

## ORIGINAL PAPER

J. C. Reif · M. L. Warburton · X. C. Xia  
D. A. Hoisington · J. Crossa · S. Taba  
J. Muminović · M. Bohn · M. Frisch  
A. E. Melchinger

## Grouping of accessions of Mexican races of maize revisited with SSR markers

Received: 11 July 2005 / Accepted: 27 March 2006 / Published online: 22 April 2006  
© Springer-Verlag 2006

**Abstract** Mexican races of maize (*Zea mays* L.) represent a valuable genetic resource for breeding and genetic surveys. We applied simple sequence repeat (SSR) markers to characterize 25 accessions of races of maize from Mexico. Our objectives were to (1) study the molecular genetic diversity within and among these accessions and (2) examine their relationships as assumed previously on the basis of morphological data. A total of 497 individuals were fingerprinted with 25 SSR markers. We observed a high total number of alleles (7.84 alleles per locus) and total gene diversity (0.61), confirming the broad genetic base of the maize races from Mexico. In addition, the accessions were grouped into distinct racial complexes on the basis of a model-based clustering approach. The principal coordinate analyses of the four Modern Incipient hybrids corroborated the proposed parental races of Chalqueño, Cónico Norteño, Celaya, and Bolita on the basis of the morphological data. Consequently, for some of the accessions, hybridizations provide a clue that can further be used to explain the associations among the Mexican races of maize.

Communicated by A. Charcosset

J. C. Reif · J. Muminović · M. Frisch · A. E. Melchinger (✉)  
Institute of Plant Breeding, Seed Science, and Population Genetics,  
University of Hohenheim, 70593 Stuttgart, Germany  
E-mail: melchinger@uni-hohenheim.de

M. L. Warburton · J. Crossa · S. Taba  
CIMMYT International, Aptdo. 370, Postal 6-641,  
06600 Mexico, DF Mexico

X. C. Xia  
Institute of Crop Breeding and Cultivation,  
Chinese Academy of Agricultural Sciences,  
Zhongguancun South Street No. 12, 100081 Beijing, China

D. A. Hoisington  
ICRISAT, Patancheru, Hyderabad 502324 Andhra Pradesh, India

M. Bohn  
University of Illinois, Urbana, IL 61801, USA

### Introduction

Genetic diversity in maize is a valuable natural resource and plays a key role in future breeding progress. Maize was domesticated about 9,000 years ago in Mexico from a form of tropical teosinte, *Zea mays* ssp. *parviglumis* (Beadle 1939; Doebley 2004). Molecular analyses suggest a single domestication event (Matsuoka et al. 2002) that resulted in a loss of diversity in maize (Eyre-Walker et al. 1998; Vigouroux et al. 2002). Following domestication, mutation generated new alleles, while recombination created novel allele combinations. Furthermore, post domestication gene flow from teosinte presumably increased the existing genetic base of maize (Doebley 2004). The genetic variation was subsequently reduced by genetic drift and selection, both natural and artificial by early farmers, which eventually resulted in a large number of races adapted to the specific environmental conditions of their habitats and desired uses by humans.

Understanding the evolutionary history and detailed knowledge of the genetic diversity within and among indigenous races of maize are of interest for fundamental research. Moreover, they are important for the efficient management of the conservation of races and their optimum utilization in plant breeding. A large-scale collection, description, and characterization of the maize germplasm complexes from Mexico began in 1943 with the initiation of the cooperative Mexican Government—Rockefeller Foundation agricultural program. Twenty-five races of maize from Mexico were initially identified and relationships were suggested mainly on the basis of morphological traits as well as archaeological data (Wellhausen et al. 1952). Some of the 25 races were assumed to be either ancient in Mexico (Ancient Indigenous) or other regions (Pre-Columbian Exotic). It was hypothesized that these ancient races served as parents of various pre-Columbian hybrid races (Prehistoric Mestizos). The races of the fourth group, Modern Incipient races, were created after the Conquest by crossing the Ancient Indigenous, Pre-Columbian, and

Prehistoric Mestizos germplasm. They have not yet reached a state of racial stability. Although these hypotheses are very appealing, they have never been examined with molecular data.

Independent of the classification and pedigree structure hypothesized by Wellhausen et al. (1952), other authors grouped Mexican races according to their similarities analyzed by morphological data, genotype by environment interactions, chromosome constitution, and isozyme markers (for review see Goodman and Brown 1988; Sanchez et al. 2000). In all studies with isozyme markers, the races did not cluster into well-defined complexes as suggested by morphological data, but rather formed a continuum with several outliers (Doebley et al. 1985; Sanchez et al. 2000). These differences among the population structure, based on morphological and molecular data were also described in maize landraces from the central valleys of Oaxaca in Mexico (Pressoir and Berthaud 2004). The authors explained these differences by considerable amount of gene flow among populations and a strong divergent selection for the quantitative traits under consideration.

We applied simple sequence repeat (SSR) markers to characterize 25 accessions of races of maize from

Mexico. Our objectives were to (1) study the molecular genetic diversity within and among these accessions, and (2) examine their relationships as assumed previously on the basis of morphological data.

## Materials and methods

### Genetic materials

We examined 25 accessions of 24 races of maize from Mexico (Table 1) that were described in detail by Wellhausen et al. (1952). Race Chalqueño was represented by two accessions. The 25 accessions have been multiplied an unknown number of generations since they were procured from the farmers' field.

