Theor Appl Genet (2009) 119:1053–1067
DOI 10.1007/s00122-009-1108-0

ORIGINAL PAPER

Microsatellite variation in maize landraces from Northwestern Argentina: genetic diversity, population structure and racial affiliations

Verónica V. Lia · Lidia Poggio · Viviana A. Confalonieri

Received: 25 September 2008 / Accepted: 30 June 2009 / Published online: 29 July 2009

Abstract The highland region or Northwestern Argentina (NWA) is one of the southernmost areas of native maize cultivation and constitutes an expansion of the peruvian Andes sphere of influence. To examine the genetic diversity and racial affiliations of the landraces cultivated in this area, 18 microsatellite markers were used to characterize 147 individuals from 6 maize races representative of traditional materials. For the whole data set, a total of 184 alleles were found, with an average of 10.2 alleles per locus. The average gene diversity was 0.571. The observed patterns of genetic differentiation suggest that historical association is probably the main factor in shaping population structure for the landraces studied here. In agreement with morphological and cytogenetic data, Bayesian analysis of NWA landraces revealed the occurrence of three main gene pools. Assessment of racial affiliations using a combined dataset including previous data on American landraces showed a clear relationship between one of these gene pools and typical Andean races, whereas the remaining two gene pools exhibited a closer association to Caribbean accessions and native germplasm from the United States, respectively. These results highlight the importance of integrating regional genetic studies if a deeper understanding of maize diversification and dispersal is to be achieved.

Communicated by A. Charcosset.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-009-1108-0) contains supplementary material, which is available to authorized users.

V. V. Lia · L. Poggio · V. A. Confalonieri
Departamento de Ecología, Genética y Evolución.
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Intendente Güiraldes y Costanera Norte s/n, 4to. Piso, Pabellón II, C1428EHA Ciudad Autónoma de Buenos Aires, Argentina

V. V. Lia · L. Poggio · V. A. Confalonieri
Consejo Nacional de Investigaciones Científicas y Técnicas, Avenida Rivadavia 1917, C1033AAJ Ciudad Autónoma de Buenos Aires, Argentina

V. V. Lia (✉)
Instituto de Biotecnología, CICVyA, INTA, Cautelar, Los Reseros y Las Cabañas s/n, B1686ICG Hurlingham, Buenos Aires, Argentina
e-mail: vlia@cnia.inta.gov.ar; verolia@ege.fcen.uba.ar

Introduction

Maize landraces have long been regarded as a valuable source of alleles to broaden the genetic basis of existing breeding programs. In practice, however, only a limited proportion (ca. 10%) of the estimated 250–300 landraces of maize of the Americas is currently being utilized for the development of new inbred lines (Goodman 1990; Hoisington et al. 1999; Taller and Bernardo 2004; Tarter et al. 2003; Warburton et al. 2002). For many years, this was mainly due to the lack of agronomic evaluation, the scarce integration between morphological and genetic data, and the poor agronomic performance of native materials (Goodman 2005). In an effort to record key agronomic traits and to secure medium and long-term seed storage, a large-scale project of characterization, description and regeneration of Latin American maize landraces (Cooperative Regeneration Project) began in 1993 under the coordination of CIMMYT. In addition to conservation initiatives, a large number of studies have documented the cytogenetic and genetic diversity of maize landraces from the Americas. However, most of the research conducted till date has been performed on a region by region basis, focusing primarily

on the central areas of diversity (i.e., Mexico and the Central Andes), with the more marginal locations being only poorly represented (e.g., Bretting et al. 1987, 1990; Doebley and Goodman 1984; Doebley et al. 1983, 1985, 1986, 1987, 1988; Goodman and Stuber 1983). Three notable exceptions to the fragmentary nature of the above mentioned studies are the cytogenetic surveys of McClintock et al. (1981), the isozyme variation studies of Sanchez et al. (2000) and the SSR analysis conducted by Matsuoka et al. (2002a) on a set of accessions encompassing the entire pre-columbian range of maize distribution.

Although large-scale regional approaches are of fundamental importance for the assessment of genetic variation, a thorough understanding of population dynamics and intra-racial variation is needed since in situ conservation cannot be achieved without more detailed micro-regional studies. In many Latin American countries, maize landraces directly descended from the crops originally grown by natives are still maintained by farmers and cultivated in a traditional manner with little or no input from commercial germplasm. Farming practices play a key role in the distribution of genetic variation, and intensive collecting and surveying of farm households are required to identify the main factors shaping population dynamics at micro-regional levels (Brush and Perales 2007; Perales et al. 2005; Pressoir and Berthaud 2004). While the race concept usually entails an abstract notion involving morphological characteristics and geographical/environmental origin, the term "population" often refers to the small plots managed by local farmers. Throughout this paper, the terms "race" and "landrace" will be used as synonyms to define a taxonomic entity delimited by a set of morphological attributes; and the term "population" will refer to an ensemble of individuals that are managed by a single farmer and that share similar morphological characteristics (i.e., can be assigned to the same landrace).

The Northern region of Argentina constitutes one of the southernmost areas of native maize cultivation, and in spite of its marginal location, it harbors ca. 70 different landraces (Cámara Hernández and Miante Alzogaray 1997, Cámara Hernández personal communication). According to Horovitz (1935), two vast agricultural systems can be distinguished within this area: (1) the highland region or Northwestern Argentina (NWA), an expansion of the peruvian Andes sphere of influence; and (2) the mesopotamic and Chaco plains or Northeastern Argentina (NEA), a lowland area thought to be more closely affiliated with landraces from Brazil and Paraguay. Genetic diversity estimates for these areas have been inferred on the basis of a restricted number of accessions (Matsuoka et al. 2002b; Sanchez et al. 2000) and only one of them allowed evaluation at the population level (Bracco et al. 2009). The extensive microsatellite survey conducted by Matsuoka et al. (2002a) enables comparison of microsatellite alleles throughout the Americas, and provides a framework for the analysis of genetic variation and racial affinities. Here, we present an analysis of microsatellite variation of maize landraces cultivated by traditional farmers in mid- to high-altitude locations of Northwestern Argentina with the aim of assessing population dynamics at a micro-regional scale. In addition, we use a model-based bayesian approach to establish racial affiliations for these landraces.

Materials and methods

Plant material

A total of 144 individuals representing 6 maize landraces (8 populations) from Northwestern Argentina (NWA) were included in this study. The selected landraces constitute the dominant types detected for this area at both the household and community levels, with a total of 22 landraces having been described for the region (Cámara Hernández, personal communication). The races Altiplano, Amarillo Grande, Amarillo Chico, Blanco, and Pisingallo are representative of traditional materials, while Orgullo Cuarenton is best categorized as an incipient modern race derived from hybridization between native germplasm and improved varieties developed in Argentina in the mid 1960s. Sampled populations were distributed spanning an altitudinal distance of approximately 2,690 m (linear distance 649 km) and were maintained by local farmers, via open-pollination, with no input from commercial germplasm. The sampling scheme was based on a relatively intensive collecting effort within each population in order to obtain more accurate estimates of allelic frequencies and genetic differentiation among farm households. Collections were performed directly from farmers' fields during 1994–1995 and preserved following standard procedures at the Laboratorio de Recursos Genéticos Vegetales N.I. Vavilov (VAV), Universidad de Buenos Aires, under the supervision of Ing. Julián Cámara Hernández. Seed germination and DNA extractions were conducted in 2000 as part of V.V. Lia doctoral research work using materials from the original collections. Voucher specimens, collection sites, racial identification, main phenotypic characteristics, cultivar altitudes, and sample sizes are given in Table 1.

A set of 18 inbred lines from M.M. Goodman Collection, North Carolina State University (A632, B164, B77, B37, Oh43, Tx303, CM37, CM105, C103, I29, K55, Va102, L317, DE811, EPI, SA24, HP301 and W64a) were used as reference to establish allele size equivalences with previously reported data (Matsuoka et al. 2002a). Lines were chosen to represent the main clusters found in the analysis of 260 temperate, tropical, and subtropical inbreds conducted by Liu et al. (2003).

