

Genetic Diversity Determined within and among CIMMYT Maize Populations of Tropical, Subtropical, and Temperate Germplasm by SSR Markers

J. C. Reif, X. C. Xia, A. E. Melchinger,* M. L. Warburton, D. A. Hoisington, D. Beck, M. Bohn, and M. Frisch

ABSTRACT

Genetic diversity in maize (*Zea mays* L.) plays a key role for future breeding progress. The main objectives of our study were to (i) investigate the genetic diversity within and among CIMMYT maize populations by simple sequence repeat (SSR) markers, (ii) examine genotype frequencies for deviations from Hardy-Weinberg equilibrium (HWE) at individual loci, and (iii) test for linkage disequilibrium (LD) between pairs of loci. Twenty-three maize populations and pools established in 1974, which mostly comprise germplasm from different racial complexes adapted to tropical, subtropical intermediate-maturity, subtropical early-maturity, and temperate megaenvironments (ME), were fingerprinted by 83 SSR markers covering the entire maize genome. Across all populations, 27% of the SSR markers deviated significantly from HWE with an excess of homozygosity in 99% of the cases. We observed no evidence for genome-wide LD among pairs of loci within each of the seven tropical populations analyzed. Estimates of genetic differentiation (G_{ST}) between populations within MEs averaged 0.09 and revealed that most of the molecular variation was found within the populations. Principal coordinate analysis based on allele frequencies of the populations revealed that populations adapted to the same ME clustered together and, thus, supported clearly the ME structure.

GENETIC DIVERSITY in maize is a valuable natural resource and plays a key role for future breeding progress. Germplasm collections as a source of genetic diversity must be well characterized for efficient management and effective exploitation. Achieving this goal in curating of gene banks is hampered by rising costs, decreasing budgets, and large collection sizes. The germplasm collection sizes should be optimized to provide maximal preservation of genetic variation and minimal redundancy with regard to genotypes, gene complexes, or possibly even genes (Kresovich et al., 1992).

Association mapping was proposed as one approach to detect genes and alleles of interest in germplasm collections (Lynch and Walsh, 1997). The resolution of association studies in a sample depends on the extent of linkage disequilibrium (LD) across the genome. LD (or the correlation between alleles of different loci) depends generally on the genealogy of the germplasm.

Besides this, drift and selection within populations can also cause LD. The genomic structure of LD must be empirically determined before embarking on association studies because it can vary among samples of germplasm. The advent of PCR-based molecular markers such as SSRs has created an opportunity for fine-scale genetic characterization of germplasm collections. Since SSR markers are highly polymorphic (Smith et al., 1997), easy to generate, and highly repeatable (Heckenberger et al., 2002), they can be used for large-scale investigations as needed in the case of genetic resources (Powell et al., 1996).

CIMMYT developed and improved from 1964 until 1973 a wide array of maize germplasm. Populations were established with materials from a single racial complex. In 1974, a major shift in the organization of the germplasm was initiated. Germplasm from different racial complexes was mixed and more than 100 populations were established to (i) reduce the large collection of germplasm from CIMMYT's gene bank to a number that can be handled efficiently in a breeding program and (ii) use the combining ability of different germplasm sources for intrapopulation improvement. In addition, 30 broad-based back-up pools were formed as an insurance against narrowing the germplasm base of the populations (CIMMYT, 1998). These pools and populations have played an important role in maize breeding and production in developing countries and have been exploited as sources of new germplasm for temperate regions (Ron Parra and Hallauer, 1997). Detailed knowledge about LD and genetic diversity of these populations would increase the efficiency of their use in breeding. However, little is known about the molecular diversity in tropical and subtropical maize populations (Warburton et al., 2002) and information about LD in this germplasm is entirely lacking.

The main objectives of our study were to characterize the population genetic structure of 23 CIMMYT maize populations as a basis for an efficient use of this germplasm in breeding programs. In particular we (i) investigated the molecular genetic diversity within and among 23 of CIMMYT's maize populations, (ii) examined genotype frequencies for deviations from Hardy-Weinberg equilibrium at individual loci, and (iii) tested for LD between pairs of loci.

J.C. Reif, A.E. Melchinger, and M. Frisch, Institute of Plant Breeding, Seed Science, and Population Genetics, Univ. of Hohenheim, 70593 Stuttgart, Germany; M. Bohn, Crop Science Dep., Univ. of Illinois, 1102 South Goodwin Avenue, Urbana, IL 61801; M.L. Warburton, D.A. Hoisington, and D. Beck, International Maize and Wheat Improvement Center (CIMMYT), Apdo. Postal 6-641 06600 Mexico D.F., Mexico; X.C. Xia, Institute of Crop Breeding and Cultivation, Chinese Academy of Agric. Sciences, Zhongguancun South Street 12, 100081, Beijing, China. 30 April 2003. *Corresponding author (melchinger@uni-hohenheim.de).

Published in Crop Sci. 44:326–334 (2004).
© Crop Science Society of America
677 S. Segoe Rd., Madison, WI 53711 USA

Abbreviations: CIMMYT, International Maize and Wheat Improvement Center; HWE, Hardy-Weinberg equilibrium; ME, megaenvironment; LD, linkage disequilibrium; MRD, modified Roger's distance; PC, principal coordinate; PCoA, principal coordinate analysis; QTL, quantitative trait locus; SSR, simple sequence repeat.