### SSR analyses

In total, 497 plants were genotyped in the Applied Biotechnology Center at the International Maize and Wheat Improvement Center (CIMMYT) with a set of 25 SSRs covering the entire maize genome. Marker names,

**Table 1** Race names, collection, codes, number of individuals per accession (No. ind.), altitude of the regions where the accession was grown (Alt.), as well as classification of accessions according to six previous studies (see footnote)

Race name	Collection	Code	No. ind.	Alt. (m)	A <sup>a</sup>	B <sup>b</sup>	C <sup>c</sup>	D <sup>d</sup>	E <sup>e</sup>	F <sup>f</sup>	G <sup>g</sup>
Arrocillo Amarillo	Puebla 262	AA	21	1,600–2,000	4	OL <sup>h</sup>	1	1	1	1	1
Bolita	Oaxa 180	BO	20	900–1,500	2	2	2	3	3	3	2
Cacahuacintle	México 212	CA	21	2,200–2,800	1	1	1	OL <sup>h</sup>	1	1	1
Celaya	Guanajuato 101	CY	20	1,200–1,800	2	2	2	4	2	2	2
Chalqueño	Puebla 260	CQ-1	21	1,800–2,300	1	1	1	1	1	1	1
	México 208	CQ-2	19	1,800–2,300	1	1	1	1	1	1	1
Chapalote	Sinaloa 2	CH	19	100–600	4	2	3	2	2	2	3
Comiteco	Chiapas 235	CM	19	1,100–1,500	2	2	2	OL <sup>h</sup>	2	2	2
Cónico	México 207	CO	21	2,200–2,800	1	1	1	1	1	1	1
Cónico Norteño	Guanajuato 102	CN	18	1,600–2,100	1	1	1	1	1	1	1
Harinoso de Ocho	Nayarit 24	HO	21	100	3	3	3	3	3	3	3
Jala	Nayarit 72	JA	13	1,000	2	2	2	OL <sup>h</sup>	3	3	2
Máiz Dulce	Guanajuato 100	MD	21	1,000–1,500	5	1	1	OL <sup>h</sup>	1	1	3
Nal-Tel	Yucatán 7	NT	21	100	4	1	2	2	2	2	2
Olotillo Blanco	Chiapas 237	OT	21	300–700	3	2	2	2	2	2	2
Olotón	Chiapas 238	OL	16	2,000–2,400	1	OL <sup>h</sup>	2	5	2	2	2
Palomero Toluquen~o	México 211	PT	21	2,200–2,800	1	1	1	1	1	1	1
Pepitilla	Guerrero 221	PE	20	100–1,700	1	2	1	1	2	2	2
Reventador	Nayarit 39	RE	21	0–1,500	4	3	3	3	3	2	3
Tabloncillo	Jalisco 263	TA	21	0–1,500	3	2	3	3	3	3	2
Tehua	Chiapas 234	TE	20	600–1,000	2	2	2	OL <sup>h</sup>	2	2	2
Tepecintle	Guatemala 234	TP	21	0–600	2	2	2	2	2	2	2
Tuxpeño	V-520-C	TU	20	0–500	2	2	2	4	2	2	2
Zapalote Chico	Oaxa 179	ZC	20	100	2	2	2	OL <sup>h</sup>	2	2	2
Zapalote Grande	Chiapas 236	ZG	21	100–600	2	2	2	2	2	2	2

<sup>a</sup>Goodman (1972) classification is based on morphological data

<sup>b</sup>Doebley et al. (1985) Classification is based on isozyme data

<sup>c</sup>Goodman and Brown (1988) classification is based on a combination of morphological, cytological, and isozyme data

<sup>d</sup>Bretting and Goodman (1989) classification is based on chromosome knob data

<sup>e</sup>Sanchez and Goodman (1992) classification is based on morphological data

<sup>f</sup>Sanchez et al. (2000) classification is based on a combination of morphological and isozyme data

<sup>g</sup>Results of our study assuming three groups

<sup>h</sup>Race could not be clustered to a specific group

BIN positions, and allele size ranges are published as supporting information at <http://www.uni-hohenheim.de/~jochreif/Landraces/SSRs.htm>. Genotyping protocols are described in detail by Reif et al. (2004). Briefly, DNA was extracted by the CTAB method and the SSR regions were amplified by PCR with fluorescent-labeled primers. PCR products were size-separated on an ABI Prism 377 DNA Sequencer (Perkin Elmer Biotechnologies, Foster City, CA, USA) equipped with the GENESCAN Software package, and then classified to specific alleles using the GENESCAN and GENOTYPER Software packages (Perkin Elmer Biotechnologies).

### Statistical analyses

We determined the number of alleles per locus for the entire set of 497 individuals, as well as for each accession separately, and examined the presence of accession-specific alleles (further referred to as unique alleles). In addition, the total gene diversity ( $H_T$ ) across all accessions and the gene diversity between individuals within each accession ( $H_S$ ) were calculated according to Nei (1987). Differences between  $H_S$  values were tested for significance by a Friedman rank sum test (Hollander and Wolfe 1973). The coefficient of gene differentiation ( $G_{ST}$ ) was calculated according to Nei (1987). The  $G_{ST}$  is the relative differentiation of the accessions. The fixation index  $F_{IS}$  for each marker was estimated separately for each accession according to Nei (1987) as one minus the observed heterozygosity divided by the expected heterozygosity under Hardy–Weinberg equilibrium.

The modified Rogers' distance (MRD) between two accessions or individuals was calculated according to Wright (1978). Standard errors of MRD were estimated by a bootstrap procedure with resampling across markers and individuals. Associations among operational taxonomic units (accessions or individuals) were revealed by principal coordinate analyses (PCoA) (Gower 1966) based on MRD values. To test the hypothetical admixture of accessions proposed by Wellhausen et al. (1952), we calculated the expected allele frequencies of the hybrid races Bolita, Celaya, Chalqueño, Comiteco, Cónico, Cónico Norteño, Jala, Tabloncillo, Tuxpeño, Zapalote Chico, and Zapalote Grande on the basis of the observed allele frequencies of their assumed parental populations. The MRDs between the hybrid accessions were calculated on the basis of (1) expected and (2) observed allele frequencies. The correlation between both MRD matrices was tested with a Mantel test (Legendre and Legendre 1998) excluding the parental populations. In addition, the allele frequencies of the hybrid accessions Bolita, Celaya, Chalqueño, Comiteco, Cónico, Cónico Norteño, Jala, Tabloncillo, Tuxpeño, Zapalote Chico, and Zapalote Grande were estimated based on the observed allele frequencies of randomly chosen parental populations. The correlation between both MRD matrices based on the (1) randomly

chosen parental populations and (2) observed allele frequencies was calculated. This procedure was repeated 100,000 times. We determined the number of runs showing a correlation coefficient higher than the one between the observed MRD values and those expected on the basis of morphological analysis (Wellhausen et al. 1952).