Table 1 Maize landraces included in the present study

Population ID	Voucher[a]	Collection site	Landrace[b]	Main phenotypic characteristics[c]	Cultivar altitude (m.a.s.l.)	Sample size
6473	VAV 6473	Susques, Dpto. Susques, Jujuy, Argentina	Altiplano	Small elliptical ears. 8–10 rows of kernels. Assortment composed of a diverse array of materials adapted to high-altitude environments, generally having yellow endosperm. High precocity	3600	18
					3000	14
6167	VAV 6167	El Puesto, Dpto. Santa Victoria, Salta, Argentina				
6485	VAV 6485	Colonia San José, Dpto. Tilcara, Jujuy, Argentina	Blanco	Large to medium-sized cylindrical ears. 12–14 rows of kernels. Semident, medium-sized floury kernels with flint lateral sides. Colorless pericarp and aleurone. Low precocity	2670	13
6480	VAV 6480	La Ciénaga de Pumamarca, Dpto. Tumbaya, Jujuy, Argentina	Amarillo Grande	Medium-sized ears. 8 and generally 10 rows of kernels. Rounded and slightly dented kernels with flint lateral sides and floury central part. Yellow endosperm and colorless aleurone. Medium precocity	2420	25
6484	VAV 6484	Tumbaya, Dpto. Tumbaya, Jujuy, Argentina	Amarillo Chico	Small ears. 8 rows of kernels. Rounded or slightly pointed kernels with a flint layer at the periphery and floury in the inner portion. Yellow endosperm and colourless aleurone. High precocity	2010	20
					1690	14
6476	VAV 6476	Termas de Reyes, Dpto. Capital, Jujuy, Argentina				
6313	VAV 6313	Los Toldillos, Dpto. Ambato, Catamarca, Argentina	Pisingallo	Medium-sized conical ears. 14-20 irregular rows of kernels. Medium to small-sized, pointed and pop kernels. Colorless pericarp, aleurone endosperm. Medium precocity	1600	16
6482	VAV 6482	La Candelaria, Dpto. Candelaria, Salta, Argentina	Orgullo Cuarentón	Large cylindrical ears, 16 rows of kernels, semident flint kernels. Yellow endosperm. Low precocity	910	24

M.a.s.l meters above sea level

[a] Voucher specimens are deposited at the "Laboratorio de Recursos Genéticos Vegetales N·I. Vavilov" (VAV), Faculty of Agronomy, University of Buenos Aires. [b]Taxonomic identification performed by Cámara Hernández based on morphological criteria. [c]Data from Torregrosa et al. (1980) and Cámara Hernández and Miante Alzogaray (unpublished)

Microsatellite typing

Eighteen microsatellite loci or SSRs were selected from a preliminary survey of 39, with only those loci with unambiguous interpretation being used for this analysis (bnlg1866, phi037, bnlg1018, phi127, phi029, bnlg1182, nc135, bnlg252, bnlg1700, bnlg1287, bnlg1165, bnlg1732, bnlg1070, phi069, phi121, bnlg1209, phi059, bnlg1360). Each chromosome pair was represented by two unlinked loci, except for pairs 8 and 9, in which a single locus was studied. Genotypic data from populations 6167, 6313, 6476, 6480, 6482, 6484, and 6485 were previously used by Lia et al. (2007) to investigate the evolutionary forces involved in the maintenance of an altitudinal cline of B chromosomes. Genotypic characterization of population 6473 represents a new addition to the previous data. Details of the used loci, including chromosomal location, putative repeat motifs and associated primers, along with DNA extraction, amplification, electrophoresis, and silver-staining procedures are given in Lia et al. (2007).

Data analysis

Allele frequencies, mean number of alleles per locus (A), allelic richness (R_S) (El Mousadik and Petit 1996) and gene diversity (H_e) (Nei 1987) were computed using Fstat 2.9.3.2 software (Goudet 1995; Goudet 2001). The presence of population-specific alleles (hereafter referred to as private alleles, i.e., alleles present in only one population and absent in the others) was examined for each population. Differences in H_e and R_S among populations were tested for significance by a Kruskal–Wallis test using STATISTICA

(Stat Soft Inc. 1993). Post hoc pairwise comparisons were performed using the Wilcoxon rank–sum test.

Departures from Hardy–Weinberg (HW) proportions at individual loci were tested within each population. Alternative hypotheses of excess or defect of homozygotes were analyzed separately using the multiscore *U* test as implemented in GENEPOP 3.4 (Raymond and Rousset 1995). Estimates of Wright's fixation index F_{IS} (Wright 1978) were obtained according to Weir and Cockerham (1984) using Fstat 2.9.3.2. Significance of F_{IS} was determined using the randomization test implemented in Fstat program.

The adequacy of allele identity (e.g., F_{ST}, Wright 1978) versus allele size-based statistics (e.g., R_{ST}, Slatkin 1995) of population differentiation for the present dataset has been evaluated elsewhere (Lia et al. 2007). As a result of this analysis, F_{ST} was chosen over R_{ST} because of its lower standard error (Gaggiotti et al. 1999; Hardy et al. 2003). Global and pairwise θ, an unbiased estimate of F_{ST} (Weir and Cockerham 1984), were calculated over all loci with Fstat 2.9.3.2 software. The significance of θ values, hereafter referred to as F_{ST}, was tested by permuting genotypes rather than alleles among samples and was corrected for multiple comparisons using Bonferroni procedures at an overall $\alpha = 0.05$ (Rice 1989).

Population structure was also examined using the Bayesian model-based approach of Pritchard et al. (2000) implemented in STRUCTURE 2.1 (http://www.pritch.bsd.uchicago.edu). In this method, a number of populations (*K*) are assumed to be present and to contribute to the genotypes of sampled individuals. The genotype of each individual is a junction of the allele frequencies in these *K* populations (clusters) and the proportion of its genotype drawn from each of the *K* populations. Loci are assumed independent, and each *K* population is assumed to follow HW proportions. The number of clusters evaluated here ranged from 1 to 8. The analysis was performed using ten replicate runs per *K* value, a burn-in period length of 100,000 and a run length of 500,000. No prior information on the origin of individuals was used to define the clusters. All the analyses were run under both the independent and the correlated allele frequency models available with STRUCTURE to test their performance in identifying possible instances of subtle population structure. The first model assumes that the allele frequencies in each population are independent draws from a distribution, while the second model assumes that the frequencies in the different populations are likely to be similar probably due to shared ancestry or migration (Falush et al. 2003). Since no differences were encountered in the results produced by the different models (data not shown), all subsequent analyses were conducted on the outputs obtained from the correlated allele frequency model to avoid format incompatibilities with the C programs described below.

Evaluation of STRUCTURE results was conducted using two ad hoc criteria: (1) the rate of change in the log probability of data between successive *K* values (ΔK) (Evanno et al. 2005); and (2) the across run stability of individual predicted allele frequencies (Camus-Kulandaivelu et al. 2007). Briefly, the second method involves calculation of Euclidean distances among outputs using the C programs proba2.c and distance_output.c (Camus-Kulandaivelu and D. Manicacci, personal communication) and using the corresponding distance matrix to build a Neighbor-joining tree.

The bayesian approach was used on the SSR data from the eight populations studied here and on a combined dataset including data from Matsuoka et al. (2002a) (193 maize accessions representing the entire range of pre-Columbian cultivation). The combined dataset was restricted to the 10 SSR loci shared by both studies (bnlg252, bnlg1018, bnlg1070, bnlg1182, bnlg1209, bnlg1287, bnlg1360, bnlg1732, bnlg1866 and phi037). Due to the differences in genotyping techniques (PAGE and silver-staining vs. fluorescent capillary electrophoresis), equivalences in allele sizes were established by genotyping the 18 inbreds indicated previously in the plant material section, under the same conditions used for the landraces studied here and comparing the resulting genotypes with those obtained by Liu et al. (2003) using the same SSR and typing procedures as in Matsuoka et al. (2002a).

To assess the impact of the reduction in marker number on the estimation of population structure, independent runs of STRUCTURE were performed using solely the 193 accessions of Matsuoka et al. (2002a). The analysis was conducted under the same parameters described above considering: (1) the 78 SSR originally used by Matsuoka et al. (2002a) for population structure analysis; and (2) the 10 SSR in common with this study. The number of clusters evaluated ranged from 1 to 5.

Nei's genetic distances (Nei 1972) between populations were calculated with the software MICROSAT (Minch et al. 1996). Cluster analysis was carried out applying the Neighbor-Joining (NJ) algorithm (Saitou and Nei 1987) implemented in PHYLIP (Felsenstein 1991–2004). Branch support was estimated by bootstrapping (1,000 pseudoreplicates) with MICROSAT. Resulting trees were visualized with TREEVIEW version 1.5.2 (Page 1996).

Results

Genetic diversity

Genotyping of 18 SSR loci in 147 individuals from 8 populations revealed a total of 184 alleles, with the number of alleles per locus ranging from 5 to 31 (mean 10.2). Estimates of H_e, R_S, *A* and number of private alleles for

Table 2 Genetic variability in maize landraces from Northwestern Argentina

Population	H_e	R_S	A	Number of private alleles
6473	0.357	2.43	2.67	0
6167	0.562	4.19	4.5	0
6485	0.612	4.61	4.89	8
6480	0.568	4.35	5.33	3
6484	0.556	3.90	4.5	4
6476	0.661	5.31	5.61	12
6313	0.597	3.98	4.33	8
6482	0.652	5.01	6.11	15

each population are presented in Table 2. The mean number of alleles per locus within populations was 4.74, ranging from 2.67 (6473) to 6.11 (6482). Nei's genetic diversity was high in all populations (mean = 0.571), with 6476 and 6473 exhibiting the highest and lowest values, respectively. A similar pattern was observed for the allelic richness, which varied from 2.43 to 5.31 (mean = 4.22). Significant differences among populations were found in both R_S and H_e [R_S: H(g.l. = 7; N = 144) = 21,11889, $P < 0.004$; H_e: H(g.l. = 7; N = 144) = 17,56395, $P < 0.014$], but only those comparisons involving 6473 remained significant after Bonferroni correction in post hoc pairwise tests. The number of private alleles exhibited a rather heterogeneous distribution among populations. A total of 15 private alleles were detected for population 6482, whereas none were encountered in populations 6473 and 6167.