MATERIALS AND METHODS

Genetic Material

Twenty-three populations and pools (further referred to as populations, if not stated otherwise) of CIMMYT's maize program were investigated including tropical, subtropical, and temperate adapted material. The germplasm was grouped on the basis of its adaptation to four megaenvironments (MEs): tropical, subtropical intermediate-maturity, subtropical early-maturity, and temperate (Table 1). The populations were maintained periodically by recombining a minimum of 420 plants per population each generation for several decades.

SSR Analyses

Forty-eight individuals from each of the seven tropical populations and 21 individuals from each of the 16 subtropical and temperate populations were chosen at random and analyzed individually. DNA was extracted employing a modified CTAB procedure (Saghai-Maroo et al., 1984). We used the set of 83 SSR markers described by Warburton et al. (2002), which provides uniform coverage of the entire maize genome. Primers and PCR conditions were described in detail by Warburton et al. (2002). SSRs were multiplexed for a maximum efficiency. Fragments were separated using acrylamide gels run on an ABI 377 automatic DNA sequencer. Fragment sizes were calculated with GeneScan 3.1 (Perkin Elmer/Applied Biosystems, Foster City, CA) using the Local Southern sizing method (Elder and Southern, 1987); allele identity was assigned by Genotyper 2.1 software (Perkin Elmer/Applied Biosystems) and the two inbred lines, CML51 and CML292, as controls. Map position for all SSRs were based on the Pioneer Compos-

ite 1999 linkage map obtained from the MaizeDB website (<http://nucleus.agron.missouri.edu/>).

Statistical Analyses

The number of alleles per locus (further referred to as allelic richness) was determined for the entire set of 672 individuals analyzed and for various subsets within this collection (populations, MEs). The existence of population- or ME-specific alleles was determined. The total gene diversity (H_T) based on SSR data across all populations was decomposed into (i) gene diversity between individuals within each population (H_S) and (ii) gene diversity between populations within each ME (H_{ME}) according to Nei (1987, p. 164) and Chakraborty (1980). Confidence intervals for H_S values were obtained by a bootstrap procedure with resampling across markers and individuals. The coefficient of gene differentiation (G_{ST}) was used as a measure of genetic differentiation between populations of the same ME and different MEs and was calculated according to Nei (1987, p. 190). G_{ST} is the proportion of the total genetic diversity that is due to differences between MEs. The fixation index F_{IS} for each population was estimated according to Nei (1987, p. 164) as one minus the observed heterozygosity divided by the expected heterozygosity.

The average number of alleles, the number of unique alleles, H_S , and H_{ME} depend on the number of individuals analyzed per population. In the tropical populations, 48 individuals were sampled, whereas in the subtropical and temperate populations only 21 individuals were sampled per population. Therefore, we used a resampling strategy to obtain comparable estimates: a random sample of 21 individuals from each of the tropical populations was chosen for the above analyses,

Table 1. Description of the 23 CIMMYT maize populations analyzed in this study, grouped according to the adaptation to the four megaenvironments (MEs).

Population	Cycle	Germplasm description
Tropical ME		
P21	5	Composed of seven Tuxpeño races plus some families from P124.
P22	6	Includes Tuxpeño and ETO Blanco germplasm as well as germplasm from Central America.
P25	0	Composed of white flint selections from crosses among germplasm from Mexico, Columbia, the Caribbean, Central America, India, Thailand and the Philippines.
P29	5	Broad genetic base including Tuxpeño, Cuban flints, and ETO.
P32	5	On the basis of white flint germplasm from South America, Cuba, Mexico and the U.S. Corn Belt.
P43	5	Tuxpeño synthetic composed of 16 S ₁ lines.
P124	21	Mainly based on Tuxpeño germplasm, but includes also some materials from Central America, the Caribbean, and Zaire.
Subtropical intermediate-maturity ME		
P33	2	Contains predominantly Argentinian (Cateto) flints.
P34	5	Includes Cuban flints, ETO, Tuxpeño, and germplasm from the U.S. Corn Belt, India, and Nepal.
P42	4	An advanced generation of ETO selected for short-plant type and crossed with Illinois Corn Belt material.
P45	3	Includes U.S. Corn Belt germplasm, Tuxpeño, Cuban flints, Puerto Rico composite, and collections from the Dominican Republic.
P47	2	Consists largely of Tuxpeño germplasm plus some U.S. Corn Belt lines.
P131	14	A broad based pool including white flint segregates from Ecuador, Argentina, India, Mexico, Pool 32, and Pool 33, but also germplasm from Mexico, U.S. Corn Belt, Brazil, Uruguay, Argentina, China, Pakistan, Yugoslavia, Lebanon, Guatemala, Venezuela, Peru, Cuba, and the Dominican Republic.
P134	20	Includes germplasm from the Mexican lowlands and highlands, the U.S. Corn Belt, southern USA, Puerto Rico, Pakistan, Hungary, China, Peru, Pakistan, Lebanon, Nicaragua, and Guatemala.
Subtropical early-maturity ME		
P46	1	Represents a superior fraction (240 half-sib families) of Pool 29, which is based on germplasm from Europe, Lebanon, the USA Corn Belt, China, Indonesia, and South America.
P48	5	Composed of U.S. Corn Belt germplasm, southern European germplasm, and 54 half-sib families from Pool 30.
P127	20	Includes germplasm from the USA, China, Lebanon, Pakistan, and several European countries.
P128	14	On the basis of crosses between white dent segregates from P127 and Hungarian germplasm from Pool 30 and various other germplasm.
P130	15	Composed of germplasm from Europe, China, Lebanon, Mexico, South America, and the U.S. Corn Belt.
Temperate ME		
P139	12	Contains germplasm from the tropical lowlands and highlands, subtropical, and temperate areas.
P140	12	On the basis of germplasm from Europe.
P141	12	Includes predominantly U.S. Corn Belt germplasm plus germplasm from China, Korea, and Lebanon.
P142	12	On the basis of germplasm from Mexico, Peru, Bolivia, Pakistan, Hungary, USA, and Yemen.

sampling was repeated 1000 times, and the results were averaged.

