

# Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico

G Pressoir<sup>1,2</sup> and J Berthaud<sup>1,2</sup>

<sup>1</sup>International Maize and Wheat Improvement Center (CIMMYT), Mexico, D.F. Mexico; <sup>2</sup>Institut de Recherche pour le Développement (IRD), Montpellier, France

Assessing the impact of farmer management of maize landraces in the Central Valleys of Oaxaca, Mexico is crucial to an understanding of maize evolution, as it was first domesticated there. In this paper, we report on the impact of traditional farmer management of maize populations in this region in structuring molecular diversity and on the population dynamics of maize landraces. These populations, from a sample of local landraces cultivated by farmers in six villages, show little among-population differentiation ( $F_{st}=0.011$ ). Most surprisingly, there is no isolation by distance and small among-village differentiation ( $F_{st}=0.003$ ). For an outbreeding plant such as maize, one would expect populations to fit Hardy-Weinberg equilibrium, but significant homozygote

excess ( $F_{is}=0.13$ ) was found. This homozygote excess shows remarkable interpopulation and interlocus differences. We show that this pattern is related to variation in the mean anthesis-silking interval as well as to the flowering range or heterogeneity in flowering of a given population. A short anthesis-silking interval and high level of heterogeneity in flowering precocity will favor assortative mating. This leads to a locus-dependent population substructure giving an unusual case of Wahlund effect and inbreeding while high levels of seed exchange among farmers prevent population differentiation at both village and regional levels.

*Heredity* (2004) **92**, 88–94, advance online publication, 3 December 2003; doi:10.1038/sj.hdy.6800387

**Keywords:** assortative mating; flowering heterogeneity;  $F_{st}$ ; inbreeding; maize landraces; population structure

## Introduction

Little research has been directed towards understanding the evolutionary process behind plant evolution in agricultural ecosystems in traditional farming communities. Maize populations, like natural populations, are subject to migration and drift, to both natural selection and farmers' selection, and finally to local extinction and recolonization processes. Therefore, in order to gain a complete idea of the system, it is important to document in detail the genetic dynamics in these farmers' fields and the impact of their various practices.

Maize diversified first in the highlands of Mexico soon after domestication. Matsuoka *et al* (2002) show that the domestication of maize is based on a unique event, and that maize accessions from the highlands of Oaxaca are genetically the closest to the wild ancestors of maize (*Zea mays* ssp. *parviglumis*). They report that the basal-most maize in a phylogenetic analysis, including maize and its wild relative teosinte, are maize accessions collected in regions close to the Central Valleys of Oaxaca. Furthermore, archeological work (Benz, 2001; Piperno and Flannery, 2001) has revealed remains of the oldest known maize in the State of Oaxaca, at the Guila Naquitz cave, dating back to 4200 B.C. The Oaxaca region has also been reported to hold a large amount of variation in terms of phenotypic diversity (Bellon *et al*, 2003).

Today in Oaxaca, farmers still cultivate maize populations in a traditional manner. Saving seed from one season to the next is a well-defined practice. In addition, the Central Valleys of Oaxaca show very little presence of, or impact from, modern varieties (Smale *et al*, 1999; Bellon *et al*, 2003), with most of the area still planted in local landraces. Therefore, this region offers unique conditions for the study of the evolutionary processes that are key to maize evolution. The concept of 'landrace' is complex (Zeven, 1998) and its complete definition remains an issue of contention among some parties. For the purposes of this paper, the term landrace refers to a maize population cultivated in a traditional fashion and managed by a single farmer.

Small-scale Mexican farmer management practices are central to the evolution of maize and its diversity. Key practices include the planting of numerous maize populations within a small area. Consequently, even if desired, farmers are incapable of preventing the exchange of pollen between populations (Bellon and Brush, 1994; Louette *et al*, 1997). Furthermore, Mexican farmers commonly acquire seed from both local and distant farmers or sources, often mixing their seed with seed from other farmers or that purchased from markets (Louette and Smale, 2000).

In addition to farmer management, the biology of the species is expected to play a major role in structuring maize populations. Maize is a monoecious species that bears two inflorescence types, the staminate tassel and the pistillate ear. There is generally a delay between male and female flowering, with male flowering generally occurring before female flowering. If the delay between male and female flowering is short enough, it could

allow assortative mating to occur. Homozygous excess, and its variation across loci and populations, has been reported for maize open-pollinated populations using both isozymes and RFLP markers (Brown and Allard, 1970; Kahler *et al.*, 1986; Salanoubat and Pernes, 1986; Lefort-Buson *et al.*, 1991; Garnier-Géhé, 1992; Dubreuil and Charcosset, 1998). Assortative mating would produce a locus-dependent Wahlund effect (Nevo *et al.*, 2000), which was tested in the research reported herein.

