Annals of Applied Biology ISSN 0003-4746

RESEARCH ARTICLE

Genetic diversity of maize landraces from lowland and highland agro-ecosystems of Southern South America: implications for the conservation of native resources

M. Bracco[1,2], V.V. Lia[2,3], J.C. Hernández[4], L. Poggio[1,2] & A.M. Gottlieb[1,2]

1 Laboratorio de Citogenética y Evolución (LaCyE), Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Autónoma de Buenos Aires, Argentina
2 Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina
3 Instituto de Biotecnología, CICVyA, INTA, Buenos Aires, Argentina
4 Laboratorio de Recursos Genéticos Vegetales 'N.I. Vavilov', Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires, Argentina

Keywords
Conservation genetics; maize landraces; microsatellites; Northern Argentina.

Correspondence
Dr V.V. Lia, Instituto de Biotecnología CICVyA, INTA
Castelar Los Reseros y Las Cabañas s/n (B1686ICG)
Hurlingham, Buenos Aires, Argentina.
Email: vlia@cnia.inta.gov.ar

Received: 18 January 2012; revised version accepted: 22 February 2012.

doi:10.1111/j.1744-7348.2012.00544.x

Abstract

The North of Argentina is one of the southernmost areas of maize landrace cultivation. Two distinct centres of diversity have been distinguished within this region: Northwestern Argentina (NWA), and Northeastern Argentina (NEA). Nowadays, maize landraces from this area are faced with two main risks. On the one hand, significant structural and functional changes have modified the rural environment with the boundaries of cropland areas experiencing a rapid expansion at the expense of northern natural forests and rangelands; and on the other, native gene pools are increasingly threatened by hybrids and commercial varieties which are more attractive relative to landraces. The first step towards any conservational action is the acquisition of an inclusive knowledge of the biological resources. For this purpose, our study assesses the genetic diversity and population dynamics of maize landraces from Northern Argentina using microsatellite markers. The Northeastern lowland region (NEA) was represented by 12 landraces (19 populations). In addition, six landraces (eight populations) from the Northwestern highland region (NWA) were used for comparison. For the NEA data set, a total of 126 alleles were found, with an average of 10.5 alleles per locus. Mean H_o, H_e and R_s were 0.350, 0.467 and 2.72, respectively. Global fit to Hardy–Weinberg proportions was observed in 7 of 19 populations. Global estimates of F_{ST} revealed significant differentiation among populations. Bayesian analyses of population structure allowed the recognition of two main gene pools (popcorns versus floury landraces). When NWA was added to the analysis, three clusters were distinguished: NEA popcorns, NEA flours and NWA racial complexes. Additional information on the relationships among these groups was retrieved from cluster analyses. This study shows that lowland landraces from Northern Argentina harbour considerable levels of genetic diversity, with contributions from different gene pools. Further studies encompassing a larger number of populations from the NEA region will certainly help to detect additional genetic variation, which may prove highly valuable in germplasm conservation and management. Future conservation efforts should focus on preserving NEA popcorns, NEA floury and NWA racial complexes as different management units.

Ann Appl Biol **160** (2012) 308–321

Introduction

From a botanical perspective, the species *Zea mays* L. involves the widely known maize (*Z. mays* ssp. *mays* Iltis & Doebley) and three wild taxa [*Z. mays* spp. *parviglumis* Iltis & Doebley, *Z. mays* ssp. *mexicana* (Schrader) Iltis & Doebley, *Z. mays* ssp. *huehuetenangensis* Iltis & Doebley]. The maize cultigen was domesticated in Central America about 9000 years ago (Matsuoka *et al*., 2002a), rapidly spreading northwards and southwards the continent. Biochemical and molecular evidences suggest that this crop emerged in the Southwest of Mexico (Balsas and Lerma river valleys), a place currently inhabited by *Z. mays* ssp. *parviglumis* (Doebley, 1990; Matsuoka *et al*., 2002a). By the time Europeans arrived to America, maize cultivation was particularly abundant in the Azteca and Inca Empires (Rebourg *et al*., 2003) and since then more than 350 landraces have been described (Vigouroux *et al*., 2008). Maize diversity has been shaped by direct human intervention as well as by natural evolutionary processes that together generated the allelic combinations and valuable variation kept in modern landraces. Domesticated varieties therefore constitute a reservoir of useful genes for traditional farmers and, ultimately, for plant breeders. This germplasm is locally adapted to heterogeneous climates and edaphic conditions such as the high altitude arid regions of the Andes, or the humid temperate lowlands of Mexico and the Paranaense forests of South America.

Because of its economic importance, maize has long been the focus of genetic and evolutionary analyses. A large number of studies have documented the cytogenetic and genetic diversity of maize landraces from the Americas (McClintock *et al*., 1981; Sánchez *et al*., 2000; Matsuoka *et al*., 2002b; Reif *et al*., 2006; Vigouroux *et al*., 2008; van Heerwaarden *et al*., 2011). However, most of the research conducted to date has been performed on a region by region basis, focusing primarily on the central areas of maize diversity (i.e. Mexico and the Central Andes), with the more marginal locations being only poorly represented (Goodman & Stuber, 1983; Doebley & Goodman, 1984; Doebley *et al*., 1983, 1985, 1986, 1987, 1988; Bretting *et al*., 1987, 1990; Oliveira Freitas *et al*., 2003; van Etten & de Bruin, 2007). Particularly for southern South America, even though the cytogenetics of the landraces has been extensively investigated at the population level (Poggio *et al*., 1990, 1998, 2005; Chiavarino, 1997; Chiavarino *et al*., 2000, 2001; Gonzalez, 2004), the study of the genetic diversity has been generally restricted to a limited number of individuals either *per* accession or *per* landrace (Sánchez *et al*., 2000; Matsuoka *et al*., 2002b; Vigouroux *et al*., 2008). However, the information derived from individual-level approaches is insufficient for the implementation of effective conservation and management programmes for which a comprehensive study of landrace genetic variability, both within and among regions, is required. The importance of population dynamics studies at micro-geographic scales in the context of *in situ* conservation programmes has been only recently recognised (Perales *et al*., 2003; Pressoir & Berthaud, 2004; Latournerie *et al*., 2006). In this respect, exhaustive microevolutionary analyses of southern South American landraces have begun to arise, revealing high levels of diversity and a complex genetic structure (Bracco *et al*., 2009; Lia *et al*., 2009).