In addition, genetic relationships among individual genotypes from the accessions were analyzed with a model-based clustering approach using the Software package STRUCTURE (Pritchard et al. 2000). The number of clusters  $k$  was varied from 3 to 25 and five runs of STRUCTURE were performed per  $k$ , setting the burn-in time to 100,000 and replication number to 1,000,000. Among the five runs per  $k$ , the one with the highest maximum likelihood was used to assign individual genotypes to different clusters. All other statistical analyses were carried out with the Software Plabsoft (Maurer et al. 2004), which is implemented as an extension of the statistical Software R (R Development Core Team 2004).

---

## Results

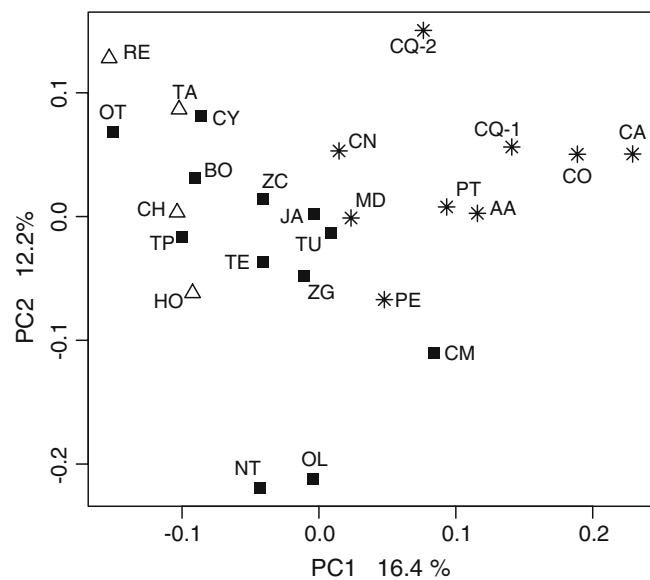
The total number of alleles detected in the 497 genotypes for the 25 SSR loci was 196, with the number of alleles per locus ranging from 4 to 14. All 25 marker loci analyzed were polymorphic across the 497 individuals. Among the 2,425 data points (497 genotypes by 25 SSRs), 10% were either missing data, true null alleles, or failed PCR amplifications. The mean number of alleles within the races was 3.44, with a range from 2.76 (Reventador) to 4.00 (Tabloncillo and Zapalote Chico) (Table 2). An average of 1.32 unique alleles was observed across the 25 accessions. Significant differences ( $P < 0.01$ ) were found among gene diversity values within races ( $H_S$ ). The  $H_S$  was lowest in Tuxpeño (0.41) and highest in Cónico Norteño and Tabloncillo (0.55), with a mean of 0.48. The  $G_{ST}$  between races averaged 0.21. The  $F_{IS}$  values ranged from 0.09 (Cónico) to 0.47 (Chalqueño) (Table 2).

Values of MRD between pairs of accessions ranged from 0.22 (Celaya  $\times$  Tabloncillo) to 0.49 (Cacahuacintle  $\times$  Reventador; Nal–Tel  $\times$  Reventador), with a mean of 0.35 and an average standard error of 0.04. Differences between estimated MRD values were significant ( $P < 0.01$ ) (<http://www.uni-hohenheim.de/~jochreif/Landraces/Distances.htm>). A significant correlation ( $r = 0.69$ ,  $P < 0.001$ ) was found between the MRD values estimated from the expected and observed allele frequencies of the hybrid accessions. For 34% of the runs based on randomly selected parental populations, the correlation coefficient was higher than 0.69, which was observed between the MRD values estimated from the expected and observed allele frequencies of the hybrid accessions. In the PCoA based on the MRD estimates of all accessions, the first two principal coordinates (PC) explained a total of 28.6% of the

**Table 2** Genetic diversity within the 25 accessions of maize from Mexico, as revealed by 25 SSRs

Race name	Code	Average no. of alleles per locus	No. of unique alleles	Gene diversity ( $H_S$ )	Fixation index ( $F_{IS}$ )
Arrocillo Amarillo	AA	3.76	0	0.50	0.31
Bolita	BO	3.72	0	0.52	0.24
Cacahuacintle	CA	3.20	4	0.43	0.27
Celaya	CY	3.64	0	0.51	0.20
Chalqueño	CQ-1	3.36	1	0.45	0.47
	CQ-2	3.60	3	0.48	0.15
Chapalote	CH	3.00	0	0.45	0.29
Comiteco	CM	3.20	0	0.48	0.32
Cónico	CO	3.92	3	0.49	0.09
Cónico Norteño	CN	3.68	1	0.55	0.26
Harinoso de Ocho	HO	2.92	0	0.43	0.22
Jala	JA	3.44	2	0.52	0.23
Máiz Dulce	MD	3.56	1	0.48	0.25
Nal-Tel	NT	3.28	1	0.42	0.11
Olotillo Blanco	OT	3.64	0	0.51	0.19
Olotón	OL	3.20	3	0.53	0.23
Palomero Toluqueño	PT	3.24	1	0.46	0.18
Pepitilla	PE	3.24	1	0.42	0.30
Reventador	RE	2.76	0	0.45	0.21
Tabloncillo	TA	4.00	4	0.55	0.22
Tehua	TE	3.64	1	0.50	0.21
Tepecintle	TP	3.40	0	0.51	0.24
Tuxpeño	TU	3.04	2	0.41	0.15
Zapalote Chico	ZC	4.00	4	0.53	0.31
Zapalote Grande	ZG	3.68	1	0.47	0.38
<i>Average/total</i>		<i>7.84</i>	<i>1.32</i>	<i>0.61</i>	<i>0.24</i>

molecular variance (Fig. 1). The two collections of Chalqueño (CQ-1 and CQ-2) clustered closely together with each other and with Cacahuacintle, Comiteco, Cónico, Palomero Toluqueño, Arrocillo Amarillo, and