The modified Roger's distance (MRD) between two populations or individuals was calculated according to Wright (1978, p. 91) and Goodman and Stuber (1983). Standard errors of MRD estimates were calculated by a bootstrap procedure with resampling across markers and individuals. Associations among operational taxonomic units were revealed by principal coordinate analysis (PCoA) (Gower, 1966) based on MRD values. PCoA were performed for (i) the 23 populations and (ii) all individuals of the populations from each ME. In the latter case, individuals with more than 30% missing values were excluded. All analyses were performed with Version 2 of the Plabsim software (Frisch et al., 2000), which is implemented as an extension to the statistical software R (Ihaka and Gentleman, 1996).

Alleles with frequencies smaller than 0.10 were pooled for each locus for tests of Hardy-Weinberg equilibrium (HWE) and LD because disequilibrium coefficients have large variances with rare alleles. The population genetic software Arlequin (Schneider et al., 2000) was used for tests of HWE at individual loci and LD between pairs of loci. Software Arlequin uses the procedure described by Guo and Thompson

(1992) to detect significant departures from HWE. LD between all pairs of loci was tested within each of the seven tropical populations using a likelihood-ratio test, whose empirical distribution is obtained by a permutation procedure (Slatkin and Excoffier, 1996). This test assumes HWE at each locus and, thus, only loci with no significant deviation from HWE were included in the analysis. An LD analysis of the 16 subtropical and temperate populations was not considered due to the small sample size of 21 individuals per population. In testing for both HWE and LD, the Bonferroni correction for multiple tests was applied (Snedecor and Cochran, 1980).

RESULTS

Population Genetic Analyses

We found a total of 666 alleles for the 83 SSR markers in the 672 genotypes (Table 2). All 83 marker loci analyzed were polymorphic across all 672 individuals. The average proportion of missing data across genotypes was 8%. At the population level, an average of 76 loci was polymorphic with a maximum of 81 (PI42) and a minimum of 63 (PI40). The percentage of loci with

Table 2. Genetic diversity within and among 23 maize populations as revealed by SSRs.

Population	No. of individuals	Avg. no. of alleles per locus	No. of unique alleles [†]	Gene diversity [‡]	Fixation index (F_{IS})	95% confidence interval for F_{IS}
Tropical germplasm						
P21	48	4.20	13	0.51 ^{ab}	0.27	(0.19,0.34)
P22	48	4.14	4	0.51 ^{ab}	0.25	(0.17,0.31)
P25	48	3.93	8	0.51 ^{ab}	0.23	(0.15,0.29)
P29	48	4.24	7	0.55 ^b	0.25	(0.17,0.32)
P32	48	3.47	3	0.45 ^a	0.31	(0.19,0.35)
P43	48	3.55	3	0.48 ^{ab}	0.28	(0.19,0.35)
PI24	48	4.22	1	0.53 ^{ab}	0.22	(0.13,0.30)
Total	336	6.07	86	0.56		
Resampled tropical germplasm§						
P21	21	3.74	12	0.50 ^{ab}	0.27	(0.18,0.35)
P22	21	3.71	6	0.50 ^{ab}	0.24	(0.15,0.31)
P25	21	3.56	7	0.50 ^{ab}	0.22	(0.15,0.30)
P29	21	3.83	5	0.54 ^{ab}	0.23	(0.15,0.32)
P32	21	3.08	2	0.45 ^{ab}	0.27	(0.19,0.35)
P43	21	3.23	3	0.47 ^{ab}	0.28	(0.19,0.38)
PI24	21	3.79	1	0.52 ^{ab}	0.21	(0.11,0.30)
Total	147	5.62	35	0.55		
Subtropical intermediate-maturity germplasm						
P33	21	3.73	4	0.55 ^{ab}	0.44	(0.30,0.50)
P34	21	3.71	3	0.52 ^{ab}	0.41	(0.28,0.47)
P42	21	3.58	2	0.54 ^{ab}	0.36	(0.24,0.43)
P45	21	3.80	4	0.57 ^b	0.37	(0.24,0.42)
P47	21	3.65	7	0.54 ^{ab}	0.37	(0.24,0.42)
PI31	21	3.78	1	0.59 ^{ab}	0.65	(0.54,0.72)
PI34	21	3.90	5	0.56 ^{ab}	0.41	(0.28,0.48)
Total	147	5.86	37	0.62		
Subtropical early-maturity germplasm						
P46	21	3.48	3	0.54 ^{ab}	0.41	(0.27,0.46)
P48	21	3.36	1	0.51 ^{ab}	0.36	(0.26,0.42)
PI27	21	3.78	8	0.55 ^{ab}	0.38	(0.25,0.44)
PI28	21	4.00	4	0.58 ^b	0.43	(0.29,0.48)
PI30	21	3.75	3	0.55 ^{ab}	0.40	(0.27,0.47)
Total	105	5.43	23	0.60		
Temperate germplasm						
P139	21	4.13	6	0.58 ^b	0.40	(0.28,0.43)
P140	21	3.81	4	0.55 ^{ab}	0.41	(0.29,0.45)
P141	21	3.69	5	0.57 ^b	0.39	(0.26,0.43)
P142	21	3.96	5	0.57 ^b	0.42	(0.30,0.47)
Total	84	5.34	22	0.61		
Grand total	672	8.02		0.62		

[†] Number of unique alleles with respect to the total number of 672 alleles found in all 23 populations.

