficient of coancestry, f_{AB}^{M}, (Bernardo 1993) was calculated using AFLPs, SSRs, and both together using the formula:

$$f_{AB}^{M} = \frac{S_{AB} - \frac{1}{2}(\delta_{A.} + \delta_{B.})}{1 - \frac{1}{2}(\delta_{A.} + \delta_{B.})}, \quad (2)$$

where S_{AB} is the similarity value obtained by Jaccard's coefficient, $\delta_{A.}$ is the average similarity between inbred A and unrelated inbreds, and $\delta_{B.}$ is the average similarity between inbred B and unrelated inbreds. This coefficient represents the probability of inbreds A and B being identical by descent (*ibd*) and is a measure of similarity. Inbreds that are *ibd* have the same allele (at a given locus) inherited from a common ancestral allele. Three unrelated inbreds (PM-308, PM-624 and PM-2837) were chosen to calculate $\delta_{A.}$ and δ_{B}, which show the proportion of alleles that are alike in state (*ais*), not *ibd*. As recommended by Bernardo (1993), negative values were set to zero.

NTSYSpc v. 2.1 (Rohlf 1997) was used to obtain the similarity matrices using Jaccard's coefficient. For computing the distance matrix, TFPGA v. 1.3 (Miller 1997) was used. The UPGMA method was applied to all of the clustering procedures and the respective cophenetic values were calculated for each of the dendrograms.

Correlations between AFLP and SSR matrices were obtained by means of Mantel's test (Mantel 1967). A bootstrap procedure (Tivang et al. 1994) with 1,000 units of re-sampling, was carried out for SSR data using medians instead of means, as recommended by Garcia et al. (2004). Modified Rogers' distance was used and each locus was considered a unit of re-sampling. Bootstrap procedures and results for AFLPs are described in Oliveira et al. (2004).

Results

Polymorphism and bootstrap validation

AFLP assays generated 638 bands, of which 569 were polymorphic. Although we used only nine primer pair combinations out of the 64 possible, the AFLP profiles could unambiguously distinguish each inbred line. The mean number of bands detected per combination was 63, ranging from 42 to 98 (Table 2), and the bootstrap procedures showed that the number of bands used was large enough for this type of study (Oliveira et al. 2004).

Two hundred and fifteen microsatellite primers were pre-screened in order to select the best amplification and genotyping conditions. The test that included 8 inbreds was the most efficient, being able to reveal the polymorphism among the inbreds while minimizing effort and time. Out of the 215 SSR primer pairs, 109 were tested using 8 lines, and 35 primer pairs were chosen. The other 15 were selected based on the test using 28 lines (58 pairs tested). None of the primers tested in 8 or 28 lines was monomorphic. The 50 loci studied were well

Table 2 Number of polymorphic AFLP bands according to the primer combinations used (table reproduced from Oliveira et al. 2004)

Primer combination	Total number of bands	Polymorphic bands
E + AAC/M + CTC	83	74
E + AAG/M + CTG	98	86
E + AAG/M + CTC	109	98
E + AAG/M + CAC	48	42
E + AAC/M + CAT	68	61
E + ACA/M + CAT	58	53
E + ACA/M + CTG	59	50
E + AAC/M + CAG	53	52
E + AAC/M + CTT	62	53
Total	638	569

distributed among the 10 maize chromosomes, ensuring a good sampling of the genome. A total of 262 bands were obtained with a mean of 5.2 alleles per locus, ranging from 2 to 14. Di- and tri-nucleotides motifs were the most abundant (Table 3). They represented 36% and 26%, respectively (AG repeats corresponded to 72% of di-nucleotides). Tetra-, penta- and hexa-nucleotides represented the remaining 38%. PIC values for SSR data ranged from 0.24 to 0.90 with a mean of 0.61 (Table 3).

Although this study used inbred lines that have been maintained for many years by self-fertilization, an unexpectedly high frequency of heterozygotes was observed—five loci showed more than 20 heterozygotes among the 85 lines (Table 3). The high number of heterozygotes seen in some SSR loci show that molecular markers are a powerful technique to investigate if inbreeding procedures are as effective at increasing homozygosity within lines as expected, and suggests that some SSR sites that were genotyped are under selection or that some pollen contamination occurred while selfing.