To allow for a more stringent comparison between the genetic diversity estimates presented here and those reported for a set of landraces encompassing the entire pre-columbian distribution (Matsuoka et al. 2002a) (hereafter referred to as American landraces), and for temperate, tropical, and subtropical maize inbreds (Liu et al. 2003), the mean number of alleles and allelic richness was re-calculated for the three data sets including only the 10 SSR shared by all the studies. For the American landraces and the inbreds, the mean number of alleles across the entire sample decreased from 27.3 (99 SSR) to 20.9 (10 SSR) and from 21.8 (94 SSR) to 17.7 (10 SSR), respectively. The opposite was observed for the populations studied here. The mean number of alleles showed an increase from 10.2 (18 SSR) to 13.1 (10 SSR), probably due to the highly variable nature of the dinucleotide-repeat SSRs from the bnlg series. In sum, the genetic diversity observed in NWA landraces accounts for 63% of that exhibited by American landraces and for 74% of that shown by the inbreds. Similar results were obtained when sample sizes were taken into account by calculating allelic richness (American landraces, R_S: 19.29; inbreds, R_S: 16.06; NWA populations, R_S: 12.92).

Population structure and genetic relationships of NWA landraces

Global estimates of F_{IS} revealed departures from HW proportions in three out of eight populations (6480, 6476, 6482), although only a few loci accounted for these deviations (Table 3). Genetic differences among the eight populations were significant ($F_{ST} = 0.146$; 95% confidence interval: 0.120–0.174). All pairwise comparisons were significantly different from zero and the average F_{ST} was 0.1431 (Table 4). The largest F_{ST} estimates were found in all pairwise comparisons involving population 6473, with the population pairs 6473–6313 ($F_{ST} = 0.3659$) and 6473–6482 ($F_{ST} = 0.2944$) exhibiting the most extreme values. The lowest levels of differentiation corresponded to the comparisons between populations 6480, 6485, 6484, 6476, and 6167 ($F_{ST} = 0.0263$–0.0955). Overall, four distinct groups appear to emerge from the genetic differentiation analysis. One group includes populations 6480, 6485, 6484, 6476, and 6167, and the three remaining groups correspond to populations 6473, 6313, and 6482, respectively.

Bayesian analysis of population structure using the model-based approach of Pritchard et al. (2000) provided support for the existence of genetic structure in our sample and yielded a stratification pattern mostly consistent with the observations derived from the traditional F_{ST} approach. However, the inference of the number of gene pools was not straightforward, given that the log-likelihood values for the data conditional on K, ln(X/K), increased progressively with larger values of K (K = 1 to K = 8) (Fig. 1a). Independent runs produced highly consistent results for K values up to K = 6, with K = 7, and K = 8 showing contrasting assignments for different runs.

At K = 2, the sample is divided into a group consisting of individuals from populations 6473, 6167, 6480, 6484, and 6485, and another group composed of individuals from populations 6482, 6476, and 6313 (Fig. 2). Average membership coefficients for both groups are very high (0.97), except for population 6476 which shows a high proportion of admixed individuals. At K = 3, population 6313 differentiates from 6482 and 6476. At K = 4, populations 6473 and 6313 are clearly differentiated, while populations 6167, 6480, 6484, and 6485 cluster together, like populations 6476 and 6482. However, under this model, population 6476 continues to exhibit considerable levels of admixture. For K = 5 and K = 6, the assignment of individuals into the new clusters did not reveal any new groupings, but resulted in a decreased mean proportion of membership for individuals from populations 6167, 6480, 6484, and 6485. Moreover, in contrast to previous K values, admixture is not apparent for population 6476.

The groupings described for K = 4 are highly concordant with the differentiation patterns detected in the F_{ST} analysis.

Table 3 F_{IS} per locus and population

Pop	Locus																	
	nc 135	phi 037	Phi 069	phi 127	phi 059	phi 029	bnlg 1700	bnlg 1165	bnlg 1018	bnlg 1209	bnlg 1287	bnlg 1070	bnlg 1182	bnlg 1866	bnlg 1732	bnlg 1360	bnlg 252	Global F_{is}
6473	0.286	−0.079	0.374	0.000	0.000	0.000	0.179	−0.086	0.075	NA	0.203	−0.091	NA	0.056	0.259	0.073	NA	0.095
6167	0.098	0.347	0.342	−0.043	−0.244	0.557	.398[a]	−0.032	0.158	0.112	−0.252	0.200	−0.040	−0.103	−0.315	0.154	−0.036	0.080
6485	0.330	−0.105	0.123	NA	0.450	−0.420	0.359	0.008	−0.043	0.571	−0.048	0.296[a]	−0.021	−0.084	−0.048	0.448[a]	−0.112	0.130
6480	0.162	0.471	0.674[a]	−0.263	−0.102	0.181	0.264	0.193[a]	−0.026	0.455*[a]	0.177	0.120	0.000	0.243	0.293	0.153[a]	0.098	0.189*
6484	0.134	−0.109	0.229	−0.091	0.202	−0.142	0.149	0.089	−0.007	0.347	0.014	0.261	NA	0.132	0.166	0.168	0.043	0.115
6476	0.020	0.623[a]	−0.128	0.006	0.497	0.141	0.534*[a]	−0.142	0.252	0.231	0.118	0.155	−0.073	0.060	0.035	−0.069	−0.161	0.134*
6313	0.268	0.448[a]	−0.241	−0.180	−0.075	0.079	−0.154	0.458	−0.066	0.351[a]	−0.083	0.156	0.113	0.030	−0.129	0.098	0.079	0.100
6482	0.071	0.574[a]	−0.040	0.352	0.686*[a]	−0.301	0.076	0.032	0.294	0.227	−0.062	0.443*[a]	0.220	0.239	0.141	0.467*[a]	0.491	0.220*

Locus phi121 was monomorphic in all populations

NA Not applicable

* Significant under randomization test ($P < 0.00035$)

[a] Significant under *U* test, alternative hypotheses of excess of homozygotes ($P < 0.001$)

However, assessment of ΔK following the methods of Evanno et al. (2005) identified the modal value of this statistic at $K = 2$ (Fig. 1b). In cases of hierarchical population structure, this method is known to preferentially detect the uppermost level of structure. Thus to explore whether substructure could be detected within each cluster, individuals with membership coefficients higher than 0.90 to either of the two groups were then analyzed independently. Admixed individuals from population 6476 were not included in this analysis. Using only the samples from populations 6482 and 6313 clearly resolved two separate entities matching exactly the populations of origin (Fig. 3a). Similarly, the analysis of individuals from populations 6473, 6484, 6480, 6485, and 6167 unambiguously detected two groups by splitting the sample into 6473 and the remaining populations (Fig. 3b). Interestingly, the groups identified by the substructure analysis (Fig. 3) were the same as those retrieved for $K = 4$, with the admixed nature of population 6476 also detected at this K value (Fig. 2).

Evaluation of output stability based on the ad hoc procedures described in Camus-Kulandaivelu et al. (2007) provided no conclusive evidence in favor of any of the K values evaluated. Even if visual inspection of structure outputs revealed contrasting assignments for $K = 7$ and $K = 8$, a clear clustering pattern was observed for all K values in the NJ tree derived from the Euclidean distance matrix, except for a single run from $K = 3$ (Fig. S1, ESM).

The NJ reticulogram based on Nei's genetic distances shows one major partition (bootstrap 94.7%), with the resulting two groups being in close agreement with the uppermost level of structure detected by the bayesian analysis (Fig. 4). Moreover, in spite of belonging to the same race as population 6484 (Amarillo chico), the intermediate position of population 6476 is also concordant with the levels of admixture identified by STRUCTURE.

Racial affiliations

To establish racial affiliations between the landraces studied here and other races from the Americas, a combined dataset including data from Matsuoka et al. (2002a) was subjected to STRUCTURE analysis (336 multilocus genotypes, 10 SSR markers). Since a total of 78 loci were originally used by these authors to infer population structure, and only 10 of them were included in the combined data matrix, the 193 maize accessions of Matsuoka et al. (2002a) were independently reanalyzed to assess the impact of the reduction in marker number, including 78 and 10 SSRs, respectively. No major differences were apparent in the STRUCTURE outputs produced by the 78 and 10 SSR data matrices ($K = 1$–4). In both cases, the sample was divided into the same groupings described by Matsuoka et al.

Table 4 Population pairwise comparisons

Pop	6473	6167	6485	6480	6484	6476	6313	6482
6473		0.171	0.225	0.284	0.230	0.289	0.704	0.562
6167	0.1686*		0.065	0.073	0.059	0.137	0.528	0.289
6485	0.1998*	0.0381*		0.043	0.073	0.154	0.451	0.264
6480	0.2239*	0.0467*	0.0263*		0.077	0.116	0.410	0.232
6484	0.1980*	0.0380*	0.0408*	0.0480*		0.180	0.573	0.321
6476	0.2303*	0.0778*	0.0718*	0.0627*	0.0955*		0.374	0.189
6313	0.3659*	0.2279*	0.1956*	0.1951*	0.2407*	0.1599*		0.417
6482	0.2944*	0.1332*	0.1164*	0.1157*	0.1533*	0.0748*	0.1673*	

Above diagonal: Nei's (1972) genetic distance (D). Below diagonal: pairwise genetic differentiation index (F_{ST})