The northern region of Argentina is one of the southernmost areas of maize landrace cultivation. Two distinct centres of diversity were distinguished within this area by the Russian botanical expert Nikolai Vavilov (1992): (a) the highland region, or Northwestern Argentina (NWA), which represents an expansion of the Peruvian Andes and (b) the mesopotamic and Chaco plains or Northeastern Argentina (NEA), a lowland area for which maize landraces were considered to be more closely affiliated with landraces from Brazil and Paraguay. Because of their remarkable phenotypical diversity and their proximity to the Peruvian Andes, the NWA maize landraces have received more attention than those of the NEA region, for which basic knowledge is still scarce (Melchiorre *et al*., 2006; Bracco *et al*., 2009). The Province of Misiones, located in the NEA region, is the second centre of genetic variability in the North of Argentina, with at least 14 native races recorded (Cámara Hernández & Miante Alzogaray, 2003; Melchiorre *et al*., 2006). Over this Province, *Guaraní* indigenous communities cultivate maize landraces descendant from crops originally grown by their ancestors with little or no input from commercial germplasm. As in pre-Hispanic times, maize is a staple cereal having relevant socio-economic values for modern indigenous communities whose farming systems are based on the provision of this crop for self-sufficiency of food needs (Martínez-Crovetto, 1968). Nowadays, maize landraces from this area are faced with two main risks. Agricultural transformation in Argentina caused significant structural and functional changes on the rural environment with the boundaries of cropland areas experiencing a rapid expansion at the expense of northern natural forests and rangelands (Viglizzo *et al*., 2011). On the other hand, as hybrids and commercial varieties become more attractive relative to landraces, the persistence of native gene pools is increasingly threatened unless special efforts are made to collect traditional germplasm or to encourage continued cultivation.

The first step towards any conservational action is the acquisition of an inclusive knowledge of the

biological resources. To explore the genetic diversity and population dynamics of maize landraces cultivated by indigenous communities from Northern Argentina, we conducted a nuclear microsatellite characterisation using a comprehensive ensemble of landraces directly collected from farmer's fields. This study is unprecedented in that sampling intensity at the population and regional levels is greatly increased in comparison to previous contributions on South American maize diversity.

Materials and methods

Twelve maize landraces (19 populations) were collected from the Province of Misiones, hereafter the NEA region (98–525 m above sea level, m.a.s.l.). Taxonomic identification was performed based on morphological criteria. Samples were obtained directly from indigenous farmers at their villages. These small communities maintain their stocks via open-pollination with little or no input from commercial germplasm. Since only a limited number of cobs are saved for sowing every year, sample sizes were determined by seed availability from different households. Each population was named with its corresponding herbarium identification number. Those populations collected from the same village, identified as the same landrace and for which a low number of individuals *per* population was available, were bulked and labelled with the initials of the corresponding landrace (i.e. BA, Co and Ro). 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; 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). Voucher specimens, collection sites, racial identification and sample sizes are given in Table 1 and Fig. 1.

Seed germination and DNA extraction were performed as described in Bracco *et al*. (2009). A set of 12 highly discriminant dinucleotide microsatellite or simple sequence repeat (SSR) markers evenly distributed throughout the genome was selected for genotyping (*bnlg1866, phi037, bnlg1182, bnlg252, bnlg1287, bnlg1732, bnlg1209, bnlg1070, bnlg1360, bnlg105, bnlg1108* and *bnlg1329*). Their sequence, genetic map position and repeat motif are available at the MaizeGDB website (http://www.maizegdb.org/ssr.php). PCR reactions, denaturing polyacrilamide gel electrophoresis, silver-staining and allele size identification were done as previously reported (Bracco *et al*., 2009). Genotypic characterisation of populations 6596, 6618, 6605, 6614 and 6607 at loci *bnlg105*, *bnlg1108* and *bnlg1329*, represents a novel addition to previous data from Bracco *et al*. (2009).

SSR data from Lia *et al*. (2009) were used for comparison between NEA and NWA landraces. This dataset consists of eight populations from six NWA landraces distributed along an altitudinal cline of approximately 2690 m.a.s.l. (Amarillo chico: VAV6484, VAV6476; Amarillo grande: VAV6480; Blanco: VAV6485; Altiplano: VAV6167, VAV6473; Orgullo cuarentón: VAV6482 and Pisingallo: VAV6313). Loci in common between NEA and NWA dataset were: *bnlg1866, phi037, bnlg1182, bnlg252, bnlg1287, bnlg1732, bnlg1209, bnlg1070* and *bnlg1360*.

Individuals showing more than 50% missing data were excluded from the analyses. In summary, the NEA dataset comprises 336 individuals and the combined dataset NEA + NWA comprises 473 individuals. For the NEA dataset (12 loci) the overall probability of identity (PI) is 7.3×10^{-11} and the probability of identity considering the genetic similarity among siblings is 1.2×10^{-4}. For the NEA + NWA dataset (nine loci) the overall PI is 2.6×10^{-10} and the probability of identity considering the genetic similarity among siblings is 3.6×10^{-4}. Both estimates were computed according to Waits *et al*. (2001) using GeneAlEx 6 (Peakall & Smouse, 2006).

Data analysis

Allele frequencies as well as the number of alleles *per* locus *per* population (A), allelic richness (R_s) (El Mousadik & Petit, 1996) and total gene diversity or heterozygosity across populations (H_e) (Nei, 1987) were calculated using the software Fstat 2.9.3.2 (Goudet, 1995, 2001). Estimates of observed heterozygosity *per* population (H_o) were obtained by direct counts from the raw data matrix. 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 among groups differentiation test implemented in the Fstat software. 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.

Departures from Hardy–Weinberg (HW) proportions at individual loci were tested within each population. Estimates of Wright's fixation index F_{IS} (Wright, 1978) were obtained according to Weir & Cockerham (1984) using Fstat program. Significance of F_{IS} was determined using the randomisation test implemented in Fstat and then modified according to Bonferroni procedures at an overall $\alpha = 0.05$ (Rice, 1989).

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 was evaluated using

Ann Appl Biol **160** (2012) 308–321 © 2012 The Authors
Annals of Applied Biology © 2012 Association of Applied Biologists

Table 1 Maize landraces from Northeastern Argentina included in this study

Landrace	Population ID	Voucher[a]	Collection Site at Misiones, Argentina	Sample Size
Pipoca colorado	6596	VAV6596	Aldea Alecrín, Dpto. Eldorado	20
Pipoca colorado	6618	VAV6618	Aldea Alecrín, Dpto. Eldorado	30
Pipoca amarillo	6605	VAV6605	Piñalito, Dpto. San Pedro	20
Pororó azul	6607	VAV6607	Aldea Alecrín, Dpto. Eldorado	25
Pororó grande	6614	VAV6614	Aldea Guavirá Poty, Dpto. San Pedro	20
Blanco ancho	6586	VAV6586	Aldea Alecrín, Dpto. Eldorado	10
Blanco ancho	6615	VAV6615	Aldea Guavirá Poty, Dpto. San Pedro	16
Blanco ancho	6602	VAV6602	Pozo Azul, Dpto. San Pedro	15
Blanco angosto	Ba	VAV6599; VAV6587	Aldea Alecrín, Dpto. Eldorado	30
Blanco angosto	6610	VAV6610	Pozo Azul, Dpto. San Pedro	20
Tupí blanco	6592	VAV6592	Pozo Azul, Dpto. San Pedro	16
Overo	6595	VAV6595	Aldea Alecrín, Dpto. Eldorado	8
Overo	6617	VAV6617	Aldea Guavirá Poty, Dpto. San Pedro	10
Overo	6601	VAV6601	Pozo Azul, Dpto. San Pedro	16
Colorado	Co	VAV6594; VAV6589	Aldea Alecrín, Dpto. Eldorado	20
Colorado	6604	VAV6604	Pozo Azul, Dpto. San Pedro	12
Azul	6597	VAV6597	Aldea Alecrín, Dpto. Eldorado	20
Variegado	6585	VAV6585	Aldea Alecrín, Dpto. Eldorado	20
Rosado	Ro	VAV6616; VAV6621	Aldea Guavirá Poty, Dpto. San Pedro	14

Dpto., Departamento.