Bootstrap analysis demonstrated that the number of loci used ($n = 50$) was appropriate to access diversity reliably among the 85 lines. If a coefficient of variation of 10% were used (mean across all CVs of each pairwise

Table 3 Microsatellites markers used in the diversity analyses and their genomic location, class of repeat, number of alleles, PIC values and number of heterozygotes

Locus	Genomic location	Repeat	No. alleles	PIC	No. heterozygotes
bnlg 149	1.00	?	7	0.75	1
bnlg 1484	1.03	di	6	0.52	2
umc 1397	1.03	penta	4	0.55	12
umc 1297	1.05	di	6	0.63	7
umc 1395	1.05	di	3	0.53	10
umc 1122	1.06	tri	4	0.51	9
umc 1774	1.10	di	3	0.60	3
umc 1630	1.11	penta	5	0.45	11
umc 1422	2.02	tri	4	0.64	1
bnlg 1621b	2.03	di	10	0.87	16
phi 083	2.04	tetra	4	0.66	12
umc 2019	2.07	tri	3	0.49	11
bnlg 2077	2.07–2.08	di	13	0.90	7
umc 1230	2.09	tri	6	0.79	8
umc 1252	2.09	tri	2	0.31	32
nc 030	3.04	di	3	0.56	2
bnlg 1951	3.06	di	6	0.75	6
bnlg 2241	3.06	di	6	0.68	4
phi 046	3.08	tetra	2	0.49	0
umc 1639	3.09	penta	5	0.76	23
phi 072	4.00–4.01	tetra	7	0.61	14
nc 135	4.01	?	4	0.56	1
umc 1943	4.02	?	4	0.32	6
phi 021	4.03	di	5	0.56	2
umc 1650	4.09	tri	2	0.42	12
umc 1325	5.00	di	6	0.72	4
umc 1416	5.00	tri	2	0.50	10
umc 1221	5.04	di	9	0.80	21
umc 1524	5.06	hexa	5	0.67	9
umc 1646	5.07	penta	5	0.36	8
umc 1792	5.08	tri	6	0.79	9
bnlg 1600	6.00	di	5	0.77	0
umc 1857	6.04	tri	6	0.80	16
umc 1653	6.07	tetra	10	0.84	5
umc 1426	7.00	penta	3	0.54	14
phi 057	7.01	tri	3	0.49	5
phi 112	7.01	di	3	0.24	0
bnlg 1666	7.04	di	12	0.88	4
umc 1161	8.06	hexa	5	0.72	6
umc 1069	8.08	penta	8	0.72	10
umc 1638	8.09	hexa	6	0.65	4
bnlg 1724	9.01	di	6	0.48	22
phi 028	9.01	tri	4	0.58	1
phi 022	9.03	tetra	3	0.52	21
umc 1357	9.05	tri	4	0.65	19
umc 1733	9.06	tetra	4	0.52	3
umc 1804	9.07	di	14	0.87	13
phi 059	10.02	tri	2	0.50	16
bnlg 2336	10.04	di	4	0.43	2
umc 1640	10.07	penta	3	0.55	5

similarities/distances, obtained through bootstrap), as recommended by Halldén et al. (1994) and Tivang et al. (1994), 20 microsatellite loci would be enough to analyze the 85 genotypes with accuracy. By using 50 loci in this study, the coefficient of variation was reduced to approximately 6%, resulting in a very reliable set of SSR data.

Clustering, correlations and diversity levels

Clusters analyses showed some differences according to the marker applied, since the two markers showed different similarity matrices, and low correlations were observed. In addition, distinct levels of divergence were observed among lines depending on the marker used. Despite these disagreements among marker types, all dendrograms were similar in the sense that they did not reveal clear clustering of lines. Well-defined divergent groups could not be seen in any of the ways used to visualize clusters. The relationships among inbreds that were suggested by the prior information on pedigrees and populations of origin could not be perfectly recovered using any of the marker types, although CIMMYT-derived inbreds tended to form groups.

AFLP data identified groups of inbreds in the dendrograms more easily than SSR data did. More distinct grouping was observed in the trees constructed with AFLP-Jaccard data (e.g., Fig. 1), than in those produced using SSR data (e.g., Fig. 2). In the AFLP-Jaccard dendrogram, using a similarity value of approximately 0.52, three groups were observed—a cluster with 42 lines (group 1) and two smaller clusters with 6 lines (group 2) and 14 lines (group 3). In addition to these clear groups, 23 other lines did not group in major clusters. The lines developed from Pop. 26 (Mezcla Amarilla) and from Pop. 36 (Cogollero) all clustered in group 1, which also contained almost all CIMMYT-derived inbreds. Three Pop. 24 lines (Antigua Veracruz) and three Pop. 27 lines (Amarillo Cristalino)—all Tuxpeño germplasm—constituted group 2. Group 3 was composed of Cateto germplasm, with the exception of line VER-266 (Tuxpeño).

Additional differences were seen using different data. Inbreds that were considered identical with one marker could be discriminated with the other. AFLP analyses showed that L-158 and L-162 had a high degree of similarity ($sim_{jac} = 0.87$), which was not observed with SSRs ($sim_{jac} = 0.66$). On the other hand, SSRs showed that AL-491 and AL-516 were genetically identical ($sim_{jac} = 1.00$), but these lines were separated using AFLPs ($sim_{jac} = 0.86$). In contrast, AFLP- and SSR-based similarity coefficients revealed the same set of lines that appeared to be most divergent from the rest. From this group, PM-308, PM-624 and PM-2837 were used as the unrelated lines for calculation of molecular coefficients of coancestry.

