Population structure and genetic diversity of maize inbreds derived from tropical hybrids

**E.C.M. Lanes[1], J.M.S. Viana[2], G.P. Paes[2], M.F.B. Paula[1], C. Maia[2],
E.T. Caixeta[3] and G.V. Miranda[1]**

[1]Departamento de Fitotecnia, Universidade Federal de Viçosa,
Viçosa, MG, Brasil
[2]Departamento de Biologia Geral, Universidade Federal de Viçosa,
Viçosa, MG, Brasil
[3]Embrapa Café/Instituto de Biotecnologia Aplicada à Agropecuária,
Universidade Federal de Viçosa, Viçosa, MG, Brasil

Corresponding author: E.C.M. Lanes
E-mail: edercml@gmail.com

ABSTRACT. The objectives of this study were to identify the population structure and to assess the genetic diversity of maize inbreds. We genotyped 81 microsatellite loci of 90 maize inbreds that were derived from tropical hybrids and populations. The population structure analysis was based on a Bayesian approach. Each subpopulation was characterized for the effective number of alleles, gene diversity, and number of private alleles. We also performed an analysis of molecular variance and computed a measure of population differentiation (F_{ST}). The genetic distances were computed from the similarity index of Lynch and the dissimilarity measures proposed by Smouse and Peakall. The cluster analyses were based on the unweighted pair-group method using arithmetic averages and Tocher method. The clustering efficiency was assessed by the error rate of the discriminant analysis. We also performed a principal coordinates analysis. The population structure analysis revealed three tropical heterotic pools, which have

been used by worldwide and Brazilian maize seed companies. The degree of genetic differentiation and of intra- and inter-population genetic diversity for these tropical heterotic pools are comparable to that observed for temperate and subtropical heterotic pools. The higher allelic frequency variation within each tropical heterotic pool and the high genetic diversity between the inbreds were evidence of heterotic groups within the main tropical heterotic pools.

INTRODUCTION

Genetic diversity analysis using molecular markers has been used in plant breeding and in studies involving germplasm conservation (Belaj et al., 2012), association mapping (Li et al., 2011), population structure (Van Inghelandt et al., 2010), genetic relationships among populations (Bracco et al., 2009), association between genetic distance and heterosis or specific combining ability, geographic adaptation and pedigree information (Legesse et al., 2007, 2008), linkage disequilibrium (Stich et al., 2006), and heterotic group identification (Xia et al., 2004, 2005). Most studies with maize (*Zea mays* L. subsp *mays*) have used inbred lines. In the few studies with maize populations, because of the difficulty and cost in performing laboratory research, the sample sizes relative to markers and individuals were reduced (Santacruz-Varela et al., 2004; Bracco et al., 2009).

Maize germplasm naturally forms two major germplasm pools, temperate and tropical (including subtropical), which are based on the environmental and day length characteristics of the planting areas in the world. Molecular markers revealed differences among the genetic diversity these two groups, in which that of the maize tropical germplasm pool is higher than that in the temperate pool because the tropical germplasm pool was found to contain many rare alleles using simple sequence repeat (SSR) markers (Liu et al., 2003) and more single nucleotide polymorphism (SNP) alleles and haplotype diversity (Yan et al., 2009). This parameter is an important resource to broaden the genetic base of maize breeding populations. Many useful allelic forms for improving temperate maize may be hidden in the tropical germplasm (Yan et al., 2009). According to Laborda et al. (2005), the tropical maize is not organized like the temperate material, but the use of molecular markers can help to separate lines for breeding purposes and to make the rational exploitation of this germplasm possible.

Several previous studies demonstrated the possibilities of molecular markers because they offer a stable and reliable alternative for genetic identification and characterization of germplasm collections. Van Inghelandt et al. (2010) performed a large genetic diversity analysis in maize using 359 SSRs, 8244 SNPs, and 1537 inbred lines. The results from the analyses of SNPs and SSRs were consistent and revealed germplasm organization in four heterotic pools. The average percentage of concordance between the clustering and the heterotic groups was approximately 90%, regardless of the marker. Comparing SNP and SSR markers for genetic diversity analysis, Yang et al. (2011b) considered SSR markers to be more informative. Bracco et al. (2009) assessed the genetic diversity of six popcorn {[*Zea mays* L. ssp *mays* [syn. *Zea mays* L. ssp *everta* (Sturtev.) Zhuk.]} landraces by genotyping 131 individuals at