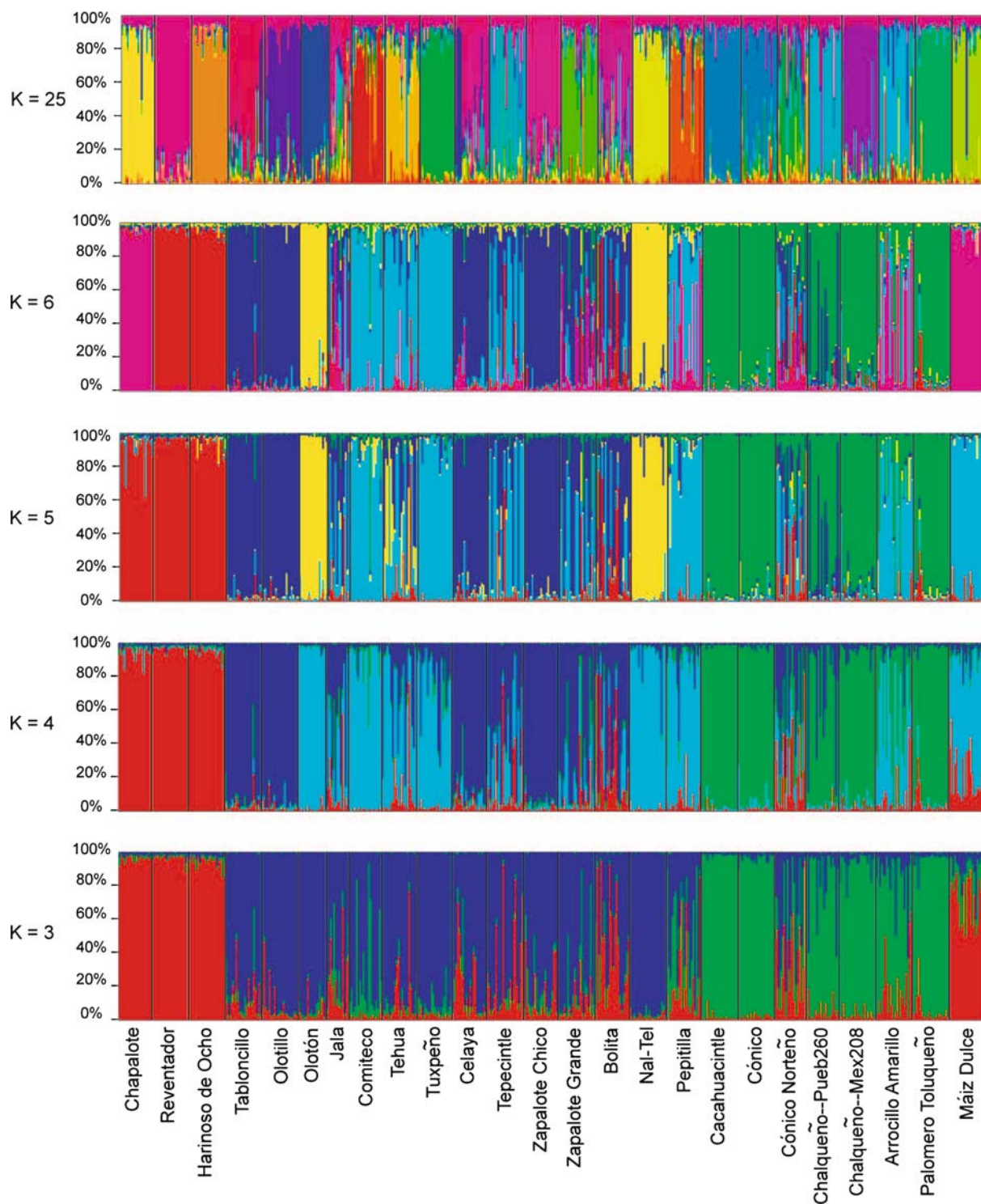
**Fig. 1** Associations among 25 accessions of maize from Mexico revealed by principal coordinate analysis based on the modified Rogers' distance determined from 25 SSR markers. Symbols reflect the classification suggested by Goodman and Brown (1988): high-elevation Mexican pyramidal races (*asterisks*), medium to low elevation races (*filled square*), and long and narrow-eared races typical of Northwestern Mexico (*triangle*)

Cónico Norteño, but were clearly separated from the other accessions by PC1. Nal-Tel and Olotón were separated from the other accessions by PC2.

By setting  $k = 25$  in the model-based clustering, 86% of the individual genotypes were grouped into their respective accessions (Fig. 2). Individuals of races Cónico and Cacahuacintle were not separated, and formed a joint cluster. For  $k = 3$ , the following three clusters were inferred: (1) Chapalote, Reventador, Harinoso de Ocho, and Máiz Dulce, (2) Tabloncillo, Olotillo Blanco, Olotón, Jala, Comiteco, Tehua, Tuxpeño, Celaya, Tepecintle, Zapalote Chico, Zapalote Grande, Bolita, Nal-Tel, and Pepitilla, and (3) Cacahuacintle, Cónico, Chalqueño, Arrocillo Amarillo, and Palomero Toluqueño. Cónico Norteño could not be clustered into a specific group, but was intermediate between all three groups. For  $k = 4$  and 5, the second cluster was split into further subgroups.