[a]Voucher specimens are deposited at the 'Laboratorio de Recursos Genéticos Vegetales N.I. Vavilov' (VAV), Facultad de Agronomía, Universidad de Buenos Aires.

Figure 1 Geographic location of the populations included in this study. A: 6610, 6601, 6592, 6604 and 6602; B: 6586, BA, 6595, Co, 6597, 6585, 6596, 6618 and 6607; C: 6605; D: 6617, 6614, 6615 and Ro; E: 6167; F: 6473; G: 6485; H: 6484; I: 6480; J: 6476; K: 6482; L: 6313. Further details on the location of NWA populations can be found in Lia *et al*. (2009).

the allele size permutation tests of Hardy *et al*. (2003) (5000 randomisations). As a result, 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 & Cockerham, 1984), were calculated over all loci with Fstat program. The significance of F_{ST} values was tested by permuting genotypes among samples, rather than alleles, as this is the preferred method when HW proportions are rejected within samples, and was corrected for multiple comparisons using Bonferroni procedures at an overall $\alpha = 0.05$. Partitioning of genetic variation within and

among populations was further assessed by analysis of molecular variance (AMOVA; Excoffier *et al*., 1992; Excoffier, 2007) using GeneAlEx 6 (Peakall & Smouse, 2006); statistical significance of each variance component was assessed based upon 999 permutations of the data.

Population structure was also examined using the Bayesian model-based approach of Pritchard *et al*. (2000) implemented in STRUCTURE 2.2 (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 (or clusters) and the proportion of its genotype drawn from each of the K populations. Since the SSR loci selected here are independent, we only assumed that each K population follows HW proportions. NEA populations were evaluated with K ranging from 1 to 10. Then, eight NWA populations were added for comparison and the number of clusters also ranged from 1 to 10. The analyses were performed using 10 replicate runs *per* K value, a burn-in period length of 10^5 and a run length of 10^6. No prior information on the origin of individuals was used to define the clusters. All the analyses were run under the correlated allele frequency models available in STRUCTURE. This model assumes that frequencies in the different populations are likely to be similar probably because of shared ancestry or migration (Falush *et al*., 2003). The run showing the highest posterior probability of the data was considered for each K value. Evaluation of STRUCTURE results was conducted using an *ad hoc* criterion based on the rate of change in the log probability of the data between successive K values (ΔK) (Evanno *et al*., 2005).

Population structure was also examined by applying the discriminant analysis of principal components (DAPC) (Jombart *et al*., 2010), a new multivariate method designed to identify and describe clusters of genetically related individuals, which is free of assumptions about HW or linkage equilibrium. Briefly, the method relies on allele data transformation using principal component analysis (PCA) as a prior step to discriminant analysis (DA). DA defines a model in which genetic variation is partitioned into a between-group and a within-group component. Groups can be defined a priori (i.e. populations, collection sites, temporal affiliations, etc.) or can be inferred using first sequential K-means (Legendre and Legendre, 1998) and model selection. DAPC was performed using the adegenet package (Jombart, 2008) for the R 2.10.1 software (R development Core Team 2009). The function DAPC was executed using the following groups: (a) the populations of origin of each individual and (b) the clusters identified by K-means. The number of clusters was assessed using the function find.clusters, evaluating a range from 1 to 40. The optimal number of clusters was chosen on the basis of the lowest associated Bayesian information criterion (BIC). One hundred principal components were retained for data transformation.

Reynolds *et al*. (1983) genetic distances between populations were calculated using PowerMarker software version 3.25 (Liu & Muse, 2005). The resulting genetic distance matrix was imported into Splitstree4 (Huson & Bryant, 2006) and used to create an unrooted tree diagram applying the Neighbour-Joining algorithm (Saitou & Nei, 1987). Branch support was estimated by bootstrapping (1000 pseudoreplicates) with PowerMarker. In addition, alternative interpopulation relationships were visualised using Splitstree4 by generating a split graph with the NeighbourNet algorithm (Bryant & Moulton, 2004). This analysis was performed using the following settings: edge fitting as ordinary least squares; equal angle as the chosen splits transformation, least squares to modify weights and four maximum dimensions as the filtering option.

Results

A total of 336 individuals belonging to 12 NEA maize landraces were SSR genotyped. The SSR loci examined exhibited high reproducibility and the resulting data matrix showed on average 4% missing data. A total of 126 alleles, 31 of them private, were detected and the average number of alleles *per* locus was 10.5 (range 4–18). Considering the NEA dataset as a whole, gene diversity (H_e) was 0.632.

Estimates of observed heterozygosity (H_o), gene diversity (H_e), allelic richness (R_s), and average number of alleles *per* locus within populations (A) were calculated for each population (Table 2). H_o, H_e, R_s and A ranged from 0.218 to 0.469 (average 0.350), from 0.253 to 0.585 (average 0.467), from 1.78 to 3.27 (average 2.72) and from 2 to 5.25 (average 3.79), respectively. Overall, population 6618 exhibited the lowest levels of gene diversity for three of the four estimates, and population 6610 showed the highest values of R_s and A. For H_e and R_s indices, significant differences were found among populations [Kruskal–Wallis test, H_e: $H_{(\text{d.f.}=18;\ N=228)} = 41.75$, $P < 0.0012$; R_s: $H_{(\text{d.f.}=18;\ N=228)} = 45.79$, $P < 0.0003$]. To identify the source of these differences, *post hoc* comparisons among collection sites (villages) were performed, showing no significant differences for either of the indices. In contrast, *post hoc* comparisons between landraces revealed that the popcorn landrace Pipoca colorado has significantly lower values of H_e and R_s (Table 3).

When NWA populations were added into the analysis, A, H_e and R_s indices were recalculated considering only

Ann Appl Biol **160** (2012) 308–321 © 2012 The Authors
Annals of Applied Biology © 2012 Association of Applied Biologists

Table 2 Genetic variability in maize landraces from Northeastern Argentina

Landrace	Population	H_o	H_e	R_s	A	Private Alleles
Pipoca colorado	6596	0.305	0.300	1.83	**2.00**	0
Pipoca colorado	6618	**0.218**	**0.253**	**1.78**	2.42	4
Pipoca amarillo	6605	0.356	0.391	2.26	3.17	**5**
Pororó grande	6614	0.231	0.292	1.89	2.67	4
Pororó azul	6607	0.263	0.287	1.91	2.50	0
Blanco ancho	6586	**0.469**	0.496	2.90	4.00	0
Blanco ancho	6615	0.271	0.427	2.48	3.42	0
Blanco ancho	6602	0.355	0.549	3.05	3.92	0
Blanco angosto	Ba	0.384	0.555	3.20	5.08	3
Blanco angosto	6610	0.377	0.562	**3.27**	**5.25**	2
Tupí blanco	6592	0.321	0.452	2.63	3.50	2
Overo	6595	0.460	**0.585**	3.19	3.67	1
Overo	6617	0.439	0.544	3.17	4.17	1
Overo	6601	0.367	0.556	3.07	4.25	1
Colorado	Co	0.388	0.559	3.10	4.58	2
Colorado	6604	0.383	0.552	**3.27**	4.67	1
Azul	6597	0.399	0.467	2.56	3.83	0
Variegado	6585	0.265	0.497	2.94	4.42	2
Rosado	Ro	0.402	0.552	3.21	4.50	3

Extreme values of each index are indicated in bold.