10 SSR loci. The analysis of molecular variance revealed that 34, 39, and 27% of the total variance were due to differences between populations, individuals within populations, and regions of adaptation, respectively. They observed Hardy-Weinberg equilibrium and no linkage disequilibrium. Furthermore, the Mantel's test revealed no relationship between genetic and geographic distances, and only one of the populations that were analyzed showed evidence of subpopulation structure. Legesse et al. (2008) analyzed nine amplified fragment length polymorphism (AFLP) primer pairs among 26 maize inbred lines and six testers to study the relationship between genetic distance and hybrid performance and heterosis. Cluster analysis separated the inbred lines from the testers, and the clustering was concordant with the pedigree. For grain yield, plant height, and days to silking, correlations between genetic distance and hybrid performance and heterosis were zero or negative. Furthermore, positive correlations identified through these analyses were considered unsatisfactory for predictive purposes. Legesse et al. (2007) genotyped 56 highland and mid-altitude maize inbred lines for 27 SSR loci. Most of the inbred lines with different adaptations were grouped in different clusters. Furthermore, the clustering was consistent with pedigree information. Stich et al. (2006) genotyped 72 elite European maize inbred lines at 452 AFLP and 93 SSR loci. The genetic diversity analysis revealed two main groups (dent and flint) and three subgroups for each group. Xia et al. (2004, 2005) evaluated the genetic diversity among maize inbred lines from different regions of adaptation and their genetic relationship to elite inbred lines from the USA, Canada, and Europe with the overall goal of identifying heterotic groups. The 155 inbred lines from tropical lowlands, 126 tropical and subtropical inbred lines from mid-altitude and highland areas, and 21 inbred lines from temperate climates were genotyped for 79 SSR markers. The analysis did not reveal clustering that was consistent with adaptation patterns.

Germplasm, which has been utilized by worldwide maize seed corporations, is recognized as improved, adapted, and divergent. Germplasm should also be the source of new inbred lines that are developed by small seed companies and are derived from commercial hybrids. However, at least for the tropical maize germplasm pool, there is no knowledge about the genetic diversity between the germplasm employed by these corporations. Our objectives were to identify the population structure and to assess the genetic diversity using maize inbreds derived from tropical hybrids.

MATERIAL AND METHODS

Inbred lines

We genotyped 90 inbred lines. Among these inbreds, 76 were derived from 27 non-transgenic tropical hybrids of the worldwide seed companies Monsanto (6; 2 from Dekalb, 3 from Agroceres Seeds, and 1 from Cargill), Dow AgroSciences (2; 1 from Agromen Seeds), Pioneer (5), Syngenta (8; 3 from Astra-Zeneca), Embrapa (Brazilian Agricultural Research Corporation) (1), and three Brazilian seed companies-Semeali Seeds, Nidera Seeds, and Balu Seeds (5). The other 14 inbreds were derived from four Brazilian varieties (from Embrapa and Cati, Integral Technical Assistance Coordination), one of which is yet commercially available. Among the hybrids, only 9 are commercial; 11, 7, and 9 are single, double, and triple crosses, respectively; 15, 9, and 3 are flint, semi-flint, and semi-dent types, respectively; and 3, 20, and 4 require more than 890 growing degree units (GDUs), 830-890 GDUs, and less than 830

GDUs. The predominant grain color of the hybrids is orange. The inbreds were developed at the Universidade Federal de Viçosa, Minas Gerais state, Brazil.

Genotyping

The lines were genotyped for 81 SSR loci chosen from the MaizeGDB database. Genomic DNA was extracted from young leaf tissue using hexadecyltrimethylammonium bromide (CTAB) following a modified protocol previously described by Doyle and Doyle (1990). Polymerase chain reactions (PCRs) were performed in a thermal cycler PTC-200 (MJ Research, Watertown, MA, USA) using a touchdown profile, and the products were analyzed by gel electrophoresis on 0.8% (w/v) agarose gels. Each PCR mixture contained 1X buffer, 1 mM $MgCl_2$, 0.5 U Taq polymerase, 0.1 μM each primer, 0.1 μM each deoxynucleotide (dATP, dTTP, dGTP, and dCTP), and 30 ng template DNA diluted with water to a total reaction volume of 20 μL. Electrophoresis was conducted in 1X TBE buffer, pH 8.0, at 50 W for 2-3 h in a Sequi-Gen GT (Bio-Rad) apparatus. Bands were visualized by silver nitrate staining.

Statistical analyses

The Structure version 2.3.3 software (Pritchard et al., 2000; Falush et al., 2003) was used for the Bayesian clustering. Five runs were performed by setting the number of clusters (K) from 2 to 10. Each run consisted of a burn-in of 10,000 followed by 50,000 Markov chain Monte Carlo replications, assuming the admixture model and correlated allelic frequencies. No prior information was used. The choice of the most likely number of clusters (K) was carried out by calculating the statistics ΔK, which is based on the rate of change in the log probability of data between successive K values, as described by Evanno et al. (2005). The populations identified by Structure were characterized for the effective number of alleles per locus, gene diversity, and number of private alleles. An analysis of molecular variance (AMOVA) was also performed. We computed a measure of population differentiation (F_{ST}).