* $P < 0.001$

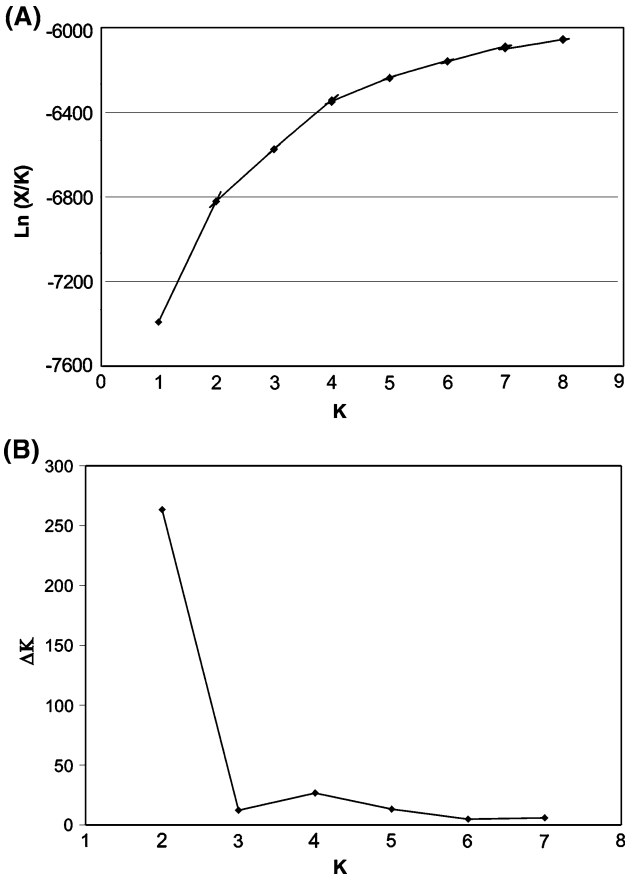

Fig. 1 Evaluation of STRUCTURE outputs for maize landraces from Northwestern Argentina. **a** Mean Ln (X/K) over ten runs for each K value. **b** Rate of change in the log probability of data between successive K values (ΔK)

(2002a), with most of the differences being restricted to the assignment of membership coefficients (data not shown).

For all subsequent analyses, American landraces were labeled following the categories proposed by Matsuoka et al. (2002a) based on geographical origin (Caribbean, Eastern and Central USA, Guatemala and Southern Mexico, Highland Mexico, Lowland Western and Northern Mexico, Northern Mexico, Southwestern USA, Core Andean South American, Other South American). As with the NWA dataset, the identification of the the optimal value of K for the combined matrix was hindered by a constant increase in ln(X/K) (Fig. 5a). The procedure of Evanno et al. (2005) detected the maximal ΔK at $K = 3$ (Fig. 5b), whereas unstability across runs was apparent for $K > 4$ using both visual inspection and the NJ tree derived from the euclidean distance matrix between STRUCTURE outputs (Fig. S2, ESM).

Memberships coefficients are high at both $K = 3$ and $K = 4$ (0.84 and 0.82, respectively), with a rather uniform genetic constitution within the categories previously defined by Matsuoka et al. (2002a), and particularly within NWA populations (Fig. 6). The clustering pattern obtained for $K = 3$ reveals a clear association between: (1) populations 6484, 6473, 6485, 6480, 6167, and the Core Andean South American accessions; (2) population 6313 and the accessions from Eastern, Central and Southwestern USA; and (3) population 6482 and the accessions from the Caribbean, Guatemala and Southern Mexico, Lowland Western and Northern Mexico, and Other South American groups. Once again, admixed individuals seem to be prevalent in population 6476. At $K = 4$, the last group is further split into two subclusters, with population 6482 showing a closer affiliation to the Caribbean accessions and several members of the Other South American group.

Discussion

Genetic diversity

Despite their marginal location, maize landraces from NWA exhibited high levels of genetic diversity. On a regional basis, the mean number of alleles per locus (10.2) was more than twice the estimates obtained for popcorn populations from NEA (4.78, Bracco et al. 2009), and considerably higher than the estimates reported for maize landraces from Mexico and Venezuela (5.76, Matsuoka et al. 2002b) and 25 accessions of Mexican landraces (7.84, Reif et al. 2006). Comparison of the mean number of alleles between NWA populations, American landraces (Matsuoka

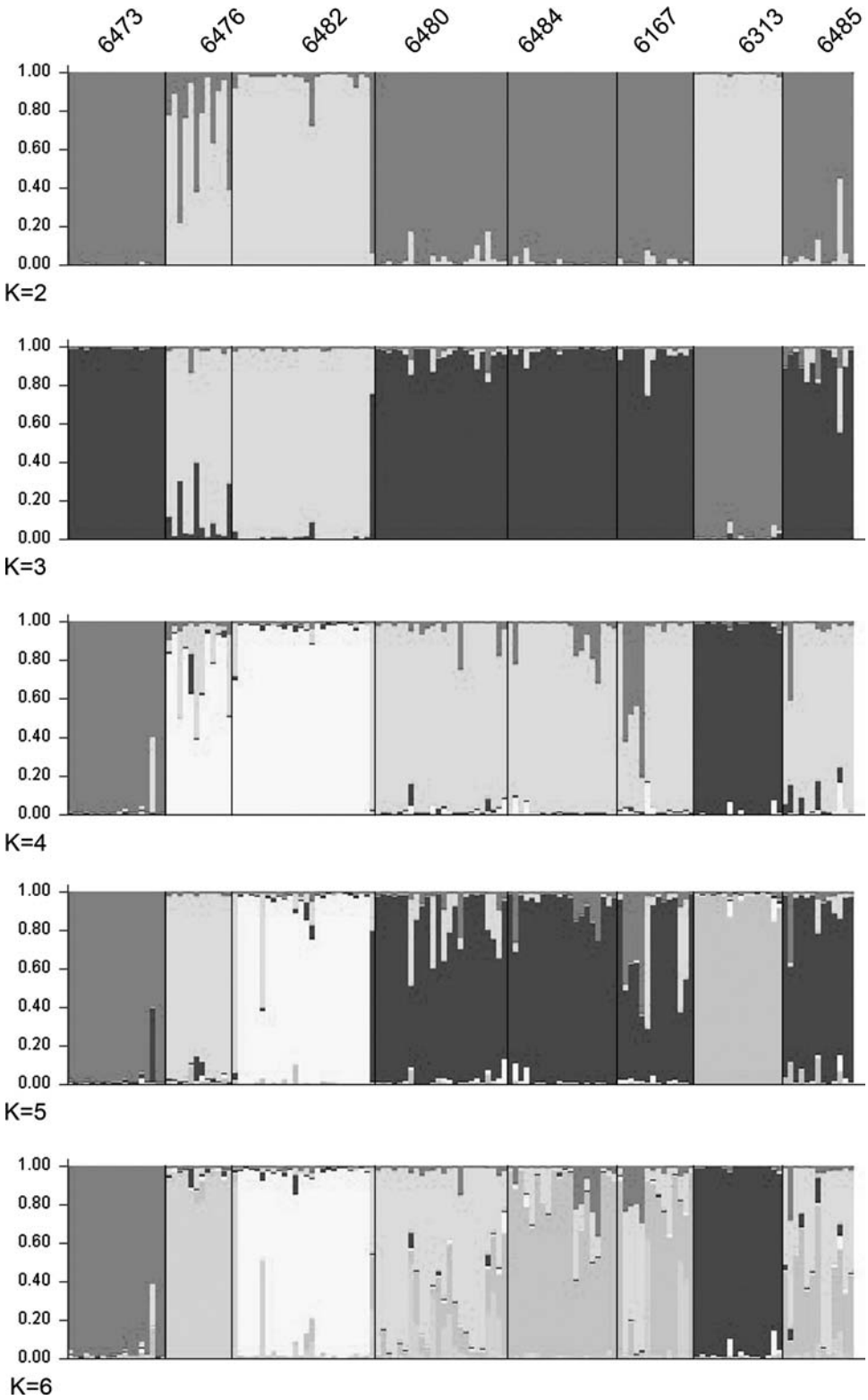

Fig. 2 Estimated population structure of maize landraces from Northwestern Argentina. Each individual is represented by a thin vertical segment, which can be partitioned into *K* colored segments that represent the individual estimated membership to the *K* cluster. The run with the highest ln(*X*/*K*) was chosen for graphical representation of each *K*

et al. 2002a), and the inbreds described by Liu et al. (2003) revealed a strong influence of marker choice on the assessment of genetic variation, even when only SSR data are being considered. After correcting for this bias by re-calculating the mean number of alleles and allelic richness using only the 10 SSR shared by the three studies, differences in genetic diversity were markedly reduced, with the genetic diversity observed in NWA landraces accounting for 63% of that exhibited by American landraces and for 74% of that shown by the inbreds.