In this paper, we assess the genetic diversity and population structure of maize landraces from the Central Valleys of Oaxaca. Because 83% of the total maize in this area is white kernel maize (Smale *et al.*, 1999), this study focuses on this maize type. We assess key agroecological factors and components of farmer management behind maize population dynamics. We also describe how this genetic diversity is structured, how management practices by farmers have an effect on population differentiation, and effects related to distance and seed exchange among Oaxacan farmers. In addition, we investigate the consequences of some flowering traits that could play a role in assortative mating in maize landrace populations. We use markers specifically linked to the QTLs of flowering traits, and other markers scattered throughout the genome.

## Material and methods

### Survey on farmer management

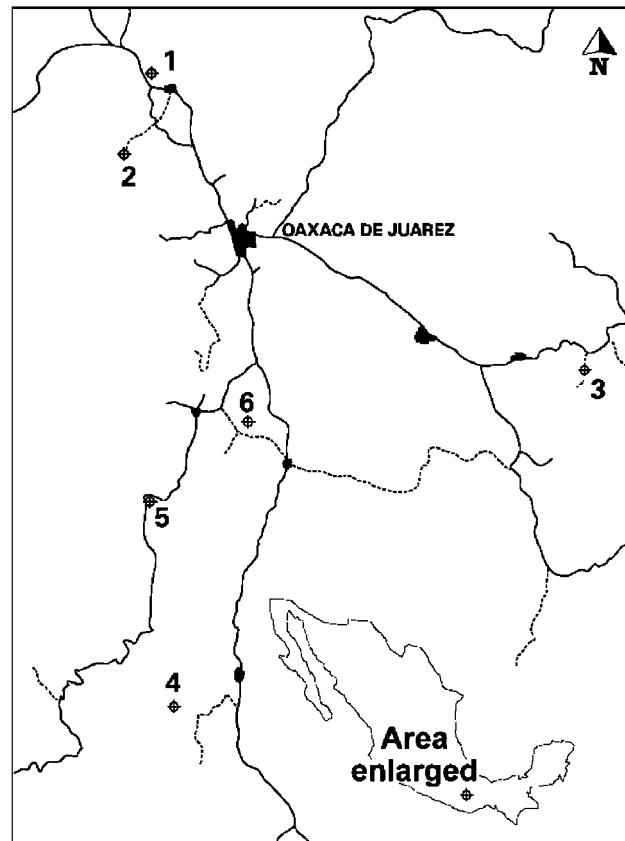
We selected six villages (Figure 1) for this study, because of their contrasting situations, in terms of ethnicity, and their maize production potential, based on an earlier study (Smale *et al.*, 1999; Bellon *et al.*, 2003). In all, 10 households per village were randomly chosen, giving a total of 60 farmers surveyed. We gathered information on farmer seed management and on seed exchange practices between and within villages.

### Material used for the genetic analysis

Of the six studied villages, Santa Ana and Huitzo showed the highest contrast in ethnicity, number of maize populations per farmer, and potential in terms of maize production (Smale *et al.*, 1999), and were therefore studied more extensively. A total of 31 populations were assayed, including field evaluation and genotyping. These included nine populations from Huitzo, three from Mazaltepec, three from San Lorenzo, three from Amatengo, three from Valdeflores, and 10 from Santa Ana. Households were selected randomly among the farmers cultivating a population of white kernel maize within each village. Sampling was carried out within a single generation. We randomly selected 20 open-pollinated families for each population and genotyped one individual per family; field evaluation for flowering traits was carried out on 18 of these families.