We used the similarity index proposed by Lynch (1990) and the unweighted and weighted measures of dissimilarity proposed by Smouse and Peakall (1999). The Lynch's index corresponds to the fraction of shared bands by two individuals (x and y), given by

$$S_{xy} = \frac{2n_{xy}}{n_x + n_y} \qquad \text{(Equation 1)}$$

where n_{xy} is the number of common fragments, and n_x and n_y are the number of distinct fragments of x and y. Regarding a locus, the Smouse and Peakall unweighted measure of dissimilarity is the square of the Euclidean distance, assuming that the distance between homozygotes for distinct alleles is 2, the distance between a homozygote and heterozygote that do not share a common allele is $\sqrt{3}$, the distance between heterozygotes that do not share a common allele is $\sqrt{2}$, the distance between a homozygote and heterozygote that share a common allele is 1, and the distance between heterozygotes that share a common allele is 1. The unweighted measure of dissimilarity is

$$d_{ij}^2 = \frac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{K_l}\left(y_{ikl} - y_{jkl}\right)^2 \qquad \text{(Equation 2)}$$

where L is the locus number, K_l is the number of alleles at locus l, and y_{ikl} is the element of the genetic vector relative to individual i and gene k at locus l (2 if homozygous for the gene, 1 if heterozygous for the gene, and 0 relative to the other alleles). Weighting by the allelic frequencies was given by

$$d_{ij}^2 = \sum_{l=1}^{L}\sum_{k=1}^{K_l} W_{kk}\left(y_{ikl} - y_{jkl}\right)^2 \qquad \text{(Equation 3)}$$

where $W_{kk} = (2K_l p_{kl})^{-1}$, and p_{kl} corresponds to the frequency of gene k at locus l.

The genetic distances were computed as 1 - S_{xy} (genetic similarity) or d_{ij}/L. For the cluster analyses, the unweighted pair-group method using arithmetic averages (UPGMA) (Johnson and Wichern, 1988) and the Tocher optimization procedure (Rao, 1952) were used. The main criterion for grouping by the Tocher method is that the average genetic distances within groups must be smaller than the average genetic distances between groups. The number of groups with UPGMA clustering was defined based on the pseudo-t^2 statistic (Calinski and Harabasz, 1974). To assess the grouping efficiency, the error rate of the discriminant analysis based on the k nearest neighbor method was used (Johnson and Wichern, 1988). This non-parametric method provides that each observation is allocated to the class that contains the highest proportion of k nearest neighbors. Isolated individuals were not included in the analyses. The a priori classification criterion was based on the grouping from UPGMA and Tocher method. The goodness-of-fit of the grouping and the similarity of the dendrograms obtained by UPGMA were assessed by the cophenetic correlation coefficient (Sokal and Rohlf, 1962) and by the Colless consensus index (Colless, 1980), respectively. The cophenetic correlation coefficient is the correlation between the distance matrix and the cophenetic matrix. The consensus index is the proportion of the possible subsets for t objects that are in the consensus for the two classifications. To test correlations between the distance matrices, we used Mantel's test. We also performed a principal coordinates analysis (PCoA; Orloci, 1978).

For these statistical analyses, NTSYS-pc 2.21 (Rohlf, 2009), SAS 9.0 (SAS Institute, 2002), GenAlEx 6.4 (Peakall and Smouse, 2006), and GENES (Cruz, 2008) software was used. The weighted measure of dissimilarity of Smouse and Peakall and the corresponding genetic distances were computed by a program developed in REALbasic 5.5 (REAL Software, 2004).

RESULTS

The 81 polymorphic markers generated 471 alleles (Table 1). The average number of alleles per marker was 5.8, and the number of alleles per marker ranged from two (umc2377, umc1057, umc1288, umc1847, umc2036, umc2098, umc2075, and umc1505) to 17 (umc1035). Based on the method proposed by Evanno et al. (2005), the analysis of

population structure revealed four subpopulations (Figure 1). In three of these populations, there is evidence of individuals with admixture ancestry. These individuals suggest crosses mainly between populations 1 and 2, and 1 and 4. Among the inbred lines, 36, 20, 4, and 11 are clearly derived from populations 1, 2, 3, and 4, respectively. The remaining 19 inbreds were categorized as having admixed ancestry, including nine with admixture between populations 1 and 2, eight with admixture between populations 1 and 4, and two that share genomes from populations 2 and 4. The estimates of population parameters indicate that gene pools 1, 2, and 4 have approximately three alleles per locus with a frequency that is distinctly different from zero (Table 2). Gene pool 1 showed the greatest gene diversity and number of private alleles, followed by gene pools 2 and 4. The F_{ST} for gene pools 1 and 2, and 1 and 4 was 0.29. The value relative to gene pools 2 and 4 was 0.04. The pattern of distribution of diversity indicated that 85.5% of the allelic frequency variation was accounted for within-population.