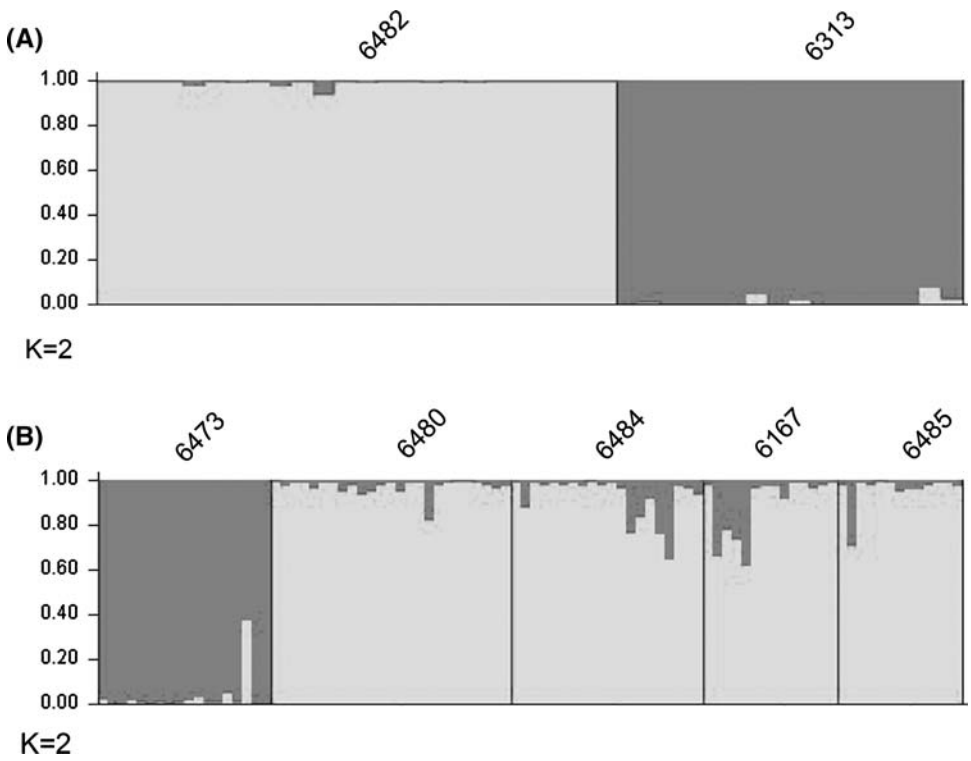

Fig. 3 Estimated population structure of the two main clusters identified for NWA landraces by the ΔK ad hoc procedure. **a** Subcluster analysis of the group 6313–6482. Membership coefficients obtained at the optimal K value ($K = 2$). **b** Subcluster analysis of the groups 6473, 6480, 6484, 6167, and 6485. Membership coefficients obtained at the optimal K value ($K = 2$)

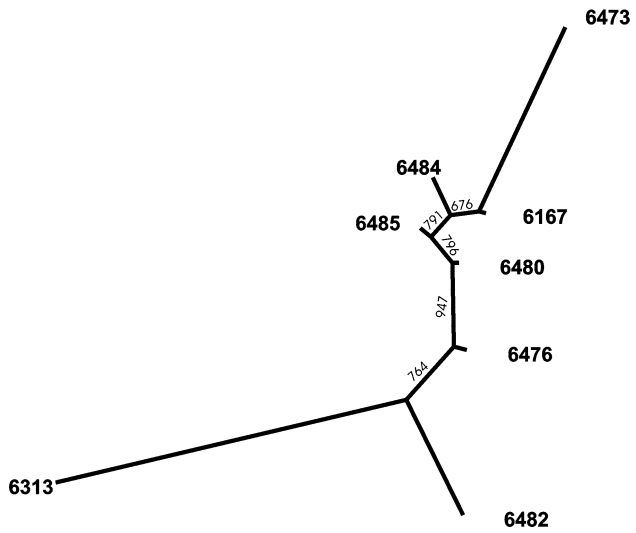

Fig. 4 Neighbor-joining network based on Nei's (1972) genetic distances between maize landraces from Northwestern Argentina. Bootstrap values are indicated beside branches

Liu et al. (2003) observed a large proportion of private alleles in the inbreds (556/2,039) and interpreted it as a function of the high mutation rate of maize SSRs. However, two of these alleles were also found in NWA landraces. Considering the small number of loci common to both studies, it is likely that many of these inbred-specific alleles are also present in other landraces not yet evaluated, suggesting that a poor characterization of the landraces involved in the origin of the inbreds together with a high mutation rate might be responsible for the observed proportion of private alleles in the lines. Indeed, while only 36 alleles were detected by Matsuoka et al. (2002a) in argentine landraces (4 accessions, 10 SSR), our sampling scheme allowed identification of 131 alleles, three of which were not described for the inbreds or the American landraces.

At the population level, the mean number of alleles and gene diversity was also high (A: 4.74, H_e: 0.571) as compared to values obtained for 25 CIMMYT populations derived from different racial complexes (A: 3.76, H_e: 0.52) (Reif et al. 2004) and for different accessions of races of maize from Mexico (A: 3.44, H_e: 0.48) (Reif et al. 2006). In contrast, the gene diversity in NWA populations was lower than the estimates obtained in maize landraces collected directly from traditional farmer fields in the Central Valleys of Oaxaca (H_e: 0.70) (Pressoir and Berthaud 2004).

Population structure

As an outbreeding species, maize populations are expected to conform to Hardy–Weinberg proportions. However, many studies have reported deviations from panmixia using different molecular markers (Dubreuil and Charcosset 1998; e.g., Kahler et al. 1986; Pressoir and Berthaud 2004; Reif et al. 2004). Most of the populations studied here (5/8) exhibited an overall fit to HW expectations, with three populations showing a global excess of homozygotes. Experimental errors, null alleles, population substructuring, selection, and assortative mating are often invoked to

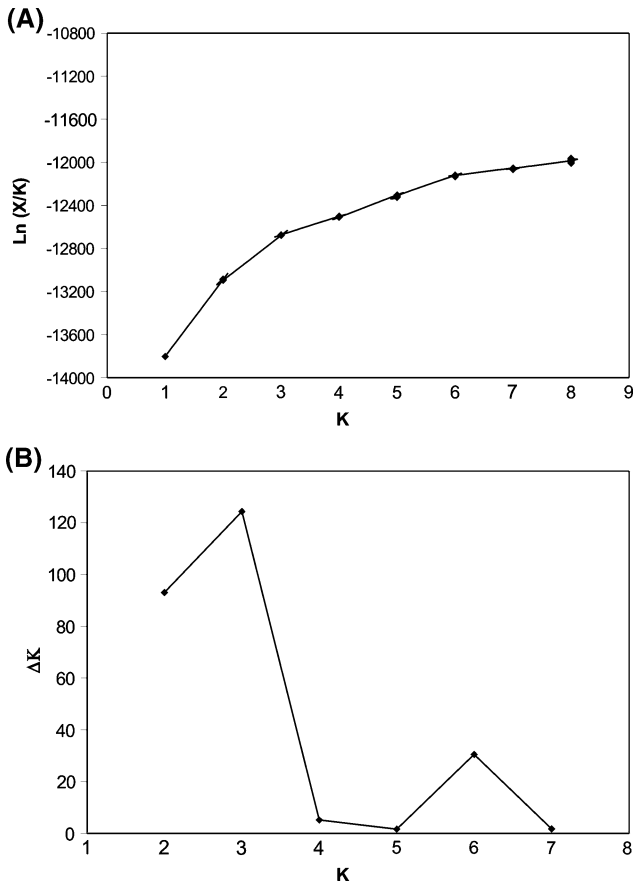

Fig. 5 Evaluation of STRUCTURE outputs for a combined dataset including maize landraces from Northwestern Argentina and a set of American landraces from Matsuoka et al. (2002a). **a** Mean ln(*X*/*K*) over ten runs for each *K* value. **b** Rate of change in the log probability of data between successive *K* values (Δ*K*)

explain the observed homozygote excess or deficit in allogamous species. A detailed inspection of Table 3 shows that the deviations are only apparent for a few loci, even in populations exhibiting global departures from panmixia. While experimental errors are always a possibility, no evidences of null alleles were found in any of the loci showing significant F_{IS} values. Assuming that biological causes underlay the observed deviations, such a heterogeneous behavior across loci is consistent with both selection and assortative mating, but is not concordant with what would be expected for processes affecting the genome as a whole (i.e., population substructuring or admixture). Pressoir and Berthaud (2004) found that those SSR showing the highest F_{IS} values in mexican maize landraces (phi011, phi024, and phi452693) corresponded to loci that map close to genes or QTLs involved in flowering time or anthesis-silking interval (dwarf8, indeterminate1, and dwarf9) (Colasanti and Sundaresan 1998; Gale and Devos 1998; Thornsberry et al. 2001). Interestingly, phi037 and bnlg1700, the two loci responsible for the deviations in population 6476 are both located in the same chromosome regions as phi011 and phi024, and map closer than them to indeterminate1 and dwar9 (Genetic 2005 1 and Pioneer Composite 1999 5 maps at MaizeGDB). However, no linkage disequilibrium has been reported between the remaining loci and any known genes associated to flowering traits.