In order to compare all the similarity/distance matrices to each other, correlations among each combination of marker type and coefficient were estimated (Table 4). As expected for lines selfed for several generations, the results obtained by Jaccard's coefficient and MRD for SSR data showed a high correlation ($r = -0.95$, a negative value as a consequence of a correlation made between a similarity and a distance matrix), indicating that these coefficients grouped lines in a very similar way. Nevertheless, as mentioned earlier, AFLP and SSR disagreed in some groupings and a low correlation ($r = 0.43$) was obtained for these markers using Jaccard's coefficient. The clusters obtained on the basis of molecular coefficient of coancestry also showed discrepancies when different marker systems were used. The correlation for both markers, considering this coefficient, was slightly higher ($r = 0.48$), but also demonstrated substantial lack of agreement. Correlations between the molecular coefficient of coancestry and Jaccard's coefficient were high for both markers (AFLP-Jaccard/AFLP-f^{M}_{AB} = 0.88 and SSR-Jaccard/SSR-f^{M}_{AB} = 0.87). In addition, an interesting characteristic was noticed when matrices containing both data were compared to AFLP and SSR separately. Correlations among AFLP/SSR-Jaccard and AFLP-Jaccard, and AFLP/SSR-f^{M}_{AB} and AFLP-f^{M}_{AB} were extremely high ($r = 0.96$ and $r = 0.88$, respectively), but correlations for AFLP/SSR-Jaccard and SSR-Jaccard, and AFLP/SSR-f^{M}_{AB} and SSR-f^{M}_{AB} were low ($r = 0.64$ in both cases). This phenomenon was also obvious in the dendrograms constructed: groupings with AFLP were almost the same as those obtained by means of both markers (data not shown).

Table 4 also lists mean similarity/distance values obtained in the seven ways used to compare the clustering of the 85 inbred lines. The analyses made on the basis of Jaccard's coefficient presented different means, revealing different levels of diversity. It was observed that microsatellites (mean = 0.26) exposed more diversity than did AFLPs (mean = 0.51). The range of variation was also greater for SSRs (0.08–1.00) when contrasted with AFLPs (0.35–0.87). Even when being analyzed as a dominant marker (using a binary matrix and not allelic frequencies), the microsatellites were more effective in showing divergence as a result of higher degree of polymorphism and multiallelism. The molecular coefficient of coancestry proposed by Bernardo (1993) revealed very low mean similarity values, owing to the great number of negative similarities estimated with this coefficient. Since these negative estimates were all considered zero, this results in reducing the mean level of observed diversity. The matrices obtained with this coefficient with AFLPs or SSRs have very similar means, but the clusters displayed were not the same.

In Fig. 3, the frequencies of each similarity interval for the different coefficients are graphically displayed. The great majority of AFLP similarities are restricted to values between 0.4 and 0.6, whereas most of those based on SSRs, range from 0.1 to 0.4 (Jaccard). The differences in the values in which most similarities range demonstrate that the level of diversity displayed by SSRs is