Table 3 *P* values of *post hoc* comparisons of genetic diversity estimates between NEA landraces

	Blanco ancho	Blanco angosto	Overo	Colorado	Pipoca colorado
Blanco ancho	–	0.442	0.479	0.516	0.029*
Blanco angosto	0.665	–	0.872	0.929	0.001*
Overo	0.485	0.872	–	0.924	0.003*
Colorado	0.644	0.949	0.903	–	0.003*
Pipoca colorado	0.034*	0.014*	0.001*	0.006*	–

Above diagonal: R_s *P* values; below diagonal: H_e *P* values. *Post hoc* comparisons of H_e and R_S indices were performed among those landraces represented by more than one population.

the nine SSR common to both datasets. The total number of alleles was 135 and overall gene diversity was 0.719. The average A value was 3.86 for NEA populations and 5.76 for NWA populations. The mean H_e and R_s for NEA populations were 0.467 and 2.79 and for NWA populations were 0.611 and 3.57, respectively, with these values being significantly different ($P_{(He)} = 0.01$; $P_{(Rs)} = 0.004$). When the number of private alleles was re-assessed, NEA populations BA, 6592, 6595, Co, 6604, 6585, Ro and 6618 presented one private allele, whereas two private alleles were found in population 6618, making a total of nine. The total number of private alleles for NWA populations was 22, ranging from two to seven, with only two populations showing no private alleles.

Population structure and cluster analysis

Global fit to HW proportions was observed in 7 of the 19 NEA populations examined in this study (Table S1), the remaining 12 populations exhibited a global homozygote excess with different loci accounting for these deviations. Global estimates of F_{ST} revealed a statistically significant degree of differentiation among the populations (global $F_{ST} = 0.261$; C.I.$_{95\%} = 0.229–0.304$; C.I.$_{99\%} = 0.221–0.321$). The largest F_{ST} was found for population pair 6618–6614 ($F_{ST} = 0.507$), whereas the less differentiated pair was Co–6604 ($F_{ST} = 0.032$). Fifty-nine percent of F_{ST} pairwise comparisons were significant (Table S2). To examine the effect of collection sites on the partitioning of genetic variation, an AMOVA was performed considering collection sites (villages) as the uppermost hierarchical level. This analysis revealed that 66% of the genetic variation resides within populations, 30% is found among populations within each village and only 4% accounts for variation among villages ($P < 0.001$).

When NEA and NWA datasets were analysed together, global F_{ST} was 0.267 (C.I.$_{95\%} = 0.224–0.346$; C.I.$_{99\%} = 0.218–0.380$). Average pairwise F_{ST} between NEA and NWA populations was 0.283 (Table S2). The

AMOVA indicated that 69% of the genetic variation resides within populations, 10% accounts for variation between regions and 21% is found among populations within each region ($P < 0.001$).

Bayesian analysis of population structure using the model-based approach of Pritchard *et al.* (2000) provided further support for the existence of genetic structure among NEA populations. The inference of the number of gene pools was not straightforward. However, the log-likelihood values for data conditional on K, $\ln(X/K)$, seemed to reach a plateau at $K = 6$. The assessment of ΔK following the methods of Evanno *et al.* (2005) identified the modal value of this statistic at $K = 2$. At this K value, clustering of individuals is highly consistent with two informal categories delimited on the basis of endosperm consistence: popcorn and floury landraces. Average membership coefficients for both groups are >98%, except for population 6605, which shows a moderate proportion of admixed individuals (Fig. 2). It is noteworthy that at $K = 6$ all popcorn populations are still highly homogeneous (average membership coefficient ≥95%), with most individuals of a population being assigned to the same cluster. In contrast, floury populations seem to have received contributions from at least two major gene pools and their individuals are also highly admixed.

A second population structure analysis was carried out combining NEA and NWA datasets. Analysis of the log-likelihood values $\ln(X/K)$ did not allow the detection of an evident plateau. Notwithstanding, the assessment of ΔK identified the modal value of this statistic at $K = 3$; the three clusters corresponded to NEA popcorns, NEA floury and NWA populations (Fig. 3).

Results from DAPC were highly concordant with the groupings inferred by the model-based Bayesian approach. For the NEA dataset, a clear distinction between popcorn and floury landraces was observed using both the actual number of populations (19) and the number of clusters selected based on the K-means algorithm and BIC values (10) (Figs 4A and S1). In addition to assigning individuals to the different clusters, DAPC also provides a visual representation of between-group relationships, thus allowing recognition of the most divergent gene pools. In this case, population 6614 was remarkably differentiated not only from the floury populations but also from the remaining popcorn landraces. For the NEA–NWA dataset, the number of clusters selected based on the K-means algorithm was 15. The results were also in general agreement with the STRUCTURE analysis. Again, population 6614 is highly differentiated from the remaining populations. NWA populations are clustered together and occupy an intermediate position between NEA floury and popcorn populations, with NEA popcorn

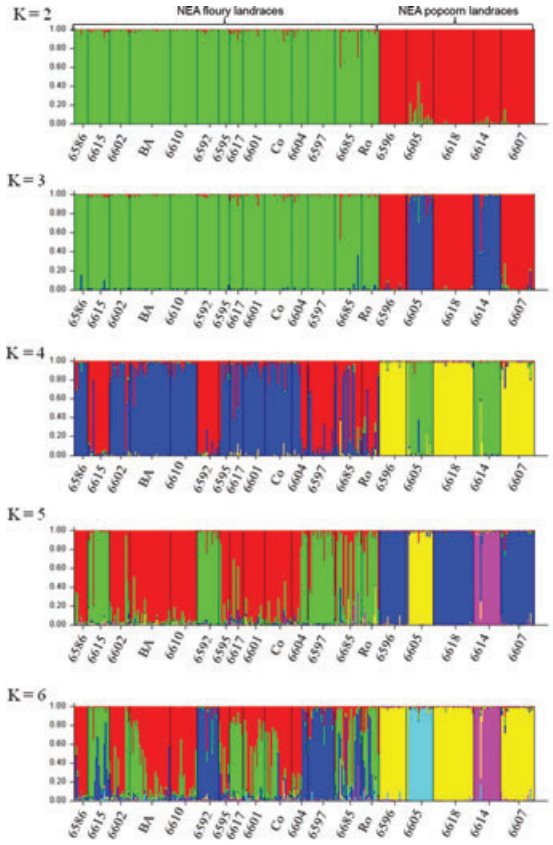

Figure 2 Estimated population structure of maize landraces from Northeastern Argentina. Each individual is represented by a thin vertical segment, which can be partitioned into K coloured segments that represent the individual estimated membership to K cluster. The run with the highest $\ln(X/K)$ was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

population 6605 showing a close association to the NWA group (Figs 4B and S2).