[‡] Gene diversity values followed by the same letters are not different at the 0.05 significance level according to a bootstrap procedure.

[§] A random sample of 21 individuals from each tropical population was chosen. Sampling was repeated 1000 times and the results averaged.

significant ($P < 0.01$) deviations from HWE varied from 14% (P48) to 40% (PI34) with an average of 27% (Fig. 1). In 99% of the cases, deviations from HWE were attributable to an excess of homozygosity. F_{IS} values ranged from 0.21 (PI24) to 0.65 (PI31) (Table 2) across germplasm types. The number of unique alleles for the various MEs ranged from 22 (temperate ME) to 86 (tropical ME). Significant differences ($P < 0.05$) were found among the H_s values within the 23 populations with a range from 0.45 (P32) to 0.59 (PI31) and a mean of 0.54. Gene diversity between populations within a given ME (H_{ME}) was highest in the subtropical intermediate-maturity populations (0.62) and smallest in the tropical populations (0.56). The coefficient of gene differentiation G_{ST} between the populations within MEs averaged 0.09 with little variation among the four MEs. Gene differentiation G_{ST} between the MEs was 0.02.

Relationships between Populations

Values of MRD between pairs of populations averaged 0.28 and ranged from 0.20 (P22 \times P124) to 0.41 (P32 \times P48) with significant differences ($P < 0.01$) between MRD estimates (Table 3). The average MRD between all pairs of populations within MEs ranged from 0.22 (temperate ME) to 0.26 (subtropical intermediate-maturity ME) and averaged 0.25. The average MRD between all pairs of populations of different MEs was maximum for tropical \times subtropical early-maturity populations (0.32) and minimum for subtropical early-maturity \times temperate populations (0.24).

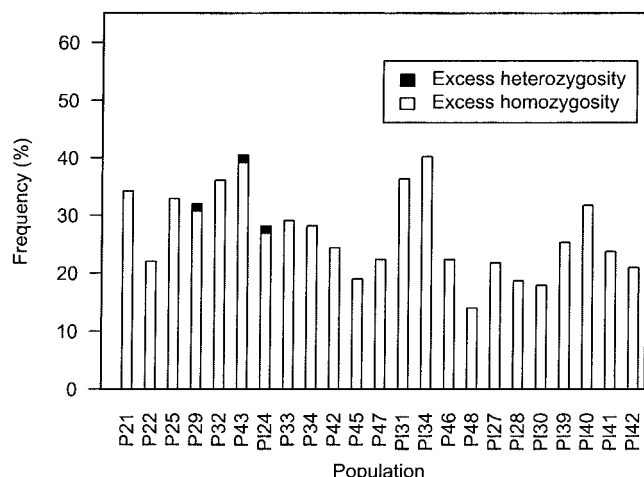


Table 3. Modified Roger's distances between populations (average standard error 0.02).

[illegible]

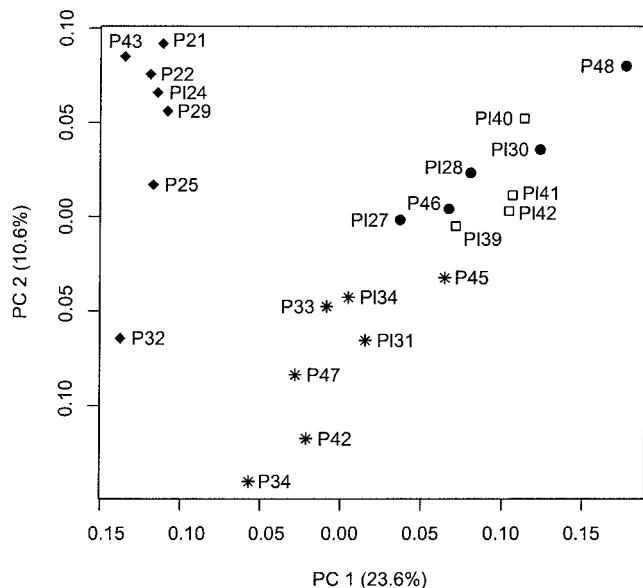


Fig. 2. Principal coordinate analysis of the 23 maize populations based on modified Roger's distance calculated from the allele frequencies of the populations. PC1 and PC2 are the first and second principal coordinates, respectively. Tropical (◆), subtropical intermediate-maturity (*), subtropical early-maturity (●), and temperate germplasm (□).

three ME populations by PC1. The subtropical intermediate-maturity populations were separated from the subtropical early-maturity and temperate populations by PC2. The temperate populations were positioned adjacent to the subtropical early-maturity populations.

In the tropical populations, individuals of P21, P22, P29, and PI24 clustered together and were clearly separated from individuals of P25, P32, and P43 (Fig. 3). Within the subtropical intermediate-maturity populations, individuals from P34 and P42 clustered together. Individuals of PI31 were widely spread across the two PCs, whereas individuals of P45 and P47 formed largely separated groups. In the subtropical early-maturity populations, PC1 clearly separated individuals of PI27 and P46 from individuals of PI30 and P48, whereas individuals of PI28 were positioned in between these two groups. The individuals of the four temperate pools were widely scattered across both PCs, with PI39 being most clearly separated from PI42.