### Simple sequence repeat genotyping

A total of 11 microsatellite markers were assayed. SSR primers were selected from the maizeDB database of public SSRs, and included the following: phi011, phi227562, phi96100, phi101049, phi029, phi093, phi024, phi452693, phi034, phi014, and umc1061. They consist of tri- or tetranucleotide repeats. Markers were selected according to their chromosomal locations, in



**Figure 1** Location of sampled villages in the Central Valleys of Oaxaca around the city of Oaxaca de Juarez (Mexico). Villages are numbered from 1 to 6. Altitude, latitude, and longitude are also given: 1. Huitzo, 1730 masl (meters above sea level) 17°15'N 96°51'W; 2. Mazaltepec, 1700 masl 17°06'N 96°52'W; 3. San Lorenzo, 1830 masl 16°51'N 96°16'W; 4. Amatengo, 1310 masl 16°30'N 96°47'W; 5. Valdeflores, 1447 masl 16°45'N 96°49'W; 6. Santa Ana, 1520 masl 16°50'N 96°42'W.

order to provide for genome-wide coverage, and also by the size of the amplification product, to allow multiplexing on an automated DNA sequencer. Sequences and mapping positions can be downloaded at <http://www.agron.missouri.edu/ssr.html>. Of the 11 microsatellites used for this study, most did not exhibit a stepwise variation. Three of these markers (phi011, phi024, and phi452693) map close to genes or QTLs involved in flowering time or anthesis-silking interval (Veldboom *et al.*, 1994; Ribaut *et al.*, 1996; Gale and Devos, 1998; Thornsberry *et al.*, 2001).

### Cytoplasm genotyping

All populations were also characterized with chloroplastic markers. We used a polymorphic chloroplast set of primers surrounding a polyA repeat in the *psbK/psbI* intergenic region with the following sequences: zmcP7430-F: CGAAGCTGCTGTAAGTTTCG and zmcP7430-R: AA-GACTTCTCGGCTCTTATCCA (Provan *et al.*, 1999).

### Analysis of the among-population genetic structure

Because villages were not randomly selected, the genetic description given is specific to our sample. Overall,  $F_{st} = \theta$  (Weir and Cockerham, 1984) was calculated for

the entire set of 31 populations. Jackknifing over populations and loci was used to provide a confidence interval, according to Weir (1996).  $\theta$  values were estimated using GDA 1.1 software (Lewis and Zaykin, 2002), which performs hierarchical F-statistics. Different levels of population subdivision were tested as suggested in Weir (1996). For a random mating population (within sample) or random distribution of individuals (between samples), F-statistics are expected to be null.

A matrix of pairwise  $F_{st}/(1-F_{st})$  was estimated as well as a matrix of geographic distances  $\ln(\text{Dist})$  (Rousset, 1997) between villages (geographic coordinates provided in Figure 1) to test for isolation by distance. A Mantel test (Sokal and Rohlf, 1995) was used to test for the independence of the matrices.

The ratio of pollen to seed flow, according to Ennos (1994) is:

$$r = \frac{(1/F_{STm})(1 + F_{IS}) - 2(1/F_{STc})}{(1/F_{STc})}$$

where  $F_{STm}$  is the  $F_{ST}$  calculated for microsatellite markers and  $F_{STc}$  is calculated for cytoplasmic markers.

#### Analysis of the genetic variation within populations

We estimated the expected unbiased heterozygosity (Nei, 1987)  $H_e = 2n(1 - \sum_i p_i^2)/(2n-1)$  for microsatellite markers and  $H_e = n(1 - \sum_i p_i^2)/(n-1)$  for cytoplasmic markers. Homozygote excess was estimated according to Weir and Cockerham (1984) within population and for the whole set of populations. For pairwise linkage disequilibrium among loci, the within-population correlation coefficient  $R$  (Weir, 1979) was calculated and tested by permuting genotypes within locus within population using the GENETIX 4.02 software (Belkhir *et al.*, 2001).

#### Within-population genetic variation in relation to temporality in flowering

In order to characterize populations for flowering, 18 open-pollinated families were sampled for each of the 31 populations. Up to 12 progenies per family were evaluated (giving a total of  $31 \times 18 \times 12 = 6888$  plants evaluated). Field layout was a two-replicate design with hierarchical structure (population plots randomly assigned and family plots randomly assigned within populations). The experiment was carried out at the CIMMYT experimental station at El Batán, Texcoco, Mexico. Days to silking (DS) and days to anthesis (DA) were assessed. Measurements were carried out under well-watered conditions and were averaged over family. Anthesis-silking interval (ASI) results from the operation  $ASI = DS - DA$ . Flowering range (FLOR) was estimated for each population  $FLOR = DS_{\max} - DS_{\min}$  where  $DS_{\max}$  and  $DS_{\min}$  are the maximum and minimum DS in a given population. The flowering range within a given population is an estimate of the heterogeneity in flowering within this population.