Maize population dynamics is inextricably linked to human activities. In addition to biological determinants, farmers' management practices greatly influence the distribution of genetic diversity and the levels of gene flow. In Northwestern Argentina, maize is primarily a subsistence crop, produced on small plots receiving little to no input from commercial varieties. The levels of genetic differentiation found among the populations in this study are generally lower than the mean observed for other outcrossing natural populations (F_{ST}: 0.22) (cf. Nybom 2004). In contrast, the F_{ST} estimates obtained for white kernel maize landraces of the Central Valleys of Oaxaca (F_{ST}: 0.011) were markedly lower than those of NWA populations and even more lower when compared to the landraces studied by Reif et al. (2006) (G_{ST}: 0.22). Although distinct dynamics of field management and interchange practices among farmers from different agroecosystems may partially explain these discrepancies, the homogeneity of the races under study in terms of the number of racial complexes involved in their origin also needs to be taken into consideration. Indeed, only a single racial complex was included in the analysis of Pressoir and Berthaud (2004), while representatives of the four racial groups delimited by Wellhausen et al. (1952) were used in the analysis of Reif et al. (2006). According to their morphological and cytogenetic features (Poggio et al. 1998; Rosato et al. 1998; Cámara Hernández and Miante Alzogaray, personal communication), the landraces included in this study can be tentatively assigned to three different groups: (1) the landraces from the Andean Complex described by McClintock et al. (1981) on the basis of *knob* number and position (6473, 6476, 6480, 6485, 6484, 6167); (2) the andean popcorns (6313); and (3) the incipient modern races derived from hybridization between native germplasm and improved varieties developed in Argentina in the mid 1960s (6482). The observed patterns of genetic differentiation are in agreement with the proposed groupings, since the average F_{ST} value for pairwise comparisons within the landraces from the so-called Andean Complex was 0.11, while comparisons between groups always resulted in average F_{ST} values higher than 0.15. Moreover, it is also worth mentioning that 15 private alleles were found in population 6482. The relatively high levels of differentiation of 6473 with respect to the remaining members of the Andean Complex group may seem to contradict its inclusion within the latter. However, the analysis of allele distribution revealed that all the alleles present in 6473 are a subset of those identified in 6167,

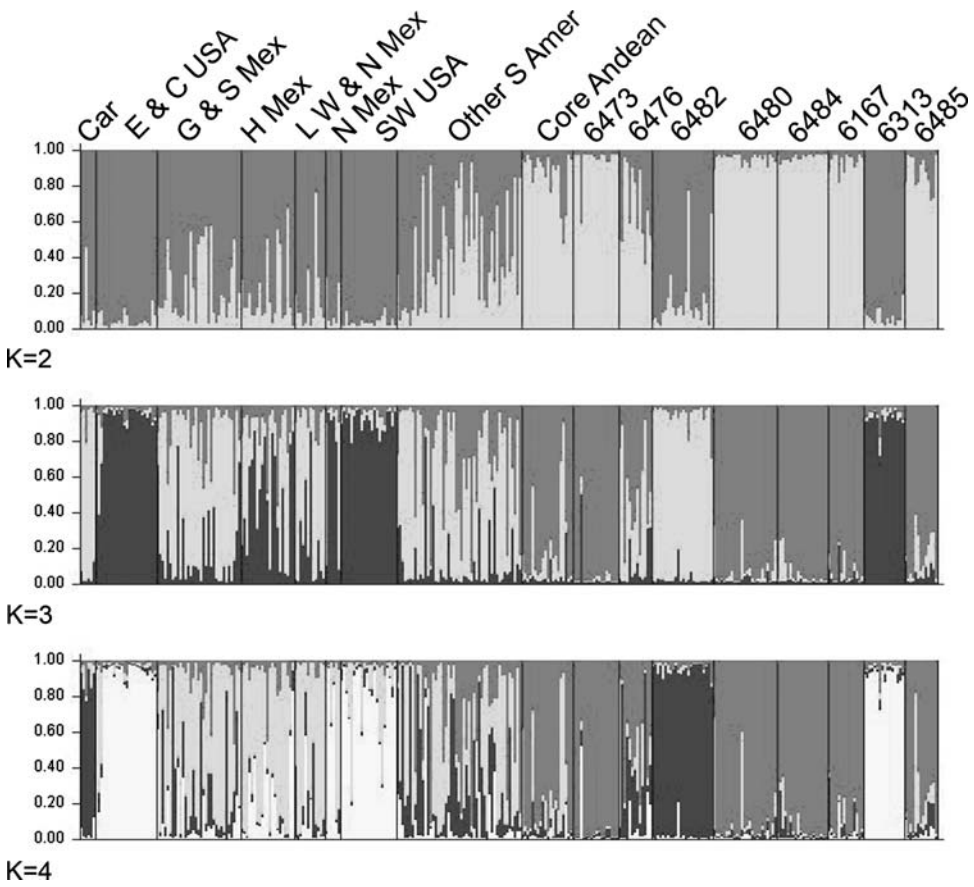

Fig. 6 Estimated population structure of maize landraces from Northwestern Argentina in the context of the American landraces from Matsuoka et al. (2002a). Each individual is represented by a thin vertical segment, which can be partitioned into *K* colored segments that represent the individual estimated membership to the *K* cluster. The run with the highest ln(*X*/*K*) was chosen for graphical representation of each *K*. *Car* Caribbean, *E & C USA* Eastern and Central USA, *G & S Mex* Guatemala and Southern Mexico, *H Mex* Highland Mexico, *LW & N Mex* Lowland Western and Northern Mexico, *N Mex* Northern Mexico, *SW USA* Southwestern USA, *Other S Amer* Other South American, *Core Andean* Core Andean South American

6484, 6476, 6480, and 6485. This observation, together with the reduced levels of variability detected in 6473 and its marginal location at 3,600 m.a.s.l., suggests that the observed degree of differentiation is more likely the result of strong genetic drift than evidence for the existence of a distinct gene pool.

Genetic differentiation measures may be interpreted as a result of ongoing gene flow, the consequence of historical association or a combination of both (Nielsen and Slatkin 2000). In practice, however, distinguishing between these scenarios is not possible always. The strong correlation between the levels of genetic differentiation and the racial affiliations proposed on the basis of morphological and cytogenetic evidence suggests that the historical association is probably the main factor in shaping population structure for the landraces studied here, although a certain amount of gene flow has also been detected (Lia et al. 2007).

Bayesian analysis of population structure was also consistent with the occurrence of three main gene pools. The uppermost level of structure unambiguously discriminated the landraces of the Andean Complex from those of the Andean popcorns and Incipient races, except for the case of population 6476 which exhibited a highly admixed genetic constitution, probably derived from pollen mixing among landraces cultivated on the same plot (Fig. 2). Further subcluster analysis also revealed a clear distinction between the races Pisingallo (6313) and Orgullo Cuarentón (6482), while a highly uniform genetic constitution was apparent for populations 6480, 6484, 6167, and 6485. Again, as in the F_{ST} analysis, the differences in allele frequencies between population 6473 and the remaining populations from the Andean Complex led to its assignment into a separate cluster (Fig. 3b). Although Bayesian analysis clearly identified what seems to be the uppermost level of structure for the populations studied here, the difficulties encountered in determining the optimal value of *K* may obscure the detection of further substructuring with real biological significance. It has been shown that both the log-likelihood values for the data conditional on *K*, ln(*X*/*K*), and ΔK perform well in identifying *K* when genetic differentiation is strong and perform poorly when differentiation is weak (Waples and Gaggiotti 2006). Moreover, the underlying structure model is not well suited to situations in which allele frequencies vary gradually across a region (i.e., populations exhibiting isolation by distance) (Pritchard and Wen 2003). In our case, both low levels of differentiation among certain populations (Table 4) and isolation by distance may account for the difficulties in inferring *K*. An isolation by distance pattern was found for populations 6476, 6482, 6480, 6484, 6167, 6313, and 6485 in the analysis of an

altitudinal cline of B chromosomes (Lia et al. 2007), however this pattern did not persist when population 6473 was included in the analysis (data not shown).

Racial affiliations

Racial classification of maize landraces has been established primarily based on their morphology and geographic origin (for a review see Goodman and Brown 1988). Numerical analyses of Latin American landraces allowed identification of 14 racial complexes, which resulted almost entirely concordant with the affiliations proposed using traditional morphometric analysis (Goodman and Bird R 1977). From a genetic standpoint, Anderson and Cutler (1942) defined a race as "a group of individuals with a significant number of genes in common, major races having a smaller number in common than do sub-races". Numerous genetic studies have been carried out with the main goal of describing the patterns of genetic diversity, but with little to no discussion of racial affiliations. One of the main problems in interpreting the large body of information available is the regional, and to some extent fragmentary nature of most of the genetic analysis conducted till date. In addition, the comparisons among regions are often precluded by a lack of homogeneity in the names of races (i.e., the same race is given different names depending on the area of cultivation) and by the inconsistencies in the conceptual definition of a race, particularly among researchers from different countries (Goodman and Bird 1977). Thus, while the genetic variation of the races of maize from Mexico (Doebley et al. 1985), Guatemala (Bretting et al. 1990) or Southwestern US (Doebley et al. 1983), to cite a few examples, has been described in detail, integration among studies has rarely been accomplished. Despite these limitations, several racial complexes appear to emerge consistently from either morphological, cytogenetic or molecular data. One such entity is the Andean Complex (McClintock et al. 1981) or the Central Andean Complex (Goodman and Bird R 1977). According to these authors, most of the Andean races fall into this group and are characterized by small, rounded, often grenade-shape ears and by elliptically-shaped kernels, usually floury and highly colored. Cytogenetically, these races present two small knobs at chromosome arms 6L and 7L. The race Pisingallo, or Pisinkalla, is also associated to mid- to high-altitude Andean regions. However, neither its chromosome constitution nor its morphological characteristics support its inclusion within the Andean Complex, being frequently assigned to a group defined as Southern Popcorns (Goodman and Bird R 1977). As previously mentioned, four of the races analyzed here—Altiplano (6473), Amarillo Grande (6484), Amarillo chico (6476–6480), and Blanco (6485)—can be tentatively assigned to the Andean Complex; whereas Pisingallo (6313) and Orgullo Cuarenton (6482) can be best categorized as Southern Popcorns and incipient races, respectively.

Combined analysis of NWA and American landraces using the bayesian model-based approach of Pritchard et al. (2000) retrieved essentially the same groupings described in the population structure analysis of Matsuoka et al. (2002a) (namely, US maize, Andean maize and all other South American and Mexican maize) and revealed additional affiliations between southern South American and Northamerican germplasm. Bayesian analysis confirmed the assignment of the races Altiplano, Amarillo grande, Amarillo chico, and Blanco to the Andean Complex, while the race Orgullo Cuarentón exhibited a close affiliation with germplasm from the Caribbean, and Pisingallo was clearly associated with native materials from the US (Eastern and Central USA and Southwestern USA) (Figs. 6, 7).