Table 1. Characterization of the simple sequence repeat (SSR) primers and number of alleles.

| Marker	Bin	Repetition	Alleles	Marker	Bin	Repetition	Alleles
umc2071	1.01	$(ATGT)_5$	6	umc1894	5.02	$(GA)_8$	5
umc2124	1.03	$(TA)_6$	9	umc1852	5.03	$(TC)_8$	7
bnlg2295	1.04	$(AG)_{30}$	6	phi085	5.06	AACGC	6
umc1243	1.05	$(CAT)_4$	3	umc2308	5.08	$(CGGCG)_4$	5
umc1281	1.05	$(CAA)_6$	6	umc1153	5.09	$(TCA)_4$	7
umc1035	1.06	$(CT)_{19}$	17	phi126	6.00	AG	12
bnlg615	1.07	-	6	phi077	6.01	AG	10
umc1431	1.09	$(GCA)_5$	3	umc2314	6.01	$(GT)_6$	5
umc1726	1.10	$(TG)_6$	3	umc1178	6.02	$(GGC)_6$	6
umc2245	2.01	$(CAA)_7$	4	umc1656	6.02	$(CGGT)_7$	6
bnlg1064	2.03	$(AG)_{16}$	6	umc1014	6.04	$(GA)_{12}$	7
bnlg1175	2.04	$(AG)_{38}$	10	umc2319	6.04	$(GAGGAG)_5$	5
umc2254	2.05	$(AGAG)_4$	6	bnlg1136	6.07	$AG_{(14)}$	7
umc2402	2.07	$(CAT)_4$	6	umc1066	7.01	$(GCCAGA)_5$	3
phi127	2.08	AGAC	5	umc1016	7.02	$(CT)_{25}$	8
umc2184	2.09	$(GCG)_5$	5	phi034	7.02	CCT	6
phi101049	2.10	AGAT	12	umc2098	7.03	$(CAG)_5$	2
umc2214	2.10	$(CTT)_4$	5	umc1301	7.03	$(GCA)_5$	3
umc1931	3.01	$(AC)_8$	6	phi116	7.06	ACTG/ACG***	6
umc2377	3.01	$(GAC)_4$	2	phi119	8.02	AG	5
umc2257	3.01	$(CTCCT)_4$	5	umc1904	8.03	$(TAAGC)_5$	5
umc1057	3.02	$(CGG)_6$	2	umc2075	8.04	$(AGCCAG)_4$	2
umc2258	3.03	$(GTG)_5$	5	bnlg240	8.06	-	7
umc2259	3.03	$(CCG)_6$	4	phi033	9.01	AAG	4
umc2266	3.06	$(AAGGAG)_4$	4	phi065	9.03	CACTT	4
umc1690	3.07	$(GCA)_4$	3	umc1271	9.03	$(CGG)_4$	4
umc1136	3.09	$(GCA)_5$	7	bnlg1714	9.04	AG(25)	5
bnlg1754	3.09	AG(20)	11	umc2343	9.05	$(TGTGTGTG)_4$	10
umc1232	4.00	$(ACAG)_4$	3	bnlg1191	9.07	$AG_{(39)}$	11
umc1017	4.01	$(CT)9(CA)_8$	7	bnlg1129	9.07	$AG_{(12)}$	5
umc1288	4.02	$(TCCA)_4$	2	umc1505	9.07	$(AAAAC)_4$	2
phi021	4.03	AG	3	umc1152	10.01	$(ATAG)_6$	7
umc1031	4.05	$(CT)_6AT(CT)_9$	16	phi052	10.02	AAG	3
umc1847	4.07	$(CGC)_6$	2	mmc0501	10.02	$(GA)_{36}$	8
umc1899	4.08	$(AG)_8$	4	umc1336	10.03	$(ACCAG)_4$	4
umc1612	4.08	$(GAGT)_6$	4	bnlg1074	10.05	$AG_{(14)}$	6
umc1050	4.11	$(AAT)_6$	7	umc1078	10.05	$(GT)_{13}$	9
bnlg1006	5.00	$AG_{(20)}$	11	bnlg1028	10.06	$AG_{(12)}$	6
umc2036	5.01	$(GTC)_4$	2	umc1045	10.06	GGA_3GAG_5	5
umc1761	5.02	$(GCA)_5$	4	umc1249	10.07	$(TG)_6$	5
umc1587	5.02	$(AT)_8$	6				
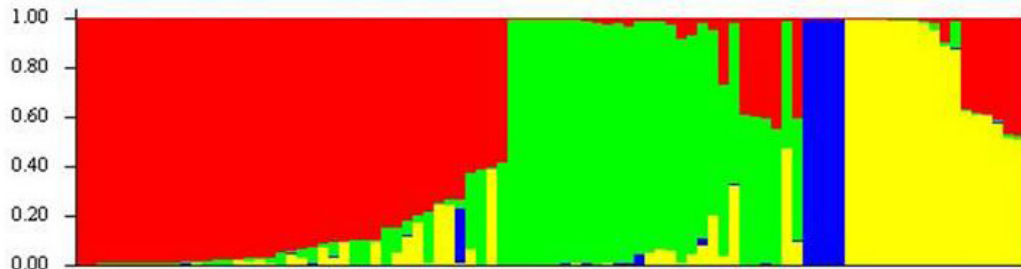
Figure 1. Population structure based on tropical maize inbred lines derived from hybrids and populations.