Genetic structure within populations was assessed by temporal autocorrelation analysis in the same manner as spatial autocorrelation analysis (Hardy and Vekemans, 1999). We used an estimate of Wright's coefficient of relationship,  $\rho_{ij}$ , between pairs of individuals (Hardy and Vekemans, 1999) that corresponds to Moran's  $I$ -statistic using individual allele frequencies (Dewey and Heywood, 1988). The genotype of each individual (one

individual per family) was used as the variable of interest to calculate  $\rho_{ij}$ . The family mean flowering time was used in the same way as other authors have used spatial coordinates (Hardy and Vekemans, 2001). Therefore, we define as the divergence in flowering time between pairs of families  $\text{Div} = |DA_1 - DA_2|$  where  $DA_1$  and  $DA_2$  are the DA mean values of families 1 and 2, respectively. Regression analysis of  $\rho_{ij}$  on  $\text{Div}$  was carried out using the SPAGeDi 1.0 software (Hardy and Vekemans, 2002). The probabilities under the hypothesis that there is no relation between  $\text{Div}$  and  $\rho_{ij}$ ,  $\Pr$  values for  $|\text{obs}| = \exp$ , were estimated after 10 000 random permutations of temporal locations. It is equivalent to carrying out a Mantel test.

## Results

#### Farmer management of maize populations

Of the 60 farmers surveyed, six reported having participated in seed exchange between different villages during the last 10 years. Of these six reported cases, four correspond to the foundation of a new population and two to the mixing of seeds of the pre-existing population with seeds from another village. In all, 12 of the 60 farmers also reported that during the last 10 years they have mixed the seed of the pre-existing population with seed provided by another farmer from the same village, in order to provide enough seed for the next generation.

#### Principal coordinate analysis for molecular markers

All tested SSR loci were polymorphic in all populations. Principal coordinate analysis provides little evidence for population differentiation. Principal coordinates one and two together explain less than 6% of the total variation. Projection of the populations over the first two planes shows a uniform distribution and no grouping of populations. Furthermore, all populations show an overlapping of their distribution. Microsatellite polymorphism appears to be continuous and not related to geographic origin of the studied populations.

#### Among-population genetic structure

$F_{st}$  values obtained both for nuclear and cytoplasmic (chloroplast) markers are indicated in Table 1. All populations, even those separated by up to 100 km, were found to share chloroplast DNA haplotypes. There was no statistical significance for isolation by distance for either microsatellites or cytoplasmic markers. Furthermore, we observed low among-village  $F_{st}$  values, significantly lower than among-population  $F_{st}$  values. This could be explained by long distance gene flow (ie, seed exchange between villages). Supporting the predominance of seed *vs* pollen flow is an estimate of the ratio of pollen to seed flow (Ennos, 1994), which is inferior to 1 with  $r = 0.55$  where  $F_{STc} = 0.028$  and  $F_{STn} = 0.011$ . To estimate  $r$ , we did not consider among-village differentiation; we considered the inbreeding coefficient  $F_{is} = 0$ , as we will show later in this paper that the observed excess of homozygotes does not correspond to consanguineous mating.

#### Genetic variation within populations

Genetic diversity estimated over all populations is  $H_e = 0.71$  for microsatellite markers and  $H_e = 0.49$  for

**Table 1** Among-population genetic structure

		Within-village differentiation	Among-village differentiation
$\theta$	SSRs all loci	0.011 <sup>s</sup>	0.003 <sup>s</sup>
	Cytoplasmic markers	0.022 <sup>s</sup>	0.007 <sup>ns</sup>

<sup>ns</sup>Not significantly different to zero.<sup>s</sup>Significantly different to zero.Estimates of among- and within-village components of  $F_{st}$  ( $\theta$ ) for microsatellites (SSRs) markers and cytoplasmic markers.**Table 2** Population structure by loci

Locus	Chromosomal location (bin)	f	$\theta$
phi 011	1.09	0.261 ± 0.039	0.015 ± 0.009
phi 227562	1.11	0.059 ± 0.025	0.009 ± 0.004
phi 96100	2.00	0.075 ± 0.029	0.017 ± 0.008
phi 101049	2.08	0.092 ± 0.024	0.003 ± 0.003
phi 029	3.04	0.126 ± 0.046	0.004 ± 0.007
phi 093	4.08	0.143 ± 0.020	0.011 ± 0.008
phi 024	5.01	0.252 ± 0.035	0.004 ± 0.005
phi 452693	6.04	0.347 ± 0.029	0.011 ± 0.008
phi 034	7.02	0.079 ± 0.027	0.010 ± 0.006
phi 014	8.04	0.093 ± 0.037	0.019 ± 0.009
umc 1061	10.06	0.002 ± 0.021	0.020 ± 0.010
Overall markers		0.138 ± 0.031	0.011 ± 0.002