DISCUSSION

Research on genetic diversity in maize with molecular markers has mostly concentrated on temperate inbred lines and their pedigree relationships as well as assignment to heterotic groups (Melchinger, 1999). Only a few studies have investigated the genetic diversity and structure of traditional maize populations from Europe (Gauthier et al., 2002) and the U.S. Corn Belt (Labate et al., 2003). In contrast to these populations, which have originated from different geographic regions and maintained separately by farmers and early breeders, the 100 populations and 30 pools at CIMMYT have a fairly short evolutionary history. These germplasm groups were created by breeders in 1974 mostly by in-

termating different racial complexes (Vasal et al., 1999). The intermixing of diverse germplasm within populations complicates detection of relationships among these populations based on pedigree information. Hence, we employed molecular markers to analyze associations among the CIMMYT maize pools and populations.

Hardy-Weinberg Equilibrium

The Hardy-Weinberg law describes the fundamental observation that in a large random-mating population both gene frequencies and genotype frequencies are constant across generations assuming absence of migration, mutation, and selection. The genotype frequencies are determined by the gene frequencies (Falconer and Mackay, 1996, p. 5). CIMMYT breeders have maintained their populations by planting a minimum of 20 rows and 21 plants per row. All plants consistent with the varietal description were shoot bagged and pollen from 10 rows was bulked to pollinate plants in the other 10 rows and vice versa. A minimum of 300 to 350 typical ears from the pollinated plants was chosen to represent each population. Considering the procedure to maintain the germplasm, it was expected that the populations would be in HWE after one generation of random mating. However, all 23 maize populations deviated significantly from HWE (Fig. 1) and showed a deficit of heterozygous loci (Table 2). This is in agreement with previous reports on other maize populations. Labate et al. (2000) investigated two random-mated maize populations and found that 27% of tests for deviation from Hardy-Weinberg equilibrium were significant, with deviations occurring due to an excess of homozygosity of 72 and 87%. Dubreuil and Charcosset (1998) also detected an excess of homozygosity in 10 populations from Europe and the U.S. using RFLP markers. In 17 open-pollinated populations assayed at 13 enzyme marker loci, 27% of Hardy-Weinberg tests were significant, with 94% showing an excess of homozygosity (Kahler et al., 1986).

The inbreeding of the populations observed in our study can be related to various causes: (i) positive assortative matings between individuals (homogamy), (ii) artificial subgrouping of individuals from populations, (iii) selection favoring homozygotes, and (iv) experimental errors during the laboratory assay for SSRs. Even though precautions were taken to avoid positive assortative mating between individuals, it cannot be excluded entirely because late flowering plants are preferentially crossed to late ones, and early flowering plants with early ones. However, only SSRs closely linked to QTL for flowering time should show a higher degree of homozygosity than expected under HWE. Assortative mating can be one reason for an artificial subgrouping of individuals, but a closer examination of the PCoAs (Fig. 3) did not provide any clue that the deficit of heterozygous individuals could be related to a subgrouping of individuals from populations. Selection favoring homozygotes is unlikely in maize, where fitness increases with heterozygosity. The choice of SSRs with tri- and higher repeats in our study and the use of the Local Southern sizing method to estimate the fragment