Table 2. Estimates of population parameters.

Population	Inbred lines	No. of alleles/loci	Effective number of alleles/loci	Gene diversity	Private alleles
1	41	4.9	3.1	0.59	65
2	28	4.4	2.9	0.56	33
3	4	1.2	1.2	0.10	2
4	17	3.4	2.5	0.51	10

The Lynch's index provided the largest maximum and average genetic distances (0.82 and 0.62), and the lower minimum distance (0.08). The Smouse and Peakall's unweighted and weighted dissimilarity measures provided similar minimum, maximum, and average distances (0.23 and 0.20, 0.52 and 0.63, and 0.46 and 0.46, respectively). The three indices were consistent for identifying the 10% most similar inbred lines, but the Smouse and Peakall weighted dissimilarity measure exclusively identified the 10% most divergent pairs. Regardless of the clustering method, using the Lynch's index provided the lowest error rate and the greatest number of groups with at least two inbred lines (Table 3). The clustering based on the weighted measure of dissimilarity of Smouse and Peakall revealed only one group with at least 81 inbreds. Although the Mantel's correlation values demonstrated high proportionality between the genetic distances that were obtained from the Lynch and the unweighted Smouse and Peakall indexes (0.93), distinct UPGMA clustering was observed based on the magnitude of the consensus index (0.41). The cophenetic correlation values showed that the best fit between the clustering and the genetic distances matrix was obtained using the Smouse and Peakall weighted dissimilarity measure (0.83).

Table 3. Numbers of groups and non-grouped inbred lines and the associated error rate obtained from the unweighted pair-group method with arithmetic means (UPGMA) and Tocher method, from the cluster analyses based on the Lynch's index and the Smouse and Peakall's unweighted (SPu) and weighted (SPw) dissimilarity measures.

Method		Lynch	SPw	SPu
Tocher	Groups	18	1	11
	Inbreds	8	9	11
	Error rate	14.6	-	21.5
UPGMA	Groups	18	1	2
	Inbreds	5	1	0
	Error rate	8.2	-	11.1

DISCUSSION

An accurate assessment of genetic diversity can be invaluable for population genetics, germplasm conservation, and crop breeding. Relevant applications include the identification of diverse parental combinations to create segregating progeny with maximum genetic variability, introgression of desirable genes from diverse germplasm into the available genetic base, identification of adaptation patterns by linking genetic variability to environmental data, understanding the genetic relationships between major groups of inbred lines used in crosses, assigning inbred lines to specific heterotic groups, and plant varietal protection (Mohammadi and Prasanna, 2003; Reif et al., 2005; Jombart et al., 2009; Laurentin, 2009). The observed genetic diversity is affected by the data (morphological, biochemical, and molecular marker data), the sample size of markers and individuals genotyped, the criterion for defining the optimal number of groups, and the amount of data lost (Mohammadi and Prasanna, 2003).

The analysis of population structure revealed three tropical heterotic pools (1, 2, and 4), which have been used by worldwide and Brazilian maize seed companies to develop inbred lines and hybrids mainly for Brazil. Because it is improbable that these companies have actual populations under recurrent selection, these heterotic pools and their admixture should be represented by the elite inbred lines and hybrids of each company. Heterotic pool 3 is related to a white grains population from the International Maize and Wheat Improvement Center (CIMMYT) and a population derived from crosses between hybrids. However, this heterotic pool was not adequately characterized because the genetic parameter estimates were obtained from only two inbreds from each of these two populations. This is why the effective number of alleles is one.