## Discussion

Wellhausen et al. (1952) suggested that it was possible to describe in detail the relationships among races of maize from Mexico, as many of them are thought to be hybrids of still-existing races (Fig. 3). The proposed relationships among the races were based on the (1) observation of intermediate morphological traits of a race considering its possible parents and (2) evidence obtained by selfing the races and the observation of mainly two types of progenies resembling their putative parents. As only some of the proposed hybrids

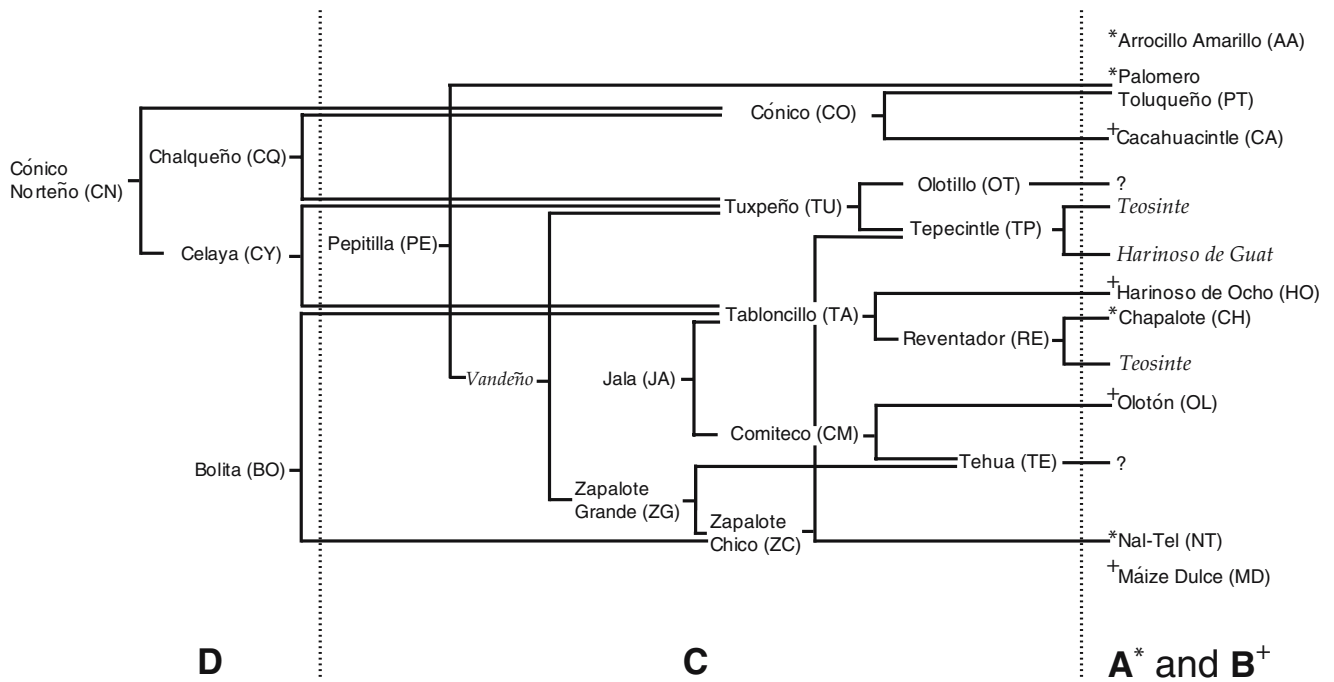


**Fig. 2** Estimated population structure of 497 individuals from 25 accessions of maize from Mexico. Each individual is represented by a thin vertical line, which is partitioned into  $k$  colored segments that represent the individual estimated membership to the  $k$  clusters

were supported by experimental data (Wellhausen et al. 1952) and because of inconsistencies between the suggested grouping of the races based on morphological and molecular data (Table 1), we employed SSR markers to analyze the associations among races of maize from Mexico.

#### Molecular diversity within accessions

We observed a higher total number of alleles per locus (7.84) than reported in previous SSR studies in maize. Analyzing 461 plants representing a diverse array of US germplasm, Labate et al. (2003) found an average of 6.5



**Fig. 3** Relationships (progenies—left side, progenitors—right side) of *A* Ancient Indigenous, *B* Pre-Columbian Exotic, *C* Prehistoric Mestizos, and *D* Modern Incipient races of maize suggested by Wellhausen et al. (1952). Races in italics are not included in this study