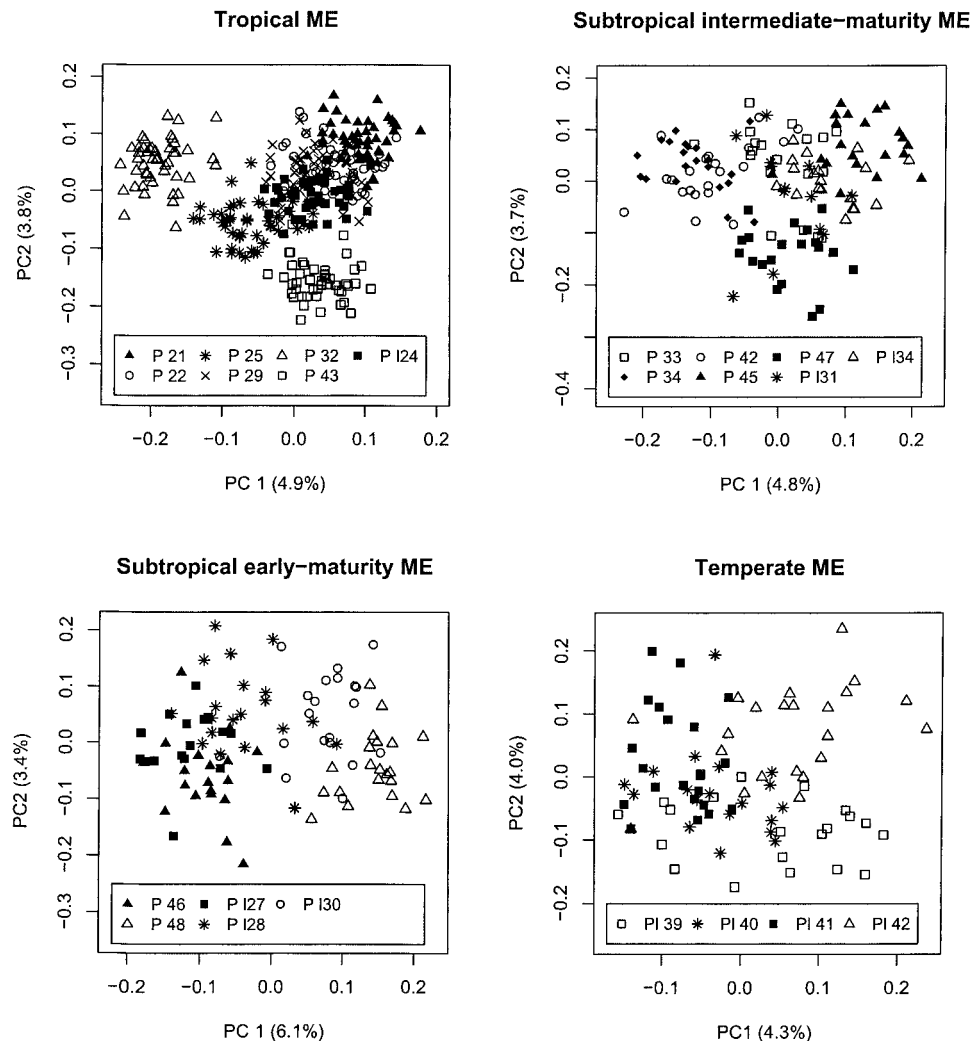


Fig. 3. Principal coordinate analyses of the individuals of the 23 populations grouped into four megaenvironments (ME) based on modified Roger's distance. PC1 and PC2 are the first and second principal coordinate, respectively, and numbers in parentheses refer to the proportion of variance explained by the principal component in the specific ME group.

sizes, which represents a conservative allele-calling procedure, reduce the laboratory error sources, which cause overestimation of heterozygosity. However, most experimental errors would lead to an overestimation of homozygotes because (i) a heterozygous locus carrying a null allele would be scored as a homozygous locus, (ii) alleles could not be detected because of competition during the PCR reaction, and (iii) the setting of the threshold of band intensity to detect alleles can be too strict.

Thus, experimental errors are probably the major cause of heterozygote deficiency within the populations apart from genuine genetic causes. To separate both sources, it would be prudent in future studies to include, besides the two inbred checks, their hybrid as a control to estimate the error rate for misscoring of heterozygous loci.

Linkage Disequilibrium

LD can result from and be maintained by epistasis (Falconer and Mackay, 1996, p. 16). It can also arise from admixture of populations with different gene fre-

quencies, or from drift in small populations. Since population admixture happened during the establishment of the tropical populations recently, this could have caused LD. However, we found that less than 0.3% of the two-locus disequilibrium tests were significant, which can be explained by type I error alone. This is in accordance with a study reported by Stuber et al. (1980), who evaluated LD among eight enzyme loci in four long-term maize selection experiments. In contrast to these results, Remington et al. (2001) reported evidence of genome-wide LD among 47 SSRs for 102 maize inbred lines from temperate and tropical regions. LD was reduced but not eliminated by grouping lines into three empirically determined subpopulations. Nevertheless, artificial population admixture within the subpopulations caused by sampling lines from different germplasm sources could be one reason for the detection of LD in this survey. On one hand, the lack of LD in our study can be explained by the low-density marker map and the decrease of LD with successive generations of intermating since the establishment of the populations.

On the other hand, the sample size of 48 individuals per population, the precision in estimating haplotype frequencies with the EM algorithm (Excoffier and Slatkin, 1995), and the elimination of loci deviating from HWE (Fig. 1) result in a low power to detect LD. Further investigations are required to examine the influence of the sample size and the structure of the population on the power of detecting LD.

Molecular Diversity of the Populations

We observed a higher total molecular allelic richness (8.02 alleles per locus) and average molecular allelic richness of the populations adapted to different MEs (5.7 alleles per locus) than reported in previous SSR studies of maize germplasm, although the Local Southern sizing method used to estimate the fragment sizes represents a conservative allele-calling procedure. Labate et al. (2003) found an average of 6.5 alleles per locus analyzing 461 plants representing a diverse array of U.S. germplasm. Matsuoka et al. (2002) found, on average, 6.9 alleles per locus for 101 maize inbred lines representing three major germplasm sources (Tropical, U.S., and Canadian/European inbreds). The total gene diversity across all populations (0.62) in our study was the same as reported by Matsuoka et al. (2002). The high molecular allelic richness (Table 1) and gene diversity values in our study confirm the broad genetic base of the populations expected from the pedigree data (Table 1).

The allelic richness and number of unique alleles were significantly higher for the tropical populations (6.07 alleles per locus, 86 alleles, respectively) than for populations adapted to the three other MEs (5.86, 5.43, and 5.34 alleles per locus, 37, 23, and 22 alleles). However, the results of the resampled tropical populations obtained from the same number of individuals per population as sampled in each of the subtropical and temperate populations clearly demonstrated the importance of the number of individuals investigated: the more individuals that are sampled, the higher is the probability of detecting rare alleles. The low difference of the pools compared with the populations with respect to average gene diversity (H_s), number of alleles, and number of unique alleles (Table 2) was surprising because (i) the pools have been assessed with a larger effective population size than were the populations and (ii) new material has been regularly introgressed into the pools. Our results indicate that a loss of rare alleles in the populations caused by drift seems to be uncommon and suggests that maintaining back-up pools is not necessary.

Genetic Structure of the Populations

PCoA based on MRD of the populations (Fig. 2) clearly supported the ME structure. PC1, which explained 23.6% of the total variance, revealed a major split between the (i) tropical, (ii) subtropical intermediate-maturity, and (iii) subtropical early-maturity and temperate ME. The position of the four temperate pools between the two groups of the subtropical early-maturity populations (P46, P127 vs. P48, P130) can be explained by the germplasm base of these populations

and pools (Table 1) and the similar selection pressure applied while adapting them to winter maize areas in the subtropics and tropics.