Camus-Kulandaivelu et al. (2006) analyzed the population structure of a set inbreds obtained directly by selfing from traditional American landraces and found that they can be organized into five groups: Northern Flint, Corn Belt Dent, Mexico, Caribbean, and Andean. More recently, a comprehensive large-scale study of New World maize races reported the existence of four main clusters: Highland Mexico, Tropical Lowland, Andean, and Northern US (Vigouroux et al. 2008). Although the groups obtained here and those described by the above-mentioned studies are not strictly concordant, they all depict the same overall pattern with differences in sample composition which is the most likely cause for the observed discrepancies. For instance, most of the races categorized as Tropical Lowland by Vigouroux et al. (2008) were not included in the Matsuoka et al. (2002a) analysis or in our combined dataset. Similarly, Corn Belt dents were not included in the study of Matsuoka et al. (2002a) or Vigouroux et al. (2008) because they were considered to be more recent in origin. Conversely, despite the differences in nomenclature, almost the same set of landraces is included in both the Northern US cluster of Vigouroux et al. (2008) and the U.S. maize group of Matsuoka et al. (2002a).

While the relationship of NWA landraces to the Andean Complex was somewhat expected, the associations observed for Orgullo Cuarentón and Pisingallo cannot be so readily explained. In the case of Orgullo Cuarentón, a relatively recent exchange of germplasm can account for its relationship with the Caribbean accessions. It has been proposed that the race Cuban Flint, the only true flint found in Cuba, was introduced from Argentina in the early 1900s as a result of accidental crossing of Argentine flint with local Cuban varieties (Goodman and Brown 1988). Considering that Orgullo Cuarentón is an incipient race which received contributions from improved varieties, the observed affiliations may well derived from the same varieties being involved in the origin of both Orgullo Cuarentón and

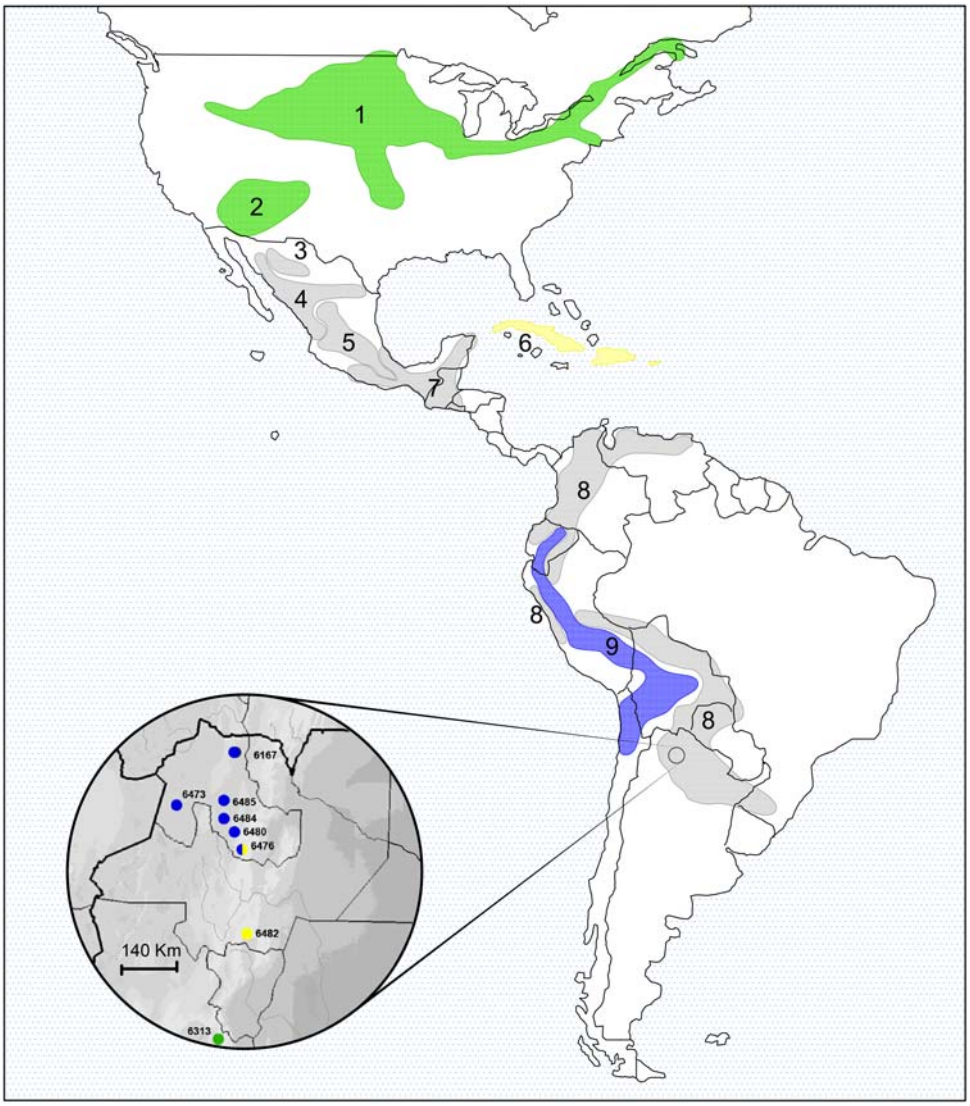

Fig. 7 Geographical origin and racial affiliations of the populations included in this study. Relationships identified by Bayesian population structure analysis among the landraces from Northwestern Argentina and the American landraces of Matsuoka et al. (2002a) are highlighted in color. Shaded areas correspond to the geographical categories defined by Matsuoka et al. (2002a). *1* Eastern and Central USA, *2* Southwestern USA, *3* Northern Mexico, *4* Lowland Western and Northern Mexico, *5* Highland Mexico, *6* Caribbean, *7* Guatemala and Southern Mexico, *8* Other South American, *9* Core Andean South American

Cuban Flint. Although appealing, this hypothesis was not supported by the isoenzymatic and karyological studies of Bretting et al. (1987), which revealed that the maize races of the Caribbean represent a supraracial group distinct from the allied races from the American mainland with no evidence of association with the Argentinian Catetos. Alternatively, it cannot be discarded that improved varieties from the US Corn Belt, which are themselves derived from germplasm from the Caribbean (Goodman and Bird R 1977), were also involved in the generation of Orgullo Cuarentón.

Santacruz-Varela et al. (2004) recently studied the phylogenetic relationships among North American popcorns and their evolutionary links to Mexican and South American popcorns, including four Argentine accessions (i.e., Pisingallo, Perlita, Perlita mediano, and Argentine Pop). Perlita, Perlita mediano, and Argentine Pop were included within a cluster mainly composed of popcorns from the lowlands of the guaranitic plains (Northestern Argentina, Southwestern Brasil, and Paraguay), whereas Pisingallo was positioned in a distantly related cluster which also included Latin American pointed popcorns (Palomero Toluqueño, Canguil, Confite Puntiagudo, and Pisinkalla) and Southwestern US pointed materials (White Rice Pop and Acoma Pueblo). Interestingly, the Latin American pointed popcorns positioned more closely to the North American pointed rice popcorns than to other Latin American popcorns, defining a cluster in close association to the North American early popcorns. It is important to note that two non-popcorn races, Tama Flint and Fairfax Brown, were also included within this last group. The former corresponds to the category of Northern Flints and Flours, whereas Fairfax Brown is a member of the Southwestern racial complex. A similar pattern of affiliations can be observed here for the race Pisingallo (6313), which according to kernel morphology and the frequency distribution of alleles at loci *phi127*, *phi121* and *phi029* corresponds to the group of Latin American pointed popcorns. Bayesian

analysis showed a clear association between this race and two groups of native materials from North America (Fig. 6). The group delimited as Eastern and Central US is mainly composed of North American Flints and Flours, and most of the members of the category Southwestern US belong to the Southwestern racial complex.

Morphological, genetic and cytogenetic evidence suggests that the North American pointed rice popcorns probably originated from the traditional races of pointed popcorns from Latin America, diffusing from the highlands of Central Mexico into Northern Mexico and then into southwestern US, thus explaining the similarities between both groups (Santacruz-Varela et al. 2004). McClintock et al. (1981) supported the view that the Pisingallo germplasm was introduced in South America at an early stage during its development following the establishment of the races of the Andean Complex, and postulated that it could be derived from both Mexican and Guatemalan maize races. On the contrary, other authors have suggested that the southern popcorns have recently descended from Paraguayan germplasm or, alternatively, that they have been developed in 1920s near Kansas City (US) and later introduced into South America (Brieger et al. 1958; Brunson and Bower 1931; Roberts et al. 1957). These timing discrepancies may reflect the coexistence in South American maize popcorns with multiple origins, as evidenced by the two distinct groups obtained in the analysis of Santacruz-Varela et al. (2004).

In conclusion, in spite of representing only a limited proportion of the traditional landraces currently cultivated in Northwestern Argentina, the landraces examined here have shown to harbor considerable levels of genetic diversity with contributions from at least three different gene pools. Further studies encompassing larger samples of landraces from this area will certainly help to unveil additional allelic variation, which may prove highly valuable, particularly in the context of adaptation to extreme environments.