The principal coordinates analysis clearly separated the four gene pools that were identified by Structure (Figure 2). Yang et al. (2011b) and Van Inghelandt et al. (2010) performed population structure analysis based on SSR and/or SNP markers. Yang et al. (2011b) analyzed high-oil maize inbred lines derived from North American and Chinese populations and obtained only two gene pools, including one with admixture. Van Inghelandt et al. (2010) observed four heterotic pools (Flint, Lancaster, Stiff Stalk, and Iodent). All founder and elite European and North American inbred lines were obtained by the French company Limagrain. Xie et al. (2008) and Wang et al. (2008) analyzed Chinese inbred lines that were genotyped for SSR loci. The population structure analyses revealed six and four heterotic pools, respectively. The inbreds analyzed by Xie et al. (2008) represent those already employed and currently in use in public and private hybrid development programs. The heterotic pools identified by Xie et al. (2008) included four North American and two Chinese germplasms. The heterotic pools obtained by Wang et al. (2008) were consistent with the pedigree information and the probable origin of the Chinese maize germplasm (Lancaster, Reid, TSPT, and P). Based on the analysis of SSR markers and five populations from Central Europe that were widely used in breeding in the pre-hybrid age, Reif et al. (2005) concluded that each population corresponds to a heterotic pool. From the SSR analysis of public inbred lines that are representative of those used in maize breeding programs in North America, Europe, South Africa, and Thailand, and others from the CIMMYT and International Institute of Tropical Agriculture, Liu et al. (2003) observed three main heterotic pools. One of them included only tropical and subtropical inbreds, one corresponded to the non-Stiff Stalk gene pool, and the last was characterized as mixed because it included sweet, popcorn, and Stiff Stalk inbreds. The population structure in tropical/subtropical, Stiff Stalk, non-Stiff Stalk, and mixed germplasms was also verified by Yang

et al. (2011a), who used maize inbreds representative of the tropical, subtropical (CIMMYT), and Chinese (temperate) gene pools and genotyped SNP loci.
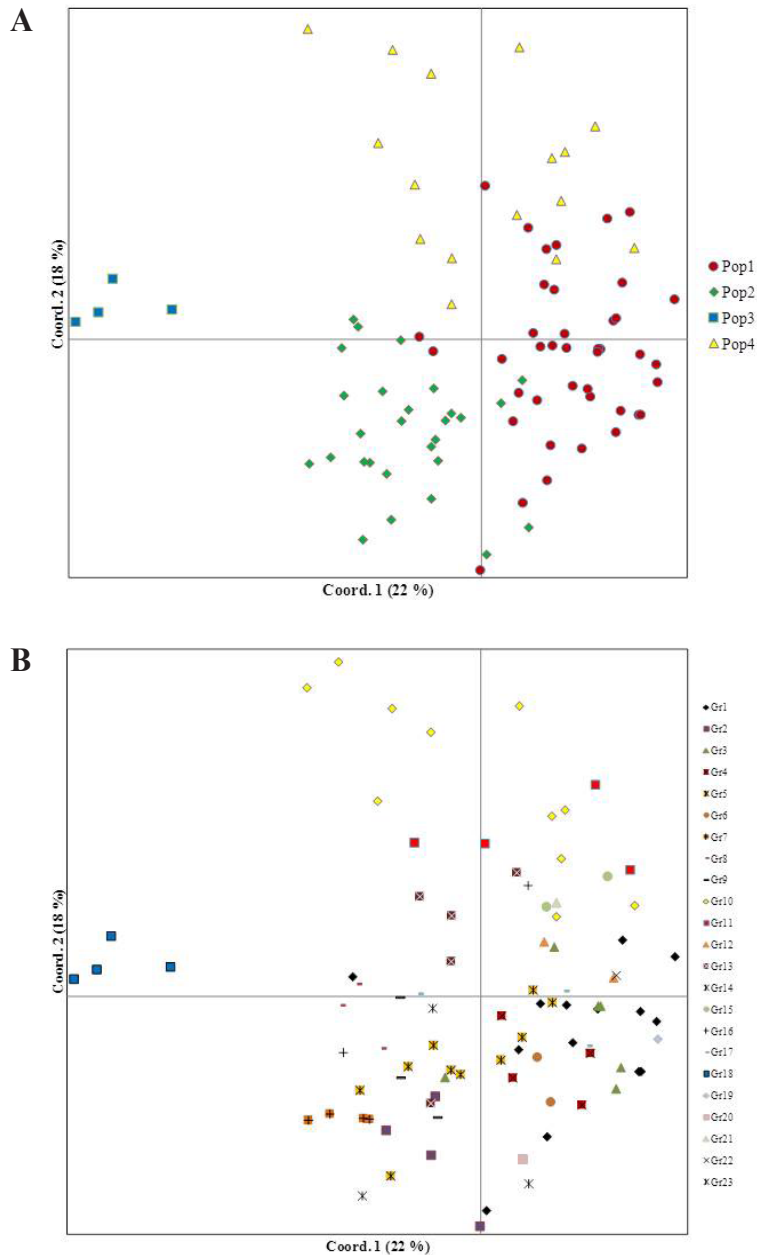