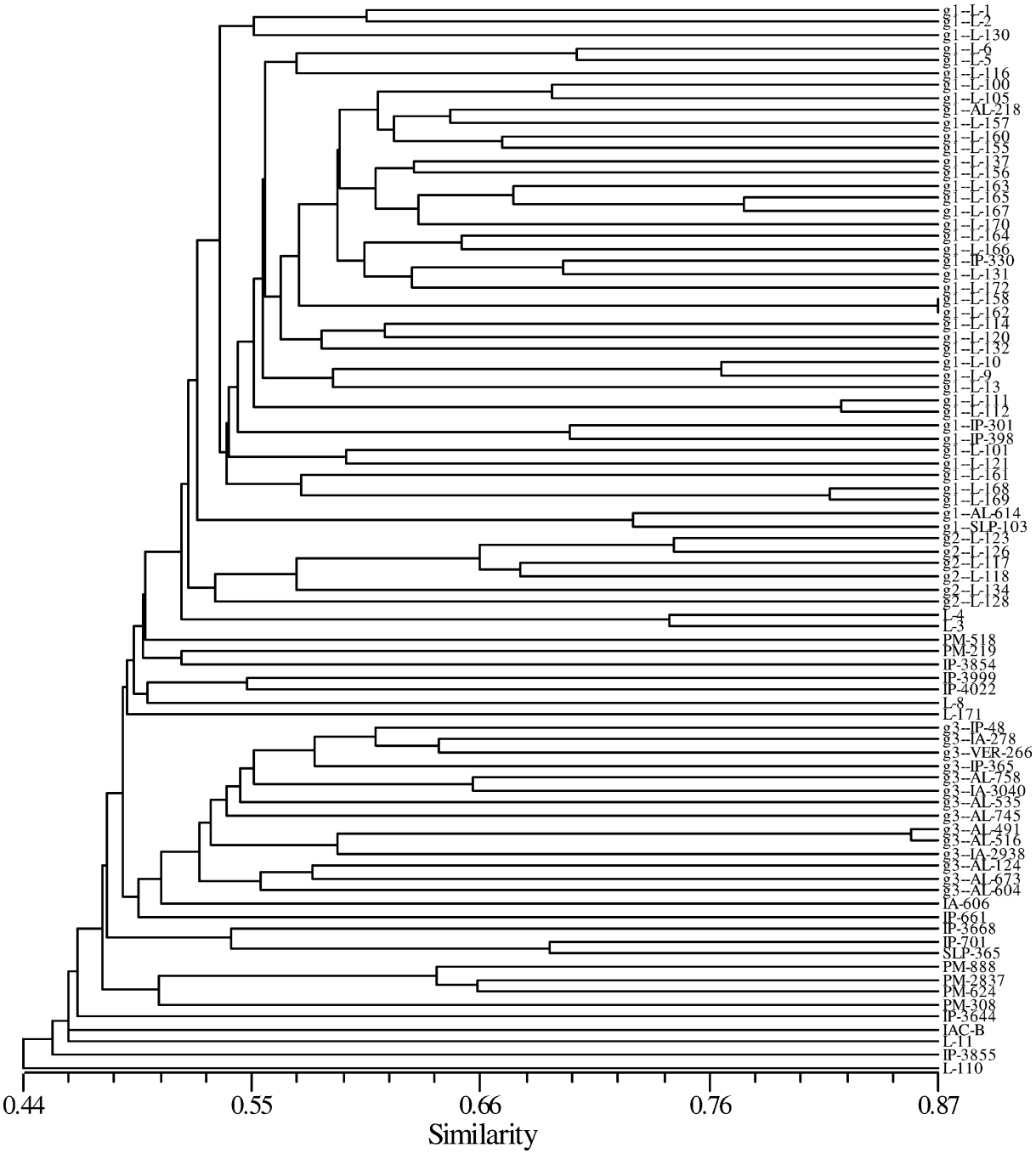

Fig. 1 UPGMA showing distribution of lines and groups (g1, g2, g3) obtained using Jaccard's coefficient and the AFLP data (cophenetic value = 0.78)

much higher than that of AFLPs, since many values observed with SSRs are very low, whereas with AFLPs, they approximate intermediate estimates. For similarities obtained with the molecular coefficient of coancestry, very similar results were reached for AFLPs and SSRs, all values ranging from 0 to 0.3. The graph exhibits the high genetic divergence of the tropical maize studied.

Interestingly, when the heterosis data from Paterniani et al. (2000) were compared to the arrangement of lines in the dendrograms, combinations of distant lines in the dendrogram showed high heterosis values, while combinations of closely related lines showed low heterosis values using the SSR-Jaccard data. The same pattern was not observed using the AFLP data.

Discussion

In our study, up to 98 polymorphic bands could be detected in a single AFLP gel with an overall polymorphic