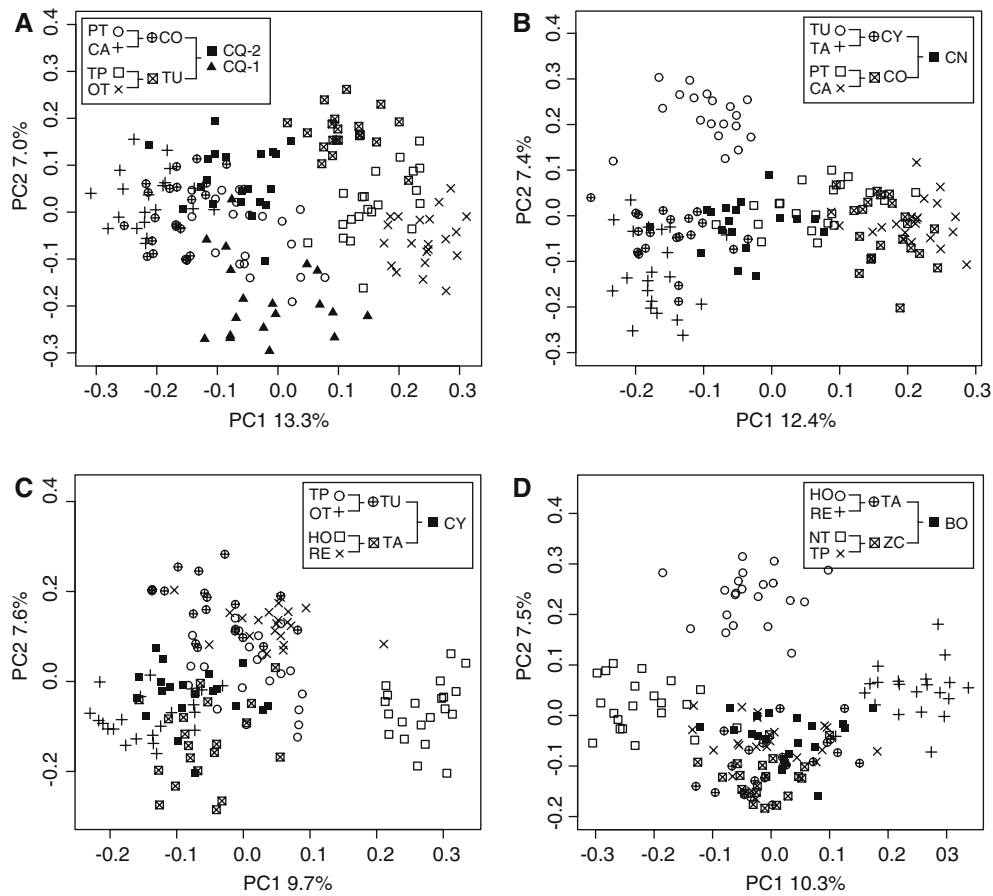
alleles per locus. Reif et al. (2005) obtained a mean of 5.9 alleles per locus for 150 maize plants from five European open-pollinated populations. The total gene diversity across all accessions (0.61) in our study was also higher than that reported by Reif et al. (2005) (0.53). The  $G_{ST}$  between accessions was in accordance with values reported by Reif et al. (2005), who fingerprinted Central European germplasm with SSRs, but lower than results on RFLP diversity in European and American material (Rebourg et al. 2003). The diversity within accessions (Table 2) was similar to the gene diversity values of US populations (Labate et al. 2003) but was significantly higher than gene diversity in the Central European germplasm (Reif et al. 2005). In conclusion, the high total number of alleles per locus and gene diversity values in our study (Table 2) confirmed the broad genetic base of the accessions of maize from Mexico. This was expected, because Mexico is regarded as the center of origin of maize (Matsuoka et al. 2002).

We calculated the correlations between  $H_S$  values of our SSR and previous isozyme studies to test for the effect of different marker systems and alternative sampling of accessions of races. The correlations were low and amounted to  $r = 0.20$  (isozyme data of Doebley et al. 1985) and  $r = 0.17$  (isozyme data of Sanchez et al. 2000). These results indicate that the genetic diversity estimates within races depend on several factors such as the (1) marker system applied, (2) markers chosen for the particular survey, (3) accessions selected for the study, and (4) individuals randomly chosen to represent a certain accession of a race. While isozymes are based on polymorphisms of enzymes, SSRs amplify DNA length polymorphisms arbitrarily distributed through-

out the genome in coding and noncoding regions. Nevertheless, more than 60% of the SSRs in our study were derived from ESTs or located closely to genes. Thus, the difference between SSRs and isozymes cannot be explained entirely by selective neutrality of SSRs. The moderate correlation between gene diversity values of both isozyme studies ( $r = 0.47$ ) and the MRD of 0.33 between the two collections of Chalqueño in our study (Fig. 1) suggest that the genetic diversity within races depends not only on the marker system and markers chosen for the particular study, but also on the choice of the collections and individual genotypes analyzed. The high level of SSR variation detected among accessions of a single race suggests that sampling less individuals but more accessions representing a certain race would likely valorize the accuracy in estimating the allele frequency of the race. Consequently, for future studies investigating the genetic diversity within races but also examining the grouping of Mexican races of maize it would be prudent to include more accessions per race.

#### Verification of complex pedigrees

A highly significant correlation ( $r = 0.69$ ,  $P < 0.001$ ) between the MRD values estimated from the observed and expected allele frequencies of the hybrid accessions proposed by Wellhausen et al. (1952) as well as the 34% of pedigrees based on randomly selected parental populations that resulted in a correlation higher than  $r > 0.69$  may both be interpreted as an indicator that the complete pedigree of the accessions is not true. For some specific triplets of the Modern Incipient hybrids, the



**Fig. 4** Relationships of modern incipient races **a** Chalqueño (*CQ*), **b** Cónico Norteño (*CN*), **c** Celaya (*CY*), **d** Bolita (*BO*), with their hypothetical parents revealed by four principal coordinate analyses based on modified Rogers' distance between individual genotypes. Codes of races are according to Table 1

relationships proposed by Wellhausen et al. (1952) seem likely (Fig. 4). Under the simplifying assumption that (1) both parental populations contribute equally to a hybrid accession and (2) mutation, drift, and selection are absent, a hybrid accession is placed in the PCoA in between its parental populations. In accordance with this expectation, the PCoA of individual genotypes of accessions involved in the pedigrees of Chalqueño, Cónico Norteño, Celaya, and Bolita revealed that all four Modern Incipient hybrids were positioned between their putative parental accessions, except for Chalqueño-México208 (Fig. 4). This corroborates the following results reported previously: (a) Chalqueño has morphological characteristics that are intermediate between Cónico and Tuxpeño. Furthermore, when Chalqueño is inbred, it breaks up into types with a Cónico-like form at one extreme and a Tuxpeño-like form at the other (Wellhausen et al. 1952, p. 169). (b) The morphological characteristics of Cónico Norteño are either similar to Cónico or Celaya. In addition, crossing Cónico with Celaya yielded a population very similar to Cónico Norteño (Wellhausen et al. 1952, p. 183). (c) The individuals of Celaya, one of the most productive agricultural races of Mexico (Crossa et al. 1990), are either alike

to Tabloncillo or Tuxpeño or are intermediate between them in their morphological characteristics (Wellhausen et al. 1952, p. 174). (d) The morphological characteristics of Bolita are intermediate between Tabloncillo and Zapalote Chico; when Bolita is inbred, it breaks up into types with a Tabloncillo-like form at one extreme and a Zapalote Chico-like form as the other (Wellhausen et al. 1952, p. 188). Summarizing, the results of the molecular analysis of the four Modern Incipient hybrids provide evidence for the proposed parental accessions of Chalqueño, Cónico Norteño, Celaya, and Bolita suggested by Wellhausen et al. (1952) on the basis of morphological data. Nevertheless, the findings must be interpreted with caution due to the restricted number of accessions sampled within highly heterogeneous races and the underlying simplifying assumption of an equal parental contribution to the hybrids. In addition, the heterogeneous races at the level of SSRs point to substantial gene flow among races, despite having certain morphological homogeneity with respect to the traits of interest to farmers (Pressoir and Berthaud 2004). The verification or falsification of the relationships suggested by Wellhausen et al. (1952) requires additional efforts fingerprinting more accessions within races.