Figure 2. Diversity between the inbred lines according to the values of the first two principal coordinates, with clusters identified by population structure analysis **A.** and unweighted pair-group method with arithmetic means (UPGMA) **B**.

The degree of genetic differentiation and of intra- and inter-population genetic diversity relative to these tropical heterotic pools is comparable to that observed for North American, European, Chinese, Thai (temperate), and subtropical heterotic pools. The F_{ST} for the gene pools identified by Yang et al. (2011a,b), Xie et al. (2008), and Liu et al. (2003) ranged from 0.06 to 0.50. The average value relative to tropical/subtropical and temperate gene pools was 0.16. In the studies of Yang et al. (2011b), Van Inghelandt et al. (2010), Wang et al. (2008), Reif et al. (2005), and Liu et al. (2003), the ranges for the estimates of the number of alleles per locus, gene diversity, and private alleles were 2.7-21.7, 0.31-0.82, and 9-765, respectively. The superior number of alleles per locus and private alleles in the former and the latter studies is due to the high number of inbreds from broad geographic origins. Compared to the tropical heterotic pools evidenced in our study, the tropical/subtropical heterotic pool characterized by Liu et al. (2003) showed a higher number of alleles per locus (13.5) and gene diversity (0.81). It is worth noting that most (50-85%) of the private alleles observed in the three tropical heterotic pools are rare, with frequencies less than 0.1. Then, the private alleles with frequencies greater than 0.1 in the tropical heterotic pools ranged from 5 to 12. The higher allelic frequency variation within each tropical heterotic pool, as indicated by the AMOVA, and the high genetic diversity between the inbred lines, as evidenced by the cluster analysis based on Lynch's index and UPGMA, suggested heterotic groups within the main tropical heterotic pools. We identified 8, 7, and 3 groups within heterotic pools 1, 2, and 4, respectively. Greater allelic frequency variation among populations (89%) was also observed by Yang et al. (2011a). From the analysis of European maize elite inbred lines that were genotyped for AFLP and SSR markers, Stich et al. (2006) observed three groups within the heterotic pools dent and flint. The dispersion of the inbred lines relative to the first principal coordinates, which explain 40% of the total variation, clearly reveals the main heterotic pools but not the heterotic groups within each main gene pool (Figure 2).

The thought that flint and dent are the main heterotic groups of the tropical germplasm that is used by maize seed companies is not corroborated by our results. Each observed heterotic pool includes the flint type, indicating flint heterotic groups. The worldwide and Brazilian maize seed companies have used all three tropical heterotic pools to develop elite inbred lines; with few exceptions, at least one sampled inbred from each company belongs to each heterotic pool. Also with few exceptions, the inbred lines with a shared genome were also sampled from hybrids of all seeds companies. Furthermore, for at least one hybrid of each seed company, the sampled inbreds were characterized as derived from the same heterotic pool or from heterotic pools 1 and 2 or 1 and 4, suggesting crosses within (among heterotic groups) and among the heterotic pools. The heterotic groups within each heterotic pool include flint and semi-flint/semi-dent types.

ACKNOWLEDGMENTS

REFERENCES

Belaj A, Dominguez-García MC, Atienza SG, Urdíroz NM, et al. (2012). Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. *Tree Genet. Genome* 8: 365-378.
Bracco M, Lia VV, Gottlieb AM, Camara HJ, et al. (2009). Genetic diversity in maize landraces from indigenous settlements of Northeastern Argentina. *Genetica* 135: 39-49.

Calinski T and Harabasz J (1974). A dendrite method for cluster analysis. *Commun. Stat.* 3: 1-27.

Colless DH (1980). Congruence between morphometric and allozyme data for Menidiu species: a reappraisal. *Syst. Zool.* 29: 288-299.

Cruz CD (2008). Programa Genes: Diversidade Genética. Editora UFV, Viçosa.