Clustering based on molecular distances shows that although most bootstrap values for internal branches were below 50%, two main partitions can be distinguished (Fig. 5A). The partition supported by the highest bootstrap value separates the NWA Andean Complex populations (6473, 6167, 6480, 6484, 6476 and 6485) from the remaining populations analysed; the second partition separates NEA popcorn populations 6605, 6596, 6618 and 6607 from all the others. All floury populations clustered together, whereas Pororó grande, Pisingallo and Orgullo Cuarentón show a central position in the unrooted NJ phylogram. The tree diagram shows a very good adjustment of the genetic distance data (fit = 93.229). The NeighbourNet, which is a visual representation of conflicting signals in the data presented as a series of parallel edges, shows highlighted partitions that

Ann Appl Biol **160** (2012) 308–321 © 2012 The Authors
Annals of Applied Biology © 2012 Association of Applied Biologists

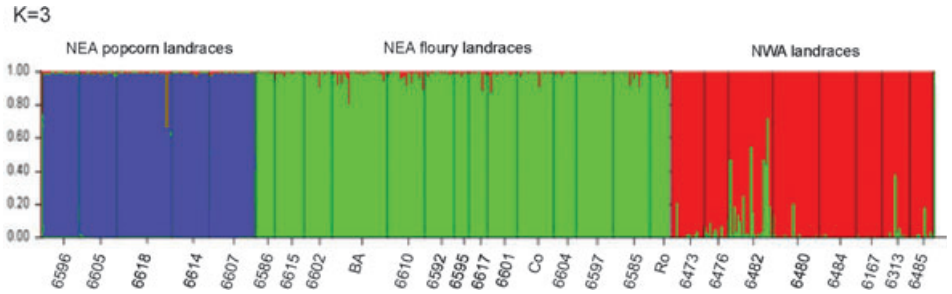

Figure 3 Estimated population structure of maize landraces from Northeastern and Northwestern Argentina. Each individual is represented by a thin vertical segment, which can be partitioned into *K* coloured segments that represent the individual estimated membership to *K* clusters. The run with the highest ln(*X*/*K*) was chosen for graphical representation of each *K*.

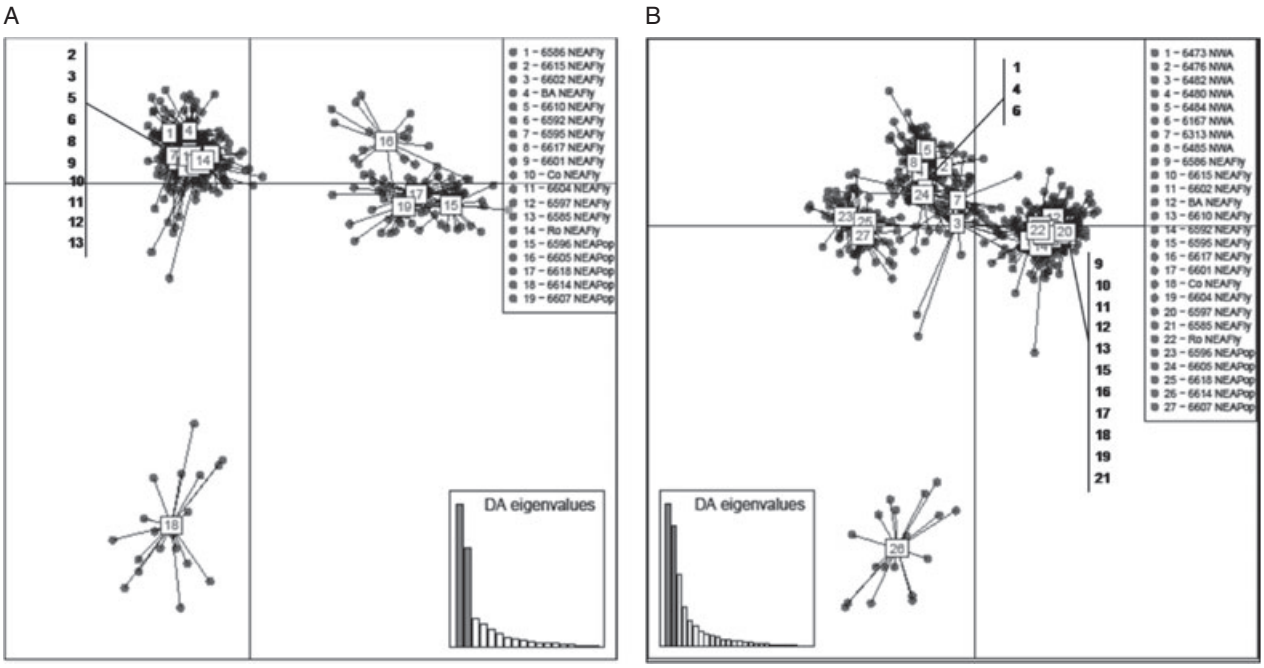

Figure 4 Scatterplots of the Discriminant Analysis of Principal Components (DAPC) using populations as prior clusters. (A) NEA datasets. (B) NEA-NWA datasets. Dots represent individual samples. Populations with overlapping centroids are listed within brackets. Insets indicate the number of principal components retained for DAPC. NEApop:NEA popcorn populations. NEAFly:NEA floury populations.

are highly concordant with the groupings identified by the NJ (fit = 95.609; Fig. 5B).

On the basis of the above results, diversity indices were calculated for each of the genetic clusters identified by Bayesian and multivariate analysis of population structure (i.e. NEA floury, NEA popcorn and NWA). Average values of diversity indices for NEA popcorns, NEA floury and NWA populations are presented in Table 4. When H_e and R_s comparisons were performed, significant differences were revealed between NEA popcorns and NWA landraces ($P_{(He)} = 0.001$; $P_{(Rs)} = 0.001$). Similar results were obtained from the comparison NEA popcorns versus NEA floury populations using the 12 loci dataset. The AMOVA indicated that 62% of the genetic variation resides within populations, 22% accounts for variation among the three groups and 16% is found among populations within each group, being these differences highly significant (P-value < 0.001).

Discussion

The present work focuses on one of the least studied areas of maize distribution, the lowland agro-ecosystems of southern South America. Over this region, maize landraces are cultivated under the influence of human and environmental selection, gene flow and genetic drift,