In most cases, the postulated hybrids of the Prehistoric Mestizos were not positioned midway between their parents, except for Cónico (Fig. 4). For some triplets, progenies overlapped with one of its parents, but were far from the second proposed parent. On one hand, this result can be explained by the possible genetic changes due to selection, drift, and mutation confirming the chronological separation of Modern from Prehistoric hybrid accessions. On the other, the observed grouping of the Prehistoric Mestizos could also be due to a wrongly proposed pedigree, deviations from an equal parental contribution to the hybrids, or a nonchronological but geographical grouping of the races. A Mantel test between the distances on the basis of the altitude of the regions, where the accession was grown, and the MRD among the accessions resulted in a moderate but significant correlation ( $r = 0.37$ ,  $P < 0.01$ ). This may be interpreted as an indicator that the grouping of accessions can be explained based on the altitude where the race was grown as already suggested by Doebley et al. (1985). Further, genotyping of conserved genomic regions could help to elucidate the relationships of the Mexican races of maize in more detail.

#### Molecular variation among accessions

The grouping of accessions into distinct complexes is complicated by the fundamental problem of not knowing how many clusters to assume. Because the accessions deviated from the Hardy–Weinberg equilibrium with most loci indicating a tendency toward an excess of homozygosity within accessions, we followed the suggestion of Pritchard et al. (2000) and did not run STRUCTURE to determine the optimal number of race complexes. An excess of homozygosity could lead to an overestimation of the number of clusters. The observed deviation from Hardy–Weinberg equilibrium was in close agreement with previous results in SSR studies with open-pollinated populations, and can be explained by both technical limitations and genuine genetic causes (for a detailed discussion see Reif et al. 2005).

Most previous studies analyzing the accessions of maize from Mexico revealed three main racial complexes with several subgroups: (1) the high-elevation Mexican pyramidal, (2) the medium- to low elevation, and (3) the long and narrow-eared races typical of northwestern Mexico (for review see Goodman and Brown 1988). Therefore, we assumed initially  $k = 3$  clusters in the STRUCTURE runs. The groupings observed in our study were only partly in agreement with previous investigations (Fig. 2; Table 1).

The accessions Arrocillo Amarillo, Cacahuacintle, Cónico, Chalqueño, and Palomero Toluqueño formed the high-elevation Mexican pyramidal group (Figs. 1, 2), which is in accordance with results of most previous studies (Table 1). The high-elevation races have red hairy leaf sheaths, sparsely branched tassels, and conical-shaped ears (Wellhausen et al. 1952) and they show

good *per se* performance when cultivated at high-elevation environments (Crossa et al. 1990). Cónico Norteño did not cluster into the high-elevation group, which is contrast to the studies of Goodman and Brown (1988) and Sanchez et al. (2000). Considering the complex pedigree of Cónico Norteño proposed by Wellhausen (Fig. 3), the intermediate grouping between the three racial complexes is not surprising.

The medium- to low elevation group contained 14 accessions (Tabloncillo, Olotillo Blanco, Olotón, Jala, Comiteco, Tehua, Tuxpeño, Celaya, Tepecintle, Zapalote Chico, Zapalote Grande, Bolita, Nal–Tel, and Pepitilla) (Fig. 2). This cluster is in close agreement with results reported by Crossa et al. (1990) and Goodman and Brown (1988), except for the races Pepitilla and Tabloncillo. The grouping of Pepitilla into the medium- to low elevation group would fit into the pedigree suggested by Wellhausen et al. (1952) (Fig. 3). The clustering of Tabloncillo into the medium- to low elevation group can again be explained by its pedigree and close relationship with Celaya (MRD = 0.22) (Fig. 1). Nevertheless, in three out of five STRUCTURE runs, Tabloncillo was grouped into the cluster of long and narrow-eared races and its classification is therefore uncertain.

In consequence, the following accessions formed the group of long and narrow-eared accessions typical of Northwestern Mexico: Chapalote, Reventador, Harinoso de Ocho, and Máiz Dulce. This clustering of the accessions Chapalote, Reventador, and Harinoso de Ocho is in close agreements with results reported by Goodman and Brown (1988) and Crossa et al. (1990). Considering the first two PCs, Máiz Dulce is closely related to Cónico Norteño (Fig. 1), which could be explained by the identical location where both accessions were collected. However, Máiz Dulce is clearly separated from the other accessions by the third PC (data not shown), which underlines its special status. The special status of Máiz Dulce might be explained by non-reciprocal cross-sterility caused by a *Ga1-s/Ga1-s* allele combination at the *Gametophyte-factor-1* locus (Nelson 1993). Thus, Máiz Dulce might be considered as a separate race, belonging to neither group. This conclusion would also fit within the pedigree proposed by Wellhausen et al. (1952), and is supported by the inconsistencies of the grouping of Máiz Dulce in previous studies (Table 1) as well as by the STRUCTURE analysis (Fig. 2). When  $k = 3$ , Máiz Dulce was classified to the long and narrow-eared accessions typical of Northwestern Mexico. With an increasing number of clusters assumed in the STRUCTURE runs from  $k = 4$  to  $k = 5$ , Máiz Dulce became part of the medium- to low elevation accessions. However, with  $k = 6$ , it was again grouped with Chapalote, which belongs to the long and narrow-eared accessions.