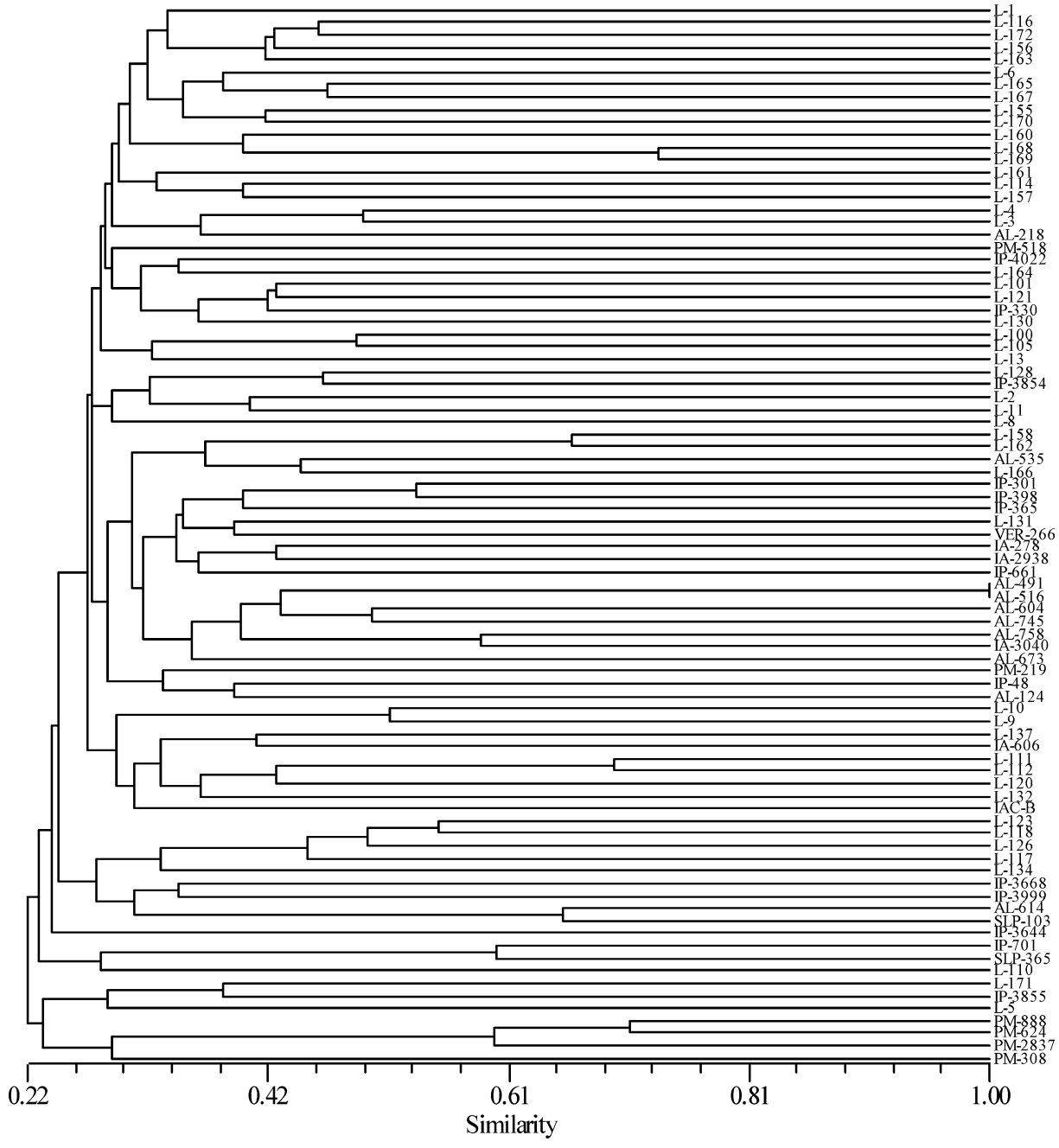

Fig. 2 UPGMA showing the distribution of lines obtained using Jaccard's coefficient and the SSR data (cophenetic value = 0.66)

Table 4 Correlations among the seven ways used to access diversity by means of Mantel's test and comparison of mean similarity/distance values according to the type of coefficient used to generate the similarity/distance matrix

	AFLP Jaccard	SSR Jaccard	AFLP/SSR Jaccard	SSR MRD	AFLP f_{AB}^{M}	SSR f_{AB}^{M}	Mean	Range
AFLP Jaccard	–						0.506	0.345–0.869
SSR Jaccard	0.43	–					0.264	0.085–1
AFLP/SSRJaccard	0.96	0.64	–				0.459	0.302–0.876
SSR MRD	−0.51	−0.95	−0.69	–			0.742	0–0.894
AFLP f_{AB}^{M}	0.88	0.46	0.86	−0.51	–		0.071	0–0.748
SSR f_{AB}^{M}	0.39	0.87	0.56	−0.83	0.48	–	0.058	0–0.663
AFLP/SSR f_{AB}^{M}	0.84	0.64	0.89	−0.67	0.96	0.68	0.072	0–0.789

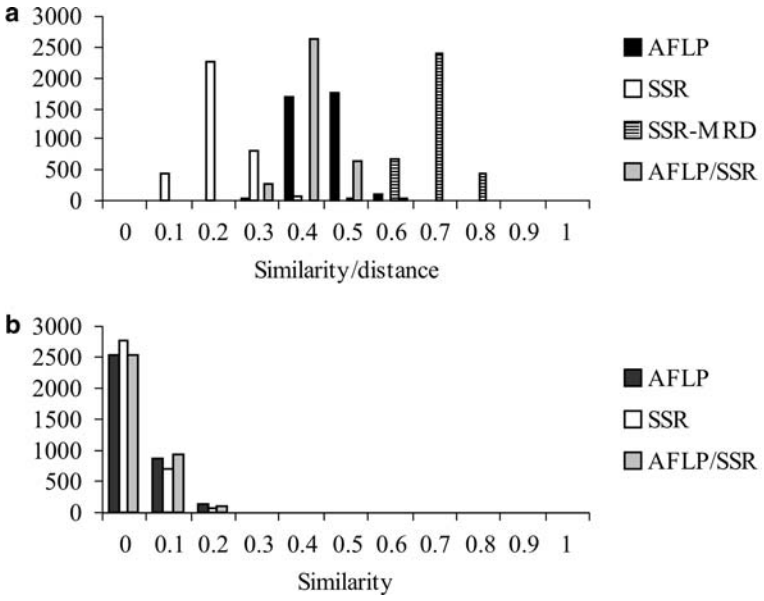

Fig. 3 Distribution of frequencies of the 3,570 pairwise values in the similarity matrices. **a** Markers analyzed using Jaccard's coefficient (and also a comparison with the distances obtained by modified Rogers' distance for SSR data). **b** Markers analyzed using the molecular coefficient of coancestry

rate of 89%. In contrast to the work of Russell et al. (1997) who found a low level of polymorphism for AFLPs in barley (48%), the tropical maize analyzed can be considered highly divergent. If each AFLP band is considered a locus, AFLPs are an efficient marker type to determine genomic variation.