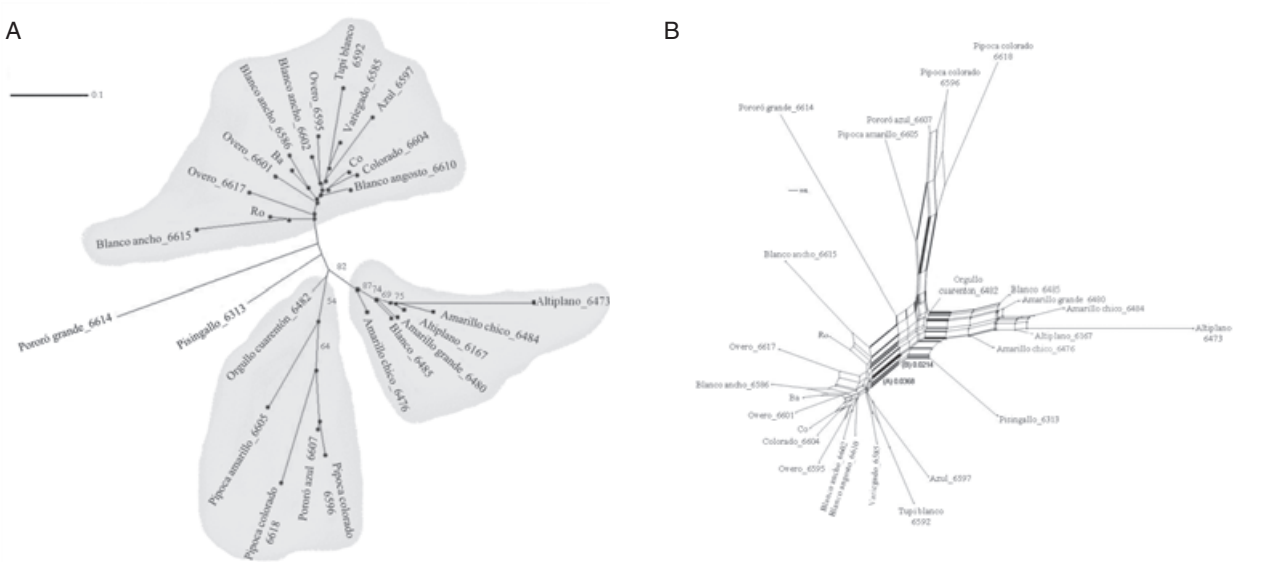

Figure 5 (A) Neighbour-joining network based on Reynolds *et al*. (1983) genetic distances between maize landraces from Northeastern and Northwestern Argentina. Bootstraps values ≥50% are indicated beside branches. NEA landraces are labelled with a point; NWA landraces are labelled with squares. (B) NeighbourNet graph based on interpopulation genetic distances. The length of the split is proportional to its weight. The weight of splits (A) and (B) are indicated.

Table 4 Average values of genetic diversity and F_{ST} indices for NEA and NWA landraces

	H_o	H_e	R_s	F_{ST}
NEA popcorns	0.300[a]	0.308[a]	1.961[a]	0.334[a]
NEA floury	0.372[b]	0.542[b]	3.087[b]	0.120[b]
NWA	0.539[b]	0.611[b]	3.572[b]	0.138[b]

a,b Groups designated by the same letter show no significant differences.

all of which interact to shape genetic diversity patterns. In this study, we assessed the genetic diversity and population dynamics of relictual maize landraces from Northern Argentina, as a necessary step towards a balanced management between consumption, improvement and conservation. These germplasm resources have been in existence since pre-Columbian times, but only recently have they been recognised as valuable reservoirs of genetic variability worth protecting.

The results presented here constitute an original contribution and differ from previous reports in two major respects. First, unlike studies that document maize genetic variability conserved in seed banks, this work surveys the genetic diversity of the germplasm currently used by aboriginal farmers. Second, the analysis is based on a microgeographic approach, which allows both to quantify genetic diversity and to assess within region population structure.

Being the area under scrutiny one of the southernmost limits of maize cultivation, it is remarkable that the levels of microsatellite variability found for NEA landraces resulted similar to those obtained from other areas of the American continent. Overall, NEA landraces exhibited a mean number of alleles *per* locus (10.5) which is within the range reported for Mexican landraces (7.84–14.9; Reif *et al*., 2006; Vigouroux *et al*., 2008) and for both the Andean and Northern United States clusters delimited by Vigouroux *et al*. (2008) (12.4 and 10.6, respectively). On a regional scale, genetic diversity was approximately 80% of the values reported for highland Mexico races (H_e: 0.814) and tropical lowland races (H_e: 0.803) and it is approximately 90% of the values obtained for the landraces of the Andean (H_e: 0.706) and Northern United States (H_e: 0.718) clusters delimited by Vigouroux *et al*. (2008). However, it should be borne in mind that the regional estimates of Vigouroux *et al*. (2008) are based on a much more extensive geographic range, with a sampling strategy that maximised detection of genetic variation. In this context, the levels of genetic diversity obtained for NEA landraces are remarkably high, particularly considering the restricted geographic area studied and the comparatively small number of landraces analysed.

At the population level, the mean number of alleles (A) and average gene diversity (H_e) obtained for NEA populations were similar to those reported for different racial complexes from tropical, subtropical and temperate maize germplasm preserved at CIMMYT (A: 3.76; H_e: 0.52; Reif *et al*., 2004) and for accessions of Mexican

landraces (A: 3.44, H_e: 0.48; Reif *et al.*, 2006). In contrast, the average gene diversity of NEA populations was approximately a third lower than the estimates obtained for landraces from traditional smallholders from the Central Valleys of Oaxaca (H_e: 0.71; Pressoir & Berthaud, 2004).

There are approximately 70 landraces described for Northern Argentina, the vast majority of which (80%) grow in NWA (Cámara Hernández & Miante Alzogaray, 1997, 2003; Melchiorre *et al.*, 2006). In agreement with the centres of diversity previously proposed by Vavilov, the distinction between NEA and NWA gene pools is supported by all the methods used here for the analysis of population structure. Surprisingly, our results indicate that the levels of genetic variability for NEA and NWA regions are not as dissimilar as expected based on the morphological variation registered. For instance, the average R_S and A obtained for NEA landraces account for 72% and 91% of the variability observed in NWA landraces, respectively. Interestingly, the floury-type landraces of NEA exhibited levels of variability similar to those of NWA populations (Table 4).

Within the NEA region, the site of collection (village) of the different maize populations was found to have only a minor influence on the distribution of the genetic diversity. Similar levels of variation were found among the villages from which samples were collected, and only a small proportion of the total genetic variation was assigned to differences among villages. Conversely, comparison of genetic diversity estimates between the two groups of landraces detected by population structure analysis (floury versus popcorn) revealed that variability in popcorn populations was slightly more than half of that observed in the floury counterpart (Table 4). These differences could stem from a differential rate of genetic erosion, with popcorns suffering more severe bottlenecks or more stringent selective regimes because of farmer preferences and/or adverse environmental conditions. Alternatively, their ancestral levels of variability may have been limited in comparison with those of the floury group.

Being maize a wind-pollinated annual crop considered to be predominantly outcrossing, it could be safe to expect random mating in most populations. However, a large proportion of NEA populations (63%) did not fit HW expectations showing global homozygote excess, with different loci accounting for the observed deviations. Although contradicting theoretical expectations, many studies have reported deviations from panmixia in maize germplasm (Kahler *et al.*, 1986; Dubreuil & Charcosset, 1998; Pressoir & Berthaud, 2004; Reif *et al.*, 2004; Lia *et al.*, 2009). Experimental errors, null alleles, selection, population substructuring and assortative mating are often invoked to explain homozygote excess or deficit in allogamous species. It is noteworthy that only one out of five popcorn populations (i.e. 6614) did not fit HW proportions, while deviations were apparent for 11 of 14 floury populations. Considering the assignments from the Bayesian analysis, these deviations are most likely caused by the population substructuring. The cluster of landraces having floury endosperm showed contributions from more than one gene pool and higher levels of individual admixture, contrasting with the homogeneity evidenced by the group of landraces with popping capacity (Fig. 2).