The high  $G_{ST}$  value (0.21) and the correct grouping of 86% of the individuals into their respective accessions showed that most accessions could be differentiated from each other with molecular markers (Fig. 2). The exceptions were the accessions Cónico and Cacahuacintle, which formed one cluster when  $k = 25$ ,



as expected from their proposed relationships. Cacahuacintle is assumed to be one of the parents of Cónico.

Summarizing, the accessions of maize from Mexico can be clustered into distinct racial complexes on the basis of the SSR markers. Our results can also be interpreted as an indicator supporting some of the relationships proposed by Wellhausen et al. (1952). Consequently, some of the associations of the Mexican accessions such as those among the four Modern Incipient hybrids could be explained in detail with hybridizations among ancient races.

### Unlocking the untapped allelic variation of the races for breeding

Novel genes and alleles of potentially useful traits can be identified during routine maintenance and systematic screening of collections, or as a spin-off of prebreeding and breeding programs carried out for other purposes. Once a desired trait has been identified, backcrossing can be used to introduce it into the elite breeding germplasm. This approach works well when the trait of interest is controlled by one or a few genes, but not for traits with polygenic inheritance such as grain yield. Hence, for the identification of genes and alleles underlying complex traits of agronomic importance, more sophisticated methods such as association mapping are required (cf. Lynch and Walsh 1997).

As a first step toward association mapping, we propose to adopt the approach suggested by Reif et al. (2005) and decompose the accessions into representative samples of homozygous lines by generating, for example, doubled haploid lines. This offers two advantages: (1) the resulting genotypes are immortal and can be subject to replicated tests for phenotyping as well as genotyping and (2) the genetic load is reduced by eliminating deleterious alleles during the inbreeding process. The lines can then serve as the base population for association mapping. The proposed strategy would facilitate a systematic utilization of the available genetic diversity in accessions of maize.

**Acknowledgments** The molecular marker analyses of this research were supported by funds from the German Federal Ministry for Economic Cooperation and Development (BMZ) Project No. 98.7860.4-001-01. The authors thank Leticia Diaz and Ana Lidia Gomez for their excellent technical assistance. The authors thank three anonymous reviewers for their valuable suggestions.

### References

- Beadle GW (1939) Teosinte and the origin of maize. *J Hered* 30:245–247
- Bretting PK, Goodman MM (1989) Karyological variation in Mesoamerican races of maize and its systematic significance. *Econ Bot* 43:107–124
- Crossa J, Taba S, Wellhausen EJ (1990) Heterotic patterns among Mexican races of maize. *Crop Sci* 30:1182–1190
- Doebley JF (2004) The genetics of maize evolution. *Annu Rev Genet* 38:37–59
- Doebley JF, Goodman MM, Stuber CW (1985) Isozymic variation in the races of maize from Mexico. *Am J Bot* 72:629–639
- Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci USA* 95:4441–4446
- Goodman MM (1972) Distance analysis in biology. *Syst Zool* 21:174–186
- Goodman MM, Brown WL (1988) Races of corn. In: Sprague GF, Dudley JW (eds) *Corn and corn improvement*. 3rd edn. Agron Monogr 18. ASA, CSSA, and SSSA, Madison, WI, pp 39–74
- Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53:325–338
- Hollander M, Wolfe DA (1973) *Nonparametric statistical inference*. Wiley, New York, pp 139–146
- Labate JA, Lamkey KR, Mitchell SH, Kresovich S, Sullivan H, Smith JSC (2003) Molecular and historical aspects of Corn Belt dent diversity. *Crop Sci* 43:80–91
- Legendre P, Legendre L (1998) *Numerical ecology*. 2nd English edn. Elsevier, Amsterdam
- Lynch M, Walsh B (1997) *Genetics and analysis of quantitative traits*. Sinauer Assoc, Sunderland, MA, p 413
- Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez Garcia J, Buckler E, Doebley J (2002) A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci USA* 99:6080–6084
- Maurer HP, Melchinger AE, Frisch M (2004) Plabsoft: Software for simulation and data analysis in plant breeding. XVIIIth EUCARPIA General Congress 2004, Tulln, Austria, pp 359–362
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Nelson OE (1993) The gametophyte factors of maize. In: Freeling M, Walbot V (eds) *The maize handbook*. Springer, Berlin Heidelberg New York
- Pressoir G, Berthaud J (2004) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity* 92:95–101
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- R Development Core Team (2004) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL <http://www.R-project.org>
- Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A (2003) Maize introduction into Europe: the history reviewed in the light of molecular data. *Theor Appl Genet* 106:895–903
- Reif JC, Xia XC, 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
- Reif JC, Hamrit S, Heckenberger M, Schipprack W, Bohn M, Melchinger AE (2005) Genetic structure and diversity of European flint maize populations determined with SSR analyzes of individuals and bulks. *Theor Appl Genet* 111:906–913
- Sanchez GJJ, Goodman MM (1992) Relationships among Mexican races of maize. *Econ Bot* 46:72–85
- Sanchez GJJ, Goodman MM, Stuber CW (2000) Isozymatic and morphological diversity in the races of maize of Mexico. *Econ Bot* 54:43–59
- Vigouroux Y, McMullen M, Hittinger CT, Houchins K, Schulz L, Kresovich S, Matsuoka Y, Doebley J (2002) Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc Natl Acad Sci USA* 99:9650–9655
- Wellhausen EJ, Roberts LM, Hernandez E (in collaboration with Mangelsdorf PC) (1952) *Races of maize in Mexico*. The Bussey Institution, Harvard University Press, Cambridge, MA
- Wright S (1978) *Evolution and genetics of populations*, Vol. IV. The University of Chicago Press, Chicago, IL, p 91