In contrast to the efficient AFLP assays, SSR assays usually detect a single locus at a time, although it is possible to multiplex several SSRs in a single gel. However, the ability to detect multiple alleles at a single locus makes SSRs more appropriate for many diversity studies. In our study, the multi-allelic nature allowed 35 loci to be selected out of the 109 tests made with eight lines, and more loci would have been genotyped if the amplification conditions had been optimized; polymorphisms were not the striking problem. The fact that these loci have been developed for temperate maize may have caused some amplifications to fail when tropical maize DNA was used.

A second problem faced during the SSR genotyping procedures was the small differences between allele fragment sizes. While the polymorphism was detectable, genotyping all 85 samples was often difficult. This was somewhat alleviated by changing from agarose to polyacrylamide gels. Moreover, the SSRs with best amplification and genotyping conditions had fewer alleles, which would bias the mean number of alleles and PIC values. Many SSR loci with a high number of alleles were found in the initial tests, but most could not be used extensively due to amplification problems.

Diversity studies of temperate maize, which does not have a genetic base as diverse as tropical maize, revealed mean values for these parameters that are similar (Smith et al. 1997; Senior et al. 1998) or even higher (Pejic et al. 1998) than those obtained in this study. Comparing our results with the work of Warburton et al. (2002) with tropical lines (mean number of alleles = 4.9), a slightly higher number of alleles was observed in our study. It should be noted, however, that the work by Warburton and colleagues also used loci not selected on the basis of number of alleles. Consequently, comparison of molecular diversity estimates between tropical and temperate lines is confounded by the selection of markers with different inherent levels of variability in the different studies.

PIC values, which reveal not only the presence of many alleles but also their frequencies, may also have been lowered by the presence of rare alleles. The presence of alleles with very low frequency may be due to high rates of mutation of the SSRs (Henderson and Petes 1992) or to the introduction of exotic germplasm (Senior et al. 1998). Matsuoka et al. (2002) found that the percentage of rare alleles in tropical inbreds was higher than that found in temperate maize. These authors also indicated that the size variation in SSRs is not always due to the addition or deletion of motifs. Many studies presume that length variation is caused only by reduction or expansion of the SSR motif, but, in diversity studies, all types of variation are important in exposing the range of genomic variability.

Regarding discrepancies seen between marker types, Manifesto et al. (2001) also described a low correlation between AFLPs and SSRs for wheat ($r = 0.27$). However, they suggested that this difference was influenced by the few loci used in their analyses ($n = 10$). This would not explain the low correlations obtained in our study, because of the large number of markers used. Previous studies (Pejic et al. 1998; Barbosa et al. 2003) demonstrated moderately high correlation values between AFLPs and SSRs, but it should be noted that the materials surveyed did not have the large amount of variability found in this study, which probably influenced the results. Even in Barbosa et al. (2003), where tropical maize was analyzed, the lines that were studied

did not show as much diversity when compared to our results. Probably, disagreements in analyzing inbreds by one marker type or another reflect the level of diversity within the material and the different levels of variation uncovered by each type of marker.

The high mutation capacity of SSR markers also may have led to the differences observed between the two techniques. According to Li et al. (2001), the inconsistency among different molecular marker data, especially for inbred lines, results from the fact that the markers evaluate different components of DNA variation, which may evolve in different ways. AFLPs, which do not mutate as fast as SSRs, show higher similarities values and this may have resulted in more consistent groups, as observed for the CIMMYT-derived lines. SSRs maximize the differences among inbreds because they can disclose the diversity in another way, showing differences at a more micro-level. This allows them to be more usefull in analyzing genetic variability, because they have the ability to detect small differences. Due to slippage events that lead to multiple alleles, each line may have been assigned to different subcategories, preventing the grouping of lines. Consequently, this higher level of polymorphism may not be appropriate for establishing pedigree relationships among cultivars. Indeed, the SSR dendrograms did not recover the relationships among lines, corroborating what Warburton et al. (2002) reported for other tropical inbreds. Markers that do not vary much genetically are more suitable for pedigree analyses, as shown by Lima et al. (2002) in sugarcane studies with AFLPs, and as confirmed by our results. In contrast, markers that are able to display greater levels of polymorphism are more efficient to study the real genetic variability among genotypes and also to distinguish them for breeding purposes.