Most of the variation was found within the populations and just a minor part (9% on average) between the populations. The higher G_{ST} values for the tropical, subtropical intermediate-maturity, and subtropical early-maturity populations (0.10, 0.09, and 0.09, respectively) than for the temperate populations (0.07) can be explained by the pedigree information (Table 1). In the temperate populations, many germplasm sources were combined to establish broad-based pools comprising different racial complexes. An analysis of G_{ST} values for individual loci revealed that the following SSRs were associated with the structuring of the germplasm: phi014, phi031, phi053, and phi112. Such a tendency may indicate that the chromosomal regions harboring these SSRs are not selectively neutral. Several studies reported QTL for the anthesis-silking interval in the vicinity of phi014 and phi031 (Ribaut et al., 1996; Veldboom et al., 1994). QTL for days to pollen were reported in chromosomal regions near phi053 and phi112 (CIM-MYT, unpublished data). This seems to be an interesting starting point for further fine-scale and association mapping approaches of the underlying genes.

The clustering observed in the tropical populations is largely consistent with the pedigree information (Table 1). P124 was formed of Tuxpeño germplasm. P21 was established from seven Tuxpeño races and some families from P124. Although P43 was derived from Tuxpeño germplasm, it did not cluster closely with P21, consistent with field data that show high levels of heterosis between P21 and P43. In addition to Tuxpeño germplasm, P22 and P29 contain other materials such as ETO or Cuban flint. However, the results of the PCoA suggest that both populations (P29 and P22) contain mainly Tuxpeño germplasm. P21 and P32 were widely separated in the PCoA, consistent with numerous reports showing substantial heterosis between Tuxpeño and ETO germplasm (Wellhausen, 1978).

In the subtropical intermediate-maturity populations, individuals of P34 and P42 clustered together, consistent with the pedigree information. P42 and P34 both contain ETO germplasm. The latter includes also Cuban flints and Tuxpeño germplasm. The wide distribution of P131 over the first and second PCs can be explained by the broad range of germplasm used in its formation (Table 1). Individuals of P131 overlapped with individuals of P47 and P134, which is again consistent with pedigree information. P47 was formed using 276 half-sibs of P132, which itself was established with germplasm from the same sources as P131. Individuals of P45 are adjacent to individuals of P33 and have an intersection with them. P45 contains mainly Tuxpeño and U.S. dents but also Cuban flint, the latter being related to Cateto flint from P33 (Goodman and Brown, 1988).

In the subtropical early-maturity populations, two clearly separated clusters were observed: individuals of P127 and P46 vs. P130 and P48. Individuals of P128 were positioned midway between these two groups, which was again in accordance with pedigree information. P127

and P46 were both established using flint germplasm from different countries. P48 was generated from 54 half-sib families of PI30, which was established from dent germplasm from Europe, China, Lebanon, South America and the U.S. Corn Belt. In contrast, PI28 was developed by mixing flint and dent germplasm from PI27 and PI30.

In the temperate populations, the individuals of the four pools were widely spread over the first and second PC and only PI42 was separated from the three other pools. This reflects nicely the selection history and the establishment of the germplasm (Table 1). PI42 was formed to introduce tropical germplasm into temperate areas, whereas PI39, PI40, and PI41 were designed to introgress temperate germplasm for the winter maize areas in the subtropics and tropics.

The analysis of the 23 maize populations clearly revealed that most of the genetic diversity is within the populations and just a minor part between the populations. This can be explained by the establishment of the populations and pools, which mostly disregarded racial complexes, and suggests that the applied procedures to handle the broad range of available germplasm was suboptimal with regard to (i) maintaining maximum genetic diversity within the populations and (ii) conserving genetic diversity between the populations. It is rather likely that desired alleles, which occurred with high frequency in just one racial complex, can be lost by mixing different germplasm sources.

Germplasm based on different racial complexes might be useful for the improvement of open-pollinated varieties. However, this germplasm is less suitable for hybrid breeding, where clearly distinct heterotic groups are advantageous (Melchinger, 1999). The reduced genetic diversity among the populations caused by admixture can only be recovered by long-term isolation or reciprocal recurrent selection programs. Therefore, only the few populations based on one racial complex (P21, P32, P33, P42, P43, and PI24) seem to be suitable for hybrid breeding programs. If no populations based on one racial complex are available for a certain ME, breeders can use either populations not adapted to the ME or landraces in their search for germplasm suitable for hybrid breeding.

ACKNOWLEDGMENTS

The molecular marker analyses of this research were supported by funds from the German "Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung" Project No. 98.7860.4-001-01. The authors thank three anonymous reviewers for their valuable suggestions.