Acknowledgments The authors wish to thank Ing. J. Cámara Hernández and A.M. Miante Alzogaray for helpfull discussion on racial characteristics. We are grateful to Dr. J. Doebley and Dr. Y. Matsuoka for kindly providing us the SSR data from Matsuoka et al. (2002a) and Liu et al. (2003). We are also in debt to Dr. Jim Holland, for generously sending us the inbred lines used to establish allele equivalencies between datasets; and to the anonymous reviewers who have greatly improved the manuscript. Several grants from the Consejo Nacional de Investigaciones Científicas y Técnicas (PIP 2296, PIP 5927), the Universidad de Buenos Aires (EX-317), and the Agencia Nacional de Promoción Científica y Tecnológica (BID 1201 OC-AR PICT 04443, BID 1728 OC-AR583 PICT 14119) are gratefully acknowledged.

References

Anderson E, Cutler HC (1942) Races of *Zea mays*: I. Their recognition and classification. Ann Mo Bot Gard 29:69–89

Bracco M, Lia V, Gottlieb A, Cámara Hernández J, Poggio L (2009) Genetic diversity in maize landraces from indigenous settlements of Northeastern Argentina. Genetica 135:39–49

Bretting PK, Goodman MM, Stuber CW (1987) Karyological and isozyme variation in West Indian and allied American mainland races of maize. Am J Bot 74:1601–1613

Bretting PK, Goodman MM, Stuber CW (1990) Isozymatic variation in Guatemalan races of maize. Am J Bot 77:211–225

Brieger FG, Gurgel JTA, Paterniani E, Blumenschein A, Alleoni MR (1958) The races of maize in Brazil and other eastern South American countries. National Academy of ScienceNational Research Council Publications, Washington, DC

Brunson AM, Bower CW (1931) Pop corn. USDA Farmers′ Bull 1679:17

Brush SB, Perales H (2007) A maize landscape: ethnicity and agro-biodiversity in Chiapas Mexico. Agric Ecosyst Environ 121:211–221

Cámara Hernández J, Miante Alzogaray AM (1997) Las razas de maíz de Jujuy y Salta Argentina. Proc 1er Taller Internacional de Recursos Fitogenéticos del Noroeste Argentino 1:19–23

Camus-Kulandaivelu L, Veyrieras JB, Madur D et al (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the Dwarf8 gene. Genetics 172(4):2449–2463

Camus-Kulandaivelu L, Veyrieras J-B, Gouesnard B, Charcosset A, Manicacci D (2007) Evaluating the reliability of structure outputs in case of relatedness between individuals. Crop Sci 47:887–890

Colasanti J,ZY, Sundaresan V (1998) The indeterminate gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize. Cell 93:593–603

Doebley J, Goodman MM (1984) Isoenzymatic variation in *Zea* (Gramineae). Syst Bot 9:203–218

Doebley J, Goodman MM, Stuber CW (1983) Isozyme variation in maize from the southwestern United States taxonomic and anthropological implications. Maydica 28:94–120

Doebley J, Goodman MM, Stuber CW (1985) Isozyme variation in the races of maize from Mexico. Am J Bot 72:629–639

Doebley J, Goodman MM, Stuber CW (1986) Exceptional genetic divergence of Northern Flint Corn. Am J Bot 73:64–69

Doebley J, Goodman MM, Stuber CW (1987) Patterns of isozyme variation between maize and Mexican Annual teosinte. Econ Bot 41:234–246

Doebley J, Wendel JF, Smith JSC, Stuber CW, Goodman MM (1988) The origin of cornbelt maize: the isozyme evidence. Econ Bot 42:120–131

Dubreuil P, Charcosset A (1998) Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. Theor Appl Genet 96:577–587

El Mousadik A, Petit R (1996) High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* (L.) Skeels) endemic to Morocco. Theor Appl Genet 92:832–839

Evanno G, Regnaut S, 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, Pritchard JK (2003) Inference of population structure: extensions to linked loci and correlated allele frequencies. Genetics 164:1567–1587

Felsenstein J (1991–2004) Phylip (Phylogeny Inference Package). University of Washington, Washington

Gaggiotti OE, Lange O, Rassmann K, Gliddon C (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. Mol Ecol 8:1513–1520

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Goodman MM (1990) Genetic and germplasm stocks worth conserving. J Hered 81:11–16

Goodman MM (2005) Broadening the U.S. germplasm base. Maydica 50:203–214
Goodman MM, Bird R McK (1977) The races of Maize IV: tentative grouping of 219 Latin American Races. Econ Bot 31:204–221
Goodman MM, Brown WL (1988) Races of corn. In: Sprague GF, Dudley JW (eds) Corn and corn improvement. Amer Soc Agron, Madison, WI, pp 33–79
Goodman MM, Stuber CW (1983) Races of maize. VI. Isozyme variation among races of maize in Bolivia. Maydica 28:169–187
Goudet J (1995) Fstat version 1.2: a computer program to calculate F-statistics. J Hered 86:485–486
Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet 1995
Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163:1467–1482
Hoisington D, Khairallah M, Reeves T et al (1999) Plant genetic resources: what can they contribute toward increased crop productivity? Proc Natl Acad Sci USA 96:5937–5943
Horovitz S (1935) Distribución geográfica de factores genéticos en los maíces autóctonos del norte argentino. Rev Arg Agron 2:133–135
Kahler AL, Hallauer AR, Gardner C (1986) Allozyme polymorphism within and among open-pollinated and adapted exotic populations of maize. Theor Appl Genet 72:592–601
Lia VV, Confalonieri VA, Poggio L (2007) B chromosome polymorphism in maize landraces: adaptive vs. demographic hypothesis of clinal variation. Genetics 177:895–904
Liu K, Goodman M, Muse S et al (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165:2117–2128
Matsuoka Y, Vigouroux Y, Goodman MM et al (2002a) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084
Matsuoka Y, Mitchell SE, Kresovich S, Goodman MM, Doebley JF (2002b) Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet 104:436–450
McClintock B, Kato TA, Blumenschein A (1981) Chromosome constitution of the races of maize, its significance in the interpretation of relationships between races and varieties in the Americas. Colegio de Postgraduados, Chapingo
Minch E, Ruiz Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL (1996) Microsat (version 1.5b): a computer program for calculating various statistics on microsatellite allele data. (Program available from: http://hpgl.stanford.edu/projects/microsat/)
Nei M (1972) Genetic distance between populations. Am Nat 106:283–292
Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
Nielsen R, Slatkin M (2000) Likelihood analysis of ongoing gene flow and historical association. Evolution 54:44–50
Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13:1143–1155
Page RDM (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358
Perales RH, Benz BF, Brush SB (2005) Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. Proc Natl Acad Sci USA 102:949–954
Poggio L, Rosato M, Chiavarino AM, Naranjo CA (1998) Genome size and environmental correlations in maize (*Zea mays* ssp. mays, Poaceae). Ann Bot (Lond) 82:107–115
Pressoir G, Berthaud J (2004) Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. Heredity 92:88–94
Pritchard JK, Wen W (2003) Documentation for STRUCTURE software: Version 2. http://pritch.bsd.uchicago.edu
Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959
Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249
Reif JC, Xia XC, Melchinger AE et al (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 J, Warburton M, Xia X et al (2006) Grouping of accessions of Mexican races of maize revisited with SSR markers. Theor Appl Genet 113:177–185
Rice W (1989) Analysing tables of statistical tests. Evolution 43:223–225
Roberts LM, Grant UJ, Ramírez ER, Hatheway WH, PLwPCM Smith (1957) Races of Maize in Colombia. National Academy of Science–National Research Council Publications, Washington, DC
Rosato M, Chiavarino AM, Naranjo CA, Cámara Hernandez J, Poggio L (1998) Genome size and numerical polymorphism for the B chromosome in races of maize (*Zea mays* ssp. mays, Poaceae). Am J Bot 85:168–174
Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
Sanchez GJ, Stuber CW, Goodman MM (2000) Isozymatic diversity in the races of maize of the Americas. Maydica 45:185–203
Santacruz-Varela A, Widrlechner MP, Ziegler KE et al (2004) Phylogenetic relationships among North American popcorns and their evolutionary links to Mexican and South American popcorns. Crop Sci 44:1456
Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462
Taller JM, Bernardo R (2004) Diverse adapted populations for improving northern maize inbreds. Crop Sci 44:1444–1449
Tarter JA, Goodman MM, Holland JB (2003) Testcross performance of semiexotic inbred lines derived from Latin American maize accessions. Crop Sci 43:2272–2278
Thornsberry JM, Goodman MM, Doebley J et al (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:203–204
Torregrosa M, Cámara Hernandez J, Solari L et al (1980) Clasificación preliminar de formas raciales de maíz y su distribución geográfica en la República Argentina. Actas II Congreso Nacional de Maíz 1:5–17
Vigouroux Y, Glaubitz JC, Matsuoka Y et al (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. Am J Bot 95:1240–1253
Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Mol Ecol 15:1419–1439
Warburton ML, Xianchua X, Crossa J et al (2002) Genetic characterization of CIMMYT inbred lines and open pollinated populations using large scale fingerprinting methods. Crop Sci 42:1832–1840
Weir B, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370
Wellhausen EJ, Roberts LM, Hernandez EMP (1952) Races of maize in Mexico. Harvard University Press, Cambridge
Wright S (1978) Evolution and the genetics of populations, volume 4: variability within and among natural populations, 3rd edn. University of Chicago Press, Chicago