Clusters made with combined AFLP and SSR data were very similar to those made with only AFLP data, but this could have been due to the greater number of AFLP than SSR bands. To test this possibility, four different subsets of the AFLP data, each with approximately the same number of bands as the complete SSR data set, were used to construct the similarity matrices, in the same way as was done for the complete data set. The mean correlation between Jaccard's coefficients based on AFLP subsets and SSR data was 0.36, for AFLP/SSR and AFLP was 0.90, and for AFLP/SSR and SSR was 0.68. These results suggest that the greater number of AFLP bands was not responsible for the disagreement between AFLP/SSR and SSR matrices. Nevertheless, the use of Jaccard's coefficient for different types of variation may cause part of this inconsistency. To consider a SSR band as a locus, and not an allele at a locus, decreases the value of SSRs as codominant markers. Analyzing it separately does not seem to imply loss of information, but when together with AFLPs, the same weight of an AFLP band is attributed to a SSR allele.

Despite the importance of the marker type in analyzing variation, the choice of statistical coefficients is a rather fundamental step in studying genetic diversity. MRD, one of the most appropriate distances when using codominant markers, was not used for AFLP data; nonetheless, the genetic variability seen with MRD was very similar to those obtained by Jaccard's coefficient and f_{AB}^{M} for SSR, which could be compared to AFLP by means of these last two coefficients. f_{AB}^{M} is a coefficient that reveals relationships among genotypes and does not focus on the diversity among them. According to Bernardo (1993), similarity values obtained by molecular markers are overestimates of true relationships among genotypes, the amount of bias depending on how divergent the material being studied is. The greater the distance among inbreds, the greater the bias created. ''Molecular similarities'' may provide wrong relationships among inbreds because of their inability to discriminate lines that are only *ais* from the ones that are *ibd*. The molecular coefficient of coancestry aims to circumvent this problem, but it was highly correlated with Jaccard's coefficient, suggesting that its use did not improve the organization of the lines. The dendrograms made with Jaccard's coefficient and with the molecular coefficient of coancestry grouped the lines in a very similar manner.

In conclusion, it was observed that tropical maize is a rich source of genetic variability, providing the necessary raw material for breeding programs. When very diverse material is under analysis, appropriate choice of molecular marker should be made, and the SSRs seem to be the most useful marker for this purpose. Considering their multi-allelic nature and their ability to uncover great levels of polymorphism, they are a good choice when the objective is assessing genetic diversity. All types of markers are able to reveal diversity, but as noted in this work, some do not access high levels of genomic variation, which may hide a part of divergence exhibited by the genotypes. Using both data together does not seem to be a good choice to analyze the genetic diversity. The use of Jaccard's coefficient, MRD and the molecular coefficient of coancestry showed very similar results in grouping genotypes when data from AFLPs and SSRs were used separately, but due to its statistical properties and simplicity, the Jaccard coefficient seems to be a good choice. The recovery of relationships among lines was optimal using AFLP data, and the prediction of better combinations of lines was achieved in the SSR dendrograms, but additional information is necessary. Well-defined heterotic groups were not observed, indicating that the tropical maize is not organized as is temperate material, but it was seen that the use of molecular markers can help to separate lines for breeding purposes and make the rational exploitation of this germplasm possible.

Acknowledgments The authors wish to thank CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) (00/11517-4) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for the financial support during the development of this research. A.P.S. and A.A.F.G. received a re-

search fellowship from CNPq. The authors would like to express their gratitude to Dr. James B. Holland and to Dr. David Hoisington who carefully revised the manuscript and kindly provided valuable comments on it.

References

Abdalla AM, Reddy OUK, El-Zik KM, Pepper AE (2001) Genetic diversity and relationships of diploid and tetraploid cottons revealed using AFLP. Theor Appl Genet 102:222–229

Barbosa AMM, Geraldi IO, Benchimol LL, Garcia AAF, Souza CL Jr, Souza AP (2003) Relationship of intra- and interpopulational tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. Euphytica 130:87–99

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

Boppenmaier J, Melchinger AE, Seitz G, Geiger HH, Herrmann RG (1993) Genetic diversity for RFLPs in European maize inbreds. III. Performance of crosses within versus between heterotic groups for grain traits. Plant Breed 111:217–226

Chin ECL, Senior ML, Shu H, Smith JSC (1996) Maize simple repetitive DNA sequences: abundance and allele variation. Genome 39:866–873

Creste S, Tulmann Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. Plant Mol Biol Rep 19:299–306

Doldi M-L, Vollmann J, Lelley T (1997) Genetic diversity in soybean as determined by RAPD and microsatellite analysis. Plant Breed 116:331–335

Enoki H, Sato H, Koinuma K (2002) SSR analysis of genetic diversity among maize inbred lines adapted to cold regions of Japan. Theor Appl Genet 104:1270–1277

Garcia AAF, Benchimol LL, Barbosa AAM, Geraldi IO, Souza Jr CL, de Souza AP (2004) Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. Genet Mol Biol 27:579–588

Gethi JG, Labate JA, Lamkey KR, Smith ME, Kresovich S (2002) SSR variation in important US maize inbred lines. Crop Sci 42:951–957