A multivariate analysis of morphological, phenological and reproductive traits was performed on a set of NEA landraces, including most of the nominal races studied here (Melchiorre *et al.*, 2006). In that study two groups of landraces were distinguished with the length of the vegetative cycle, and morphological traits associated to it (e.g. height of the plant, number of leaves, etc.), being the most important discriminant characters. In agreement with the results from population structure analysis, these two groups correspond to popcorn and floury landraces, further supporting the distinction between these categories.

Maize population dynamics is inextricably linked to human activities. Farmers' management practises greatly influence the distribution of genetic diversity and gene flow. As observed for the estimates of diversity, the distribution of the variability also seems to be different between floury and popcorn landraces. Global F_{ST} is higher in popcorn than in floury populations, with the popcorn estimate being even higher than that of the NWA populations (Table 4). Moreover, the Bayesian analysis showed that at $K = 6$ popcorn populations are subdivided into three clearly delimited groups, which are strongly concordant with the villages from which they were collected. In contrast, no evident pattern emerges from the assignments of the more heterogeneous floury landraces. These observations, along with the lower variability levels of popcorn populations, suggest that genetic drift plays a prominent role in shaping the genetic structure of NEA popcorns. Results presented here concerning population 6614 of Pororó grande suggest that it would be interesting to study more populations of this race in order to acknowledge the origin of such pronounced differentiation.

On the other hand, the heterogeneous gene pool composition inferred for the floury group is likely to be a consequence of biological determinants and farmers' management practises. The *Guaraní's* practises of planting several landraces closely mixed in small fields, should

lead to extensive gene flow among landraces. Nevertheless, the significant levels of differentiation found here among landraces suggest the existence of biological and/or human barriers to gene flow, particularly among popcorns. Two, non-exclusive main explanations for the persistence of differentiated landraces can be proposed. First, biological barriers such as differences in flowering time, or any kind of yet unidentified incompatibility, may minimize gene flow among landraces. Second, farmer's selection for particular combinations of morphological traits in their choice of seeds for sowing the next season may preserve the identity of landraces.

On a regional scale, several factors may act to explain the limited extent of gene flow between NEA and NWA regions. The occurrence of two separate expansions of maize cultivation into South America, one through the highland and the other through the lowland regions of the continent (McClintock *et al*., 1981; Oliveira Freitas *et al*., 2003), may have imposed historical barriers and prevented genetic homogenisation between these two areas. Alternatively, different cultural traditions and usage of landraces, specific ecological adaptations, and a relative isolation of the *Guaraní* communities that preserve maize landraces in the province of Misiones, may have contributed to the present differentiation. A more extensive sampling of the NEA region is still needed to determine whether the observed differentiation levels are confined to the populations studied here or if they can be thought of as the consequence of a more general pattern.

In conclusion, the landraces examined here have shown to harbour considerable levels of genetic diversity with contributions from different gene pools. From a conservation perspective, the quantification of genetic diversity within and among NEA landraces allows to assess a population's potential to adapt to fluctuating and heterogeneous environments, thus assisting in the design of protection strategies. Moreover, as shown by the results presented here, this germplasm constitutes a valuable source of novel alleles to broaden the genetic basis of breeding programmes. Further studies encompassing a larger number of populations from the other provinces of the NEA region will certainly help to detect additional genetic variation, which may prove highly valuable.

Conservation endeavours encompass the genetic management of small populations to maximise retention of genetic diversity and minimize inbreeding, as well as the delineation of management units. We believe that conservation programmes should prioritize the preservation of maize landraces from the two major regions of Northern Argentina, NEA and NWA, as different units. Within the NEA region, the two landrace groups distinguished in this work, floury and popcorn landraces, should also deserve separate management.

Acknowledgements

The authors wish to thank the *Guaraní* communities for providing their maize landraces. Several grants from the Consejo Nacional de Investigaciones Científicas y Técnicas (PIP 0342) and the Universidad de Buenos Aires (EX178) are gratefully acknowledged.

References

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

Bretting P.K., Goodman M.M., Stuber C.W. (1987) Kariological and isozyme variation in West Indian and allied American landraces of maize. *American Journal of Botany*, **74**, 1601–1613.

Bretting P.K., Goodman M.M., Stuber C.W. (1990) Isozymatic variation in Guatemalan races of maize. *American Journal of Botany*, **77**, 211–225.

Bryant D., Moulton V. (2004) Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution*, **21**, 255–265.

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

Cámara Hernández J., Miante Alzogaray A.M. (2003) Caracterización y clasificación en razas, de maíces nativos de la Provincia de Misiones, Argentina. In Proceedings of the IV Simposio de recursos genéticos para América Latina y el Caribe, Mar del Plata, Argentina.

Chiavarino, A.M. (1997) *Cromosomas B: Meiosis y Herencia en Razas Nativas de Maíz*. Buenos Aires: Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, p. 143.

Chiavarino A.M., Rosato M., Manzanero S., Jimenez G., Gonzalez-Sanchez M., Puertas M.J. (2000) Chromosome nondisjunction and instabilities in tapetal cells are affected by B chromosomes in maize. *Genetics*, **155**, 889–897.

Chiavarino A.M., González-Sánchez M., Poggio L, Puertas M.J., Rosato M., Rosi P. (2001) Is maize B chromosome preferential fertilization controlled by a single gene? *Heredity*, **86**, 743–748.

Doebley J. (1990) Molecular evidence and the evolution of maize. *Economic Botany*, **44**, 7–25.

Doebley J., Goodman M.M. (1984) Isoenzymatic variation in *Zea* (Gramineae). *Systematic Botany*, **9**, 203–218.

Doebley J., Goodman M.M., Stuber C.W. (1983) Isozyme variation in maize from the southwestern United States taxonomic and anthropological implications. *Maydica*, **28**, 94–120.

Doebley J., Goodman M.M., Stuber C.W. (1985) Isozyme variation in the races of maize from Mexico. *American Journal of Botany*, **72**, 629–639.

Doebley J., Goodman M.M., Stuber C.W. (1986) Exceptional genetic divergence of Northern Flint Corn. *American Journal of Botany*, **73**, 64–69.

Doebley J., Goodman M.M., Stuber C.W. (1987) Patterns of isozyme variation between maize and Mexican Annual teosinte. *Economic Botany*, **41**, 234–246.

Doebley J., Wendel J.F., Smith J.S.C., Stuber C.W., Goodman M.M. (1988) The origin of cornbelt maize: the isozyme evidence. *Economic Botany*, **42**, 120–131.

Dubreuil P., Charcosset A. (1998) Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. *Theoretical and Applied Genetics*, **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. *Theoretical and Applied Genetics*, **92**, 832–839.

van Etten J., de Bruin S. (2007) Regional and local maize seed exchange and replacement in the western highlands of Guatemala. *Plant Genetic Resources*, **5**, 57–70.

Evanno G., Regnaut S., Goudet J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L. (2007) Analysis of population subdivision. In D. Balding, M. Bishop, C. Cannings eds. *Handbook of Statistical Genetics*, 3rd Edn, pp. 980–1020. New York, USA: Wiley.