REFERENCES

- Chakraborty, R. 1980. Gene-diversity analysis in nested subdivided populations. *Genetics* 96:721–726.
- CIMMYT. 1998. A complete listing of maize germplasm from CIMMYT. Maize Program Special Report. Mexico DF, Mexico.
- Dubreuil, P., and A. Charcosset. 1998. Genetic diversity within and among maize populations: A comparison between isozyme and nuclear RFLP loci. *Theor. Appl. Genet.* 96:577–587.
- Elder, J.K., and E.M. Southern. 1987. Computer-aided analysis of one-dimensional restriction fragments gels. p. 165–172 *In* M.J. Bishop and C.J. Rawling (ed.) *Nucleic acid and protein sequence analysis—A practical approach*. IRL Press, Oxford, UK.
- Excoffier, L., and M. Slatkin. 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* 12:921–927.
- Falconer, D.S., and T.F.C. Mackay. 1996. *Introduction to quantitative genetics*. 4th ed. Longman Group Ltd., London.
- Frisch, M., M. Bohn, and A.E. Melchinger. 2000. Plabim: Software for simulation of marker-assisted backcrossing. *J. Hered.* 91:86–87.
- Gauthier, P., B. Gouesnard, L. Dallard, R. Redaelli, C. Rebourg, A. Charcosset, and A. Boyat. 2002. RFLP diversity and relationships among traditional European maize populations. *Theor. Appl. Genet.* 105:91–99.
- Goodman, M.M., and C.W. Stuber. 1983. Races of maize: VI. Isozyme variation among races of maize in Bolivia. *Maydica* 28:169–187.
- Goodman, M.M., and W.L. Brown. 1988. Races of Corn. p. 39–74. *In* G.F. Sprague and J.W. Dudley (ed.) *Corn and corn improvement*. 3rd ed. Agron. Monogr. 18. ASA, CSSA, and SSSA, Madison, WI.
- Gower, J.C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53:325–338.
- Guo, S., and E. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361–372.
- Heckenberger, M., A.E. Melchinger, J.S. Ziegler, L.K. Joe, J.D. Hauser, M. Hutton, and M. Bohn. 2002. Variation of DNA fingerprints among accessions within maize inbred lines with regard to the identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data. *Mol. Breed.* 10:181–191.
- Ihaka, R., and R. Gentleman. 1996. A language for data analysis and graphics. *J. of Computational and Graphical Statistics*, Vol. 5. 3:299–314.
- Kahler, A.L., A.R. Hallauer, and C.O. Gardner. 1986. Allozyme polymorphisms within and among open-pollinated and adapted exotic populations of maize. *Theor. Appl. Genet.* 72:592–601.
- Kresovich, S., J.G.K. Williams, J.R. Mc Ferson, E.J. Routman, and B.A. Schaal. 1992. Characterization of genetic identities and relationships of *Brassica oleracea* L. via a random amplified polymorphic DNA assay. *Theor. Appl. Genet.* 85:190–196.
- Labate, J.A., K.R. Lamkey, S.H. Mitchell, S. Kresovich, H. Sullivan, and J.S.C. Smith. 2003. Molecular and historical aspects of Corn Belt dent diversity. *Crop Sci.* 43:80–91.
- Labate, J.A., K.R. Lamkey, M. Lee, and W. Woodman. 2000. Hardy-Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. *Maydica* 45:243–255.
- Lynch, M., and B. Walsh. 1997. *Genetics and analysis of quantitative traits*. p. 413. Sinauer Assoc., Sunderland, MA.
- 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.
- Melchinger, A.E. 1999. Genetic diversity and heterosis. Chapter 10. *In* J.G. Coors and S. Pandey (ed.) *The genetics and exploitation of heterosis in crops*. CSSA, Madison, WI.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia University Press, New York.
- Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. *Mol. Breed.* 2:225–238.
- Remington, D.L., J.M. Thornsberry, Y. Matsuoka, L.M. Wilson, S.R. Whitt, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. (USA)* 98:11479–11484.
- Ribaut, J.M., D.A. Hoisington, J.A. Deutsch, C. Jiang, and D. Gonzalez de Leon. 1996. Identification of quantitative trait loci under drought conditions in tropical maize. I. Flowering parameters and the anthesis-silking interval. *Theor. Appl. Genet.* 92:905–914.
- Ron Parra, J., and A.R. Hallauer. 1997. Utilisation of exotic maize germplasm. *Plant Breed. Rev.* 14:165–187.
- Saghai-Marouf, M.A., K.M. Soliman, R. Jorgenson, and R.W. Allard. 1984. Ribosomal DNA spacer length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. *Proc. Natl. Acad. Sci. (USA)* 81:8014–8018.
- Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin, ver. 2.0:

- A software of population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Slatkin, M., and L. Excoffier. 1996. Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity* 76:377–383.
- Smith, J.S.C., E.C.L. Chin, H. Shu, O.S. Smith, S.J. Wall, M.L. Senior, S.E. Mitchell, S. Kresovich, and J. Ziegler. 1997. An evaluation of 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.
- Snedecor, G.W., and W.G. Cochran. 1980. *Statistical methods*. Iowa State Univ. Press, Ames.
- Stuber, C.W., R.H. Moll, M.M. Goodman, H.E. Schaffer, and B.E. Weir. 1980. Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays* L.). *Genetics* 93:225–236.
- Vasal, S.K., H.S. Cordova, S. Pandey, and G. Srinivasan. 1999. Tropical maize and heterosis. Chapter 34. *In* J.G. Coors and S. Pandey (ed.) *The genetics and exploitation of heterosis in crops*. CSSA, Madison, WI.
- Veldboom, L.R., M. Lee, and W.L. Woodman. 1994. Molecular marker-facilitated studies in an elite maize population: 1. Linkage analysis and determination of QTL for morphological traits. *Theor. Appl. Genet.* 88:7–16.
- Warburton, M.L., X. Xianchun, J. Crossa, J. Franco, A.E. Melchinger, M. Frisch, M. Bohn, and D. Hoisington. 2002. Genetic characterization of CIMMYT inbred maize lines and open pollinated populations using large scale fingerprinting methods. *Crop Sci.* 42:1832–1840.
- Wellhausen, E.J. 1978. Recent developments in maize breeding in the tropics. p. 59–91. *In* D.B. Walden (ed.) *Maize breeding and genetics*. John Wiley & Sons, New York.
- Wright, S. 1978. *Evolution and genetics of populations*, Vol. IV. p. 91. The Univ. of Chicago Press, Chicago.