Ghebru B, Schmidt RJ, Bennetzen JL (2002) Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theor Appl Genet 105:229–236

Hahn V, Blankenhorn K, Schwall M, Melchinger AE (1995) Relationships among early European maize inbreds. III. Genetic diversity revealed with RAPD markers and comparison with RFLP and pedigree data. Maydica 40:299–310

Halldén C, Nilsson N-O, Rading IM, Säll T (1994) Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. Theor Appl Genet 88:123–128

Henderson ST, Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol Cell Biol 12:2749–2757

Hoisington D, Khairallah M, González-de-Léon D (1994) Laboratory protocols: CIMMYT applied molecular genetics laboratory, 2nd edn., CIMMYT, Mexico

Jaccard P (1908) Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44:223–270

Lanza LLB, de Souza CL Jr, Ottoboni LMM, Vieira MLC, de Souza AP (1997) Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor Appl Genet 94:1023–1030

Li C, Fatokun CA, Ubi B, Singh BB, Scoles GJ (2001) Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. Crop Sci 41:189–197

Lima MLA, Garcia AAF, Oliveira KM, Matsuoka S, Arizono H, de Souza CL Jr, de Souza AP (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

Liu S, Cantrell RG, McCarty JC Jr, Stewart JMcD (2000) Simple sequence repeat-based assessment of genetic diversity in cotton race stock accessions. Crop Sci 40:1459–1469

Lu H, Bernardo R (2001) Molecular marker diversity among current and historical maize inbreds. Theor Appl Genet 103:613–617

Manifesto MM, Schlatter AR, Hopp HE, Suaréz EY, Dubcovsky J (2001) Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci 41:682–690

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Matsuoka Y, Mitchell SE, Kresovich S, Goodman M, Doebley J (2002) Microsatellites in *Zea* – variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet 104:436–450

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

Meyer AS, Garcia AAF, Souza AP, Souza CL Jr (2004) Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (Zea mays L). Genet Mol Biol 27:83–91

Miller MP (1997) TFPGA—Tools for population genetic analyses, vol 1.3. Northern Arizona University

Nei (1987) Molecular evolutionary genetics. Columbia University Press, New York

Oliveira KM, Laborda PR, Garcia AAF, Zagatto-Paterniani MEA, Souza AP (2004) Evaluating genetic relationships between tropical maize inbred lines by means of AFLP profiling. Hereditas 140:24–33

Paterniani MEAGZ, Sawazaki E, Dudienas C, Duarte AP, Gallo PB (2000) Diallel crossing among maize lines with emphasis on resistance to foliar diseases. Genet Mol Biol 23:381–385

Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G, Motto M (1998) Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs and AFLPs. Theor Appl Genet 97:1248–1255

Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using microsatellite markers. Theor Appl Genet 91:1001–1007

Priolli RHG, Mendes-Junior CT, Arantes NE, Contel EPB (2002) Characterization of Brazilian soybean cultivars using microsatellite markers. Genet Mol Biol 25:185–193

Reif JC, Melchinger AE, Xia XM, Warburton ML, Hoisington DA, Vasal SK, Srinivasan G, Bohn M, Frisch M (2003a) Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. Crop Sci 43:1275–1283

Reif JC, Melchinger AE, Xia XM, Warburton ML, Hoisington DA, Vasal SK, Beck D, Bohn M, Frisch M (2003b) Use of SSRs for establishing heterotic groups in subtropical maize. Theor Appl Genet 107:947–957

Reif JC, Xia XM, Melchinger AE,Warburton ML, Hoisington DA, Beck D, Bohn M, Frisch M (2004) Genetic diversity determined within and among CIMMYT maize populations of tropical, subtropical, and temperate germplasm by SSR markers. Crop Sci 44:326–334

Rohlf FJ (1997) NTSYSpc - Numerical taxonomy and multivariate analysis system, v. 2.1. Exerter Publications, New York

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R (1997) Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet 95:714–722

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York, pp B23

Senior ML, Murphy JP, Goodman MM, Stuber CW (1998) Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. Crop Sci 38:1088–1098

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays L.*): comparisons with data from RFLPs and pedigree. Theor Appl Genet 95:163–173

Smith JSC, Kresovich S, Hopkins MS, Mitchell SE, Dean RE, Woodman WL, Lee M, Porter K (2000) Genetic diversity among elite sorghum inbred lines assessed with simple sequence repeats. Crop Sci 40:226–232

Tivang JG, Nienhuis J, Smith OS (1994) Estimation of sampling variance of molecular marker data using the bootstrap procedure. Theor Appl Genet 89:259–264

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Fritjers A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23:4407–4414

Warburton ML, Xianchun X, Crossa J, Franco J, Melchinger AE, Frisch M, Bohn M, Hoisington D (2002) Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. Crop Sci 42:1832–1840

Wright S (1978) Evolution and the genetics of populations, vol. 4. Variability within and among natural populations. University of Chicago Press, Chicago