Excoffier L., Smouse P., Quattro J. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

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

Gaggiotti O.E., Lange O., Rassmann K., Gliddon C. (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*, **8**, 1513–1520.

Gonzalez G.E. (2004) Afinidades genómicas y mapeo cromosómico en maíz y especies relacionadas a través de estudios de citogenética clásica y de hibridación in situ. Doctoral Thesis. Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires.

Goodman M.M., Stuber C.W. (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. *Journal of Heredity*, **86**, 485–486.

Goudet J. (2001) *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)*. URL http://www2.unil.ch/popgen/.

Hardy O.J., Charbonnel N., Freville H., Heuertz M. (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, **163**, 1467–1482.

van Heerwaarden J., Doebley J., Briggs W.H., Glaubitz J.C., Goodman M.M., Sanchez Gonzalez J.J., Ross-Ibarra J. (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 1088–1092.

Huson D.H., Bryant D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jombart T. (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T., Devillard S., Balloux F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Kahler A.L., Hallauer A.R., Gardner C. (1986) Allozyme polymorphism within and among open-pollinated and adapted exotic populations of maize. *Theoretical and Applied Genetics*, **72**, 592–601.

Latournerie L., Tuxill J., Yupit E., Arias L., Cristobal J., Jarvis D.I. (2006) Traditional maize storage methods of Mayan farmers in Yucatan, Mexico: implications for seed selection and crop diversity. *Biodiversity and Conservation*, **15**, 1771–1795.

Legendre P., Legendre L. (1998) *Numerical Ecology*. 2nd English Edition edn. Amsterdam, The Netherlands: Elsevier.

Lia V.V., Poggio L., Confalonieri V.A. (2009) Microsatellite variation in maize landraces from Northwestern Argentina: genetic diversity, population structure and racial affiliations. *Theoretical and Applied Genetics*, **119**, 1053–1067.

Liu K., Muse S.V. (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, **21**, 2128–2129.

Martínez-Crovetto R. (1968) La alimentación entre los indios guaraníes de Misiones (República Argentina). *Etnobiológica*, **4**, 1–24.

Matsuoka Y., Vigouroux Y., Goodman M.M., Sánchez G.J., Buckler E.S., Doebley J. (2002a) A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 6080–6084.

Matsuoka Y., Mitchell S.E., Kresovich S., Goodman M.M., Doebley J. (2002b) Microsatellites in *Zea* – variability, patterns of mutations, and use for evolutionary studies. *Theoretical and Applied Genetics*, **104**, 436–450.

McClintock B., Kato T.A., 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, Mexico. 521 p.

Melchiorre P., Bartoloni N., Cámara H.J. (2006) Phenetic relationships among native races of maize (*Zea mays* ssp. *mays*) from North-eastern Argentina (Misiones). *Journal of Genetics and Breeding*, **60**, 173–182.

Nei M. (1987) *Molecular Evolutionary Genetics*. New York, USA: Columbia University Press.

Oliveira Freitas F., Bendel G., Allaby R.G., Brown T. (2003) DNA from primitive maize landraces and archaeological remains: implications for the domestication of maize and its expansion into South America. *Journal of Archaeological Science*, **30**, 901–908.

Peakall R., Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Perales R.H., Brush S.B., Qualset C.O. (2003) Dynamic management of maize landraces in central Mexico. *Economic Botany*, **7**, 21–34.

Poggio L., Molina M.C., Naranjo C.A. (1990) Cytogenetic studies in the genus *Zea*. 2. Colchicine induced multivalents. *Theoretical and Applied Genetics*, **79**, 461–464.

Poggio L., Rosato M., Chiavarino A.M., Naranjo C.A. (1998) Genome size and environmental correlations in maize (*Zea mays* ssp. *mays*, Poaceae). *Annals of Botany*, **82**, 107–115.

Poggio L., Gonzalez G., Confalonieri V.A., Comas C., Naranjo C.A. (2005) The genome organization and diversification of maize and its allies revisited: evidences from classical and FISH-GISH cytogenetic analysis. *Cytogenetics and Genome Research*, **109**, 259–267.

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 J.K., Stephens M., Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

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. *Theoretical and Applied Genetics*, **106**, 895–903.

Reif J.C., Xia X.C., Melchinger A.E., Warburton M.L., Hoisington D.A., 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 Science*, **44**, 326–334.

Reif J.C., Warburton M.L., Xia X.C., Hoisington D.A., Crossa J., Taba S., Muminovic J., Bohn M., Frisch M., Melchinger A.E. (2006) Grouping of accessions of Mexican races of maize revisited with SSR markers. *Theoretical and Applied Genetics*, **113**, 177–185.

Reynolds J., Weir B.S., Cockerham C.C. (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics*, **105**, 767–779.

Rice W. (1989) Analysing tables of statistical tests. *Evolution*, **43**, 223–225.

Saitou N., Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Sánchez G.J., Stuber J.C.W., Goodman M.M. (2000) Isozymatic diversity in the races of maize of the Americas. *Maydica*, **45**, 185–203.

Slatkin M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Vavilov N.I. (1992) *Origin and Geography of Cultivated Plants*, pp. 96. Great Britain: English version: Cambridge University Press.

Viglizzo E.F., Ricard M.F., Jobbágy E.G., Frank F.C., Carreño L.V. (2011) Assessing the cross-scale impact of 50 years of agricultural transformation in Argentina. *Field Crops Research*, **124**, 186–194.

Vigouroux Y., Glaubitz J.C., Matsuoka Y., Goodman M.M., Sánchez G.J., Doebley J. (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *American Journal of Botany*, **95**, 1240–1253.

Waits L.P., Luikart G., Taberlet P. (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology*, **10**, 249–256.

Weir B., Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wright S. (1978) *Variability Within and Among Natural Populations in Evolution and the Genetics of Populations*. Chicago, IL, USA: University of Chicago Press.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 F_{IS} *per* locus and population.

Table S2 Population pairwise comparisons. Above diagonal: pairwise genetic differentiation index (F_{ST}). Below diagonal: Reynolds *et al*. (1983) genetic distances.

Figure S1 (A) Scatterplot of the discriminant analysis of principal components (DAPC) for the NEA dataset using the 10 clusters identified by the *K*-means algorithm. Dots represent individual samples. Clusters composed of NEA popcorn populations are coloured with the red palette, clusters composed of NEA floury populations are coloured with the blue palette. Insets indicate the number of principal components retained for DAPC. (B) Cluster assignment according to the *K*-means algorithm.

Ann Appl Biol **160** (2012) 308–321 © 2012 The Authors
Annals of Applied Biology © 2012 Association of Applied Biologists

Figure S2 (A) Scatterplot of the discriminant analysis of principal components (DAPC) for the NEA–NWA dataset using the 15 clusters identified by the *K*-means algorithm. Dots represent individual samples. Clusters composed of NEA popcorn populations are coloured with the red palette, clusters mainly composed of NEA floury populations are coloured with the blue palette, clusters mainly composed of NWA populations are coloured with the green palette. Insets indicate the number of principal components retained for DAPC. (B) Cluster assignment according to the *K*-means algorithm.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.