Doyle JJT and Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.

Falush D, Stephens M and Pritchard JK (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.

Johnson SA and Wichern DW (1988). Applied Multivariate Statistical Analysis. 2nd edn. Prentice-Hall, New Jersey.

Jombart T, Pontier D and Dufour AB (2009). Genetic markers in the playground of multivariate analysis. *Heredity* 102: 330-341.

Laborda PR, Oliveira KM, Garcia AA, Paterniani ME, et al. (2005). Tropical maize germplasm: what can we say about its genetic diversity in the light of molecular markers? *Theor. Appl. Genet.* 111: 1288-1299.

Laurentin H (2009). Data analysis for molecular characterization of plant genetic resources. *Genet. Resour. Crop Evol.* 56: 277-292.

Legesse BW, Myburg AA, Pixley KV and Botha AM (2007). Genetic diversity of African maize inbred lines revealed by SSR markers. *Hereditas* 144: 10-17.

Legesse BW, Myburg AA, Pixley KV, Twumasi-Afriyie S, et al. (2008). Relationship between hybrid performance and AFLP based genetic distance in highland maize inbred lines. *Euphytica* 162: 313-323.

Li YH, Smulders MJM, Chang RZ, Qiu LJ, et al. (2011). Genetic diversity and association mapping in a collection of selected Chinese soybean accessions based on SSR marker analysis. *Conserv. Genet.* 12: 1145-1157.

Liu K, Goodman M, Muse S, Smith JS, et al. (2003). Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165: 2117-2128.

Lynch ML (1990). The similarity index and DNA fingerprinting. *Mol. Biol. Evol.* 7: 478-484.

Mohammadi SA and Prasanna BM (2003). Analysis of genetic diversity in crop plants-salient statistical tools and considerations. *Crop Sci.* 43: 1235-1248.

Orloci L (1978). Multivariate Analysis in Vegetation Research. Dr W. Junk B. V, The Hague.

Peakall R and Smouse PE (2006). GENALEX 6: genetic analysis in Excel Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295.

Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Rao CR (1952). An Advanced Statistical Method in Biometric Research. John Wiley & Sons, New York.

REAL Software (2004). REALbasic 5.5. Version 5.5.4. REAL Software Inc. Austin.

Reif JC, Melchinger AE and Frisch M (2005). Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Sci.* 45: 1-7.

Rohlf FJ (2009). NTSYSpc: Numerical Taxonomy and Multivariate Analysis System. Exeter Software, New York.

Santacruz-Varela A, Widrlechner MP, Ziegler KE and Salvador RJ (2004). Phylogenetic relationships among North American popcorns and their evolutionary links to Mexican and South American popcorns. *Crop Sci.* 44: 1456-1467.

SAS Institute (2002). SAS/STAT User's Guide Version 90. 1st edn. SAS Institute Inc, Cary.

Smouse PE and Peakall R (1999). Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82 (Pt 5): 561-573.

Sokal RR and Rohlf FJ (1962). The comparison of dendrograms by objective methods. *Taxon* 11: 33-40.

Stich B, Maurer HP, Melchinger AE and Frisch M (2006). Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. *Mol. Breed.* 17: 217-226.

Van Inghelandt D, Melchinger AE, Lebreton C and Stich B (2010). Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *Theor. Appl. Genet.* 120: 1289-1299.

Wang R, Yu Y, Zhao J, Shi Y, et al. (2008). Population structure and linkage disequilibrium of a mini core set of maize inbred lines in China. *Theor. Appl. Genet.* 117: 1141-1153.

Xia XC, Reif JC, Hoisington DA and Melchinger AE (2004). Genetic diversity among CIMMYT maize inbred lines investigated with SSR markers: I. lowland tropical maize. *Crop Sci.* 44: 2230-2237.

Xia XC, Reif JC, Melchinger AE, Frisch M, et al. (2005). Genetic diversity among CIMMYT maize inbred lines investigated with SSR markers: II. subtropical tropical midaltitude and highland maize inbred lines and their relationships with elite US and European maize. *Crop Sci.* 45: 2573-2582.

Xie CX, Warburton M, Li MS, Li XH, et al. (2008). An analysis of population structure and linkage disequilibrium using multilocus data in 187 maize inbred lines. *Mol. Breed.* 21: 407-418.

Yan JB, Shah T, Warburton M, Buckler ES, et al. (2009). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One* 4: e8451.

Yang X, Gao S, Xu S, Zhang Z, et al. (2011a). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* 28: 511-526.

Yang X, Xu Y, Shah T, Li H, et al. (2011b). Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. *Genetica* 139: 1045-1054.