Estimates of  $F_{is}$  ( $f$ ) and  $F_{st}$  ( $\theta$ ) (not considering among-village differentiation) according to Weir and Cockerham (1984) by loci over all populations and over all markers. 95% confidence interval provided by jackknifing over populations for individual loci and by jackknifing over loci for the overall markers value.

cytoplasmic markers. A significant departure from Hardy-Weinberg equilibrium is observed in almost all populations. Estimates of  $F_{is}$  over all populations by loci show significant variation of homozygote excess among loci (Table 2). To test the uniformity of the homozygote excess among loci, we checked the distribution of  $F_{is}$ . One empty region (gap) was found. The three loci that map close to genes or QTLs involved in flowering time or anthesis-silking interval (phi011, phi024, and phi452693) showed significantly higher  $F_{is}$  values than the others (Table 2).

We evaluated the number of significant linkage disequilibria between pairs of loci. They are not higher than expected by chance alone at  $P \leq 5\%$  (5.1% positive over 31 populations and 55 possible different pairs of loci). Using a Bonferroni correction, no significant linkage disequilibrium is observed.

#### Within-population genetic variation in relation to temporality in flowering

A summary of genetic variation and variation for flowering traits for all populations is presented in Table 3. Two populations show clear evidence of admixture of material of different flowering precocity. Population from farmer 235 (Table 3) shows a bimodal distribution of the flowering time, and the population from farmer 115 (Table 3) has a very wide flowering range compared with the distribution of the flowering range among all populations. These two populations will not be considered for regression analysis.

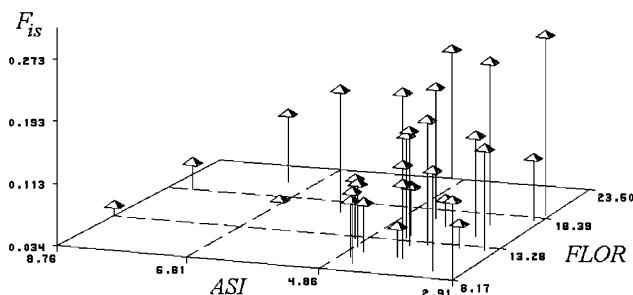
**Table 3** Genetic and flowering variation within populations

Village	Farmer's ID	$H_e$	$H_{obs}$	$F_{is}$	FLOR	ASI
Huitzo	107	0.72	0.54	0.25	16.62	3.52
Huitzo	108	0.65	0.52	0.21	17.03	4.40
Huitzo	110	0.69	0.55	0.20	12.92	3.94
Huitzo	115	0.69	0.58	0.17	32.25	3.98
Huitzo	121	0.68	0.55	0.20	16.72	5.80
Huitzo	122	0.69	0.65	0.07	15.88	4.08
Huitzo	129	0.71	0.66	0.07	12.92	3.47
Huitzo	136	0.66	0.59	0.12	17.83	3.03
Huitzo	139	0.71	0.57	0.21	23.50	5.05
Mazaltepec	211	0.70	0.61	0.13	20.75	7.23
Mazaltepec	222	0.69	0.64	0.07	18.33	8.36
Mazaltepec	235	0.68	0.65	0.05	26.48	5.34
San Lorenzo	309	0.69	0.67	0.03	18.17	6.96
San Lorenzo	314	0.66	0.55	0.17	14.72	3.47
San Lorenzo	317	0.67	0.64	0.05	13.33	8.76
Amatengo	405	0.70	0.63	0.10	10.99	4.61
Amatengo	433	0.70	0.61	0.14	12.94	4.31
Amatengo	439	0.71	0.62	0.13	8.17	2.91
Valdeflores	512	0.69	0.61	0.13	9.24	4.52
Valdeflores	533	0.69	0.64	0.07	10.81	4.08
Valdeflores	536	0.70	0.62	0.12	11.67	4.80
Santa Ana	602	0.71	0.63	0.11	10.18	4.68
Santa Ana	606	0.70	0.58	0.17	13.67	4.36
Santa Ana	610	0.72	0.64	0.11	12.92	5.02
Santa Ana	614	0.72	0.61	0.17	12.79	3.07
Santa Ana	619	0.70	0.57	0.18	12.60	4.17
Santa Ana	625	0.70	0.63	0.10	14.00	4.34
Santa Ana	630	0.72	0.57	0.21	15.22	4.63
Santa Ana	633	0.69	0.58	0.16	9.23	3.34
Santa Ana	639	0.71	0.52	0.27	18.92	2.99
Santa Ana	640	0.70	0.61	0.13	10.60	3.97

Expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_{obs}$ ), homozygote excess over all markers ( $F_{is}$ ), within-population heterogeneity in flowering time measured as flowering range (FLOR) and anthesis-silking interval (ASI) for 31 populations from six villages in the Central Valleys of Oaxaca. FLOR and ASI are given in days.

The within population range in flowering time and anthesis-silking interval differs greatly from one population to the other. The regression of the overall homozygote excess (using all 11 microsatellite markers) on mean anthesis-silking interval and flowering range is highly significant (Figure 2 and Table 4).

The population showing the most elevated  $F_{is}$  value (639 in Table 3) shows a significant correlation between the flowering distance and relatedness measured as Moran's  $I$ -statistic (temporal autocorrelogram shown in Figure 3). For this population, pairwise estimates of Moran's  $I$ -statistic ( $\rho_{ij}$ ) for the divergence in flowering (Div) class between 0.75 and 1.56 days significantly deviates from the expected value obtained by permuting locations as shown in Figure 3 ( $Pr = 0.029$ , star in Figure 3). In addition, the linear regression of  $\rho_{ij}$  on Div is significant ( $Pr = 0.0018$ ). Regression slope is  $m = -0.0244$  and the intercept,  $b = 0.1136$ . These results suggest a clear case of assortative mating (correlation coefficient,  $r = -0.243$ ) for this population. This population was tested for assortative mating because it is the only population showing a low anthesis-silking interval value, a high heterogeneity in flowering, and an elevated  $F_{is}$  value. To conduct the regression, we used only the three markers (described earlier) showing significantly higher  $F_{is}$  values than the others.

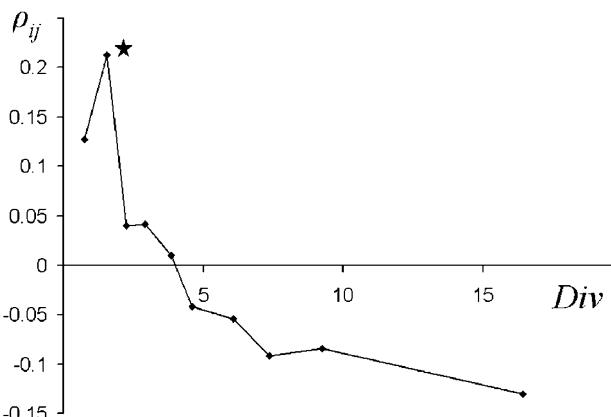


**Figure 2** Homozygote excess ( $F_{is}$ ) as a function of the flowering range (FLOR) and the anthesis-silking interval (ASI).  $F_{is}$  was calculated over all markers.

**Table 4** Regression of homozygote excess  $F_{is}$  on flowering range (FLOR) and the mean anthesis-silking interval (ASI)

Parameter	Estimate	Student's t-value for $H_0$ : Parameter = 0	Pr >  t	SE of estimate
INTERCEPT	0.142441	3.54	0.0016	0.04029086
FLOR	0.007635	2.88	0.0079	0.00265083
ASI	-0.02404	-3.71	0.0010	0.00647796

$R^2 = 0.39$ .  $F_{is}$  was calculated over all markers.



**Figure 3** Autocorrelogram showing the temporal segregation of the genotypes (Div is the divergence in flowering time) in function of the relatedness measured as Moran's  $I$ -statistic ( $\rho_{ij}$ ). Given for the population showing the most elevated overall  $F_{is}$  value and the lowest ASI value (population from farmer 639 in Table 3). Moran's  $I$ -values are computed using the individual genotypes as variables. Star symbol shows values that deviate significantly from zero ( $P < 0.05$ ). Moran's  $I$ -values are computed with SSR markers phi011, phi024, and phi452693. These loci showed significantly higher  $F_{is}$  values than the others.

## Discussion

### Low level of among-population differentiation

Village and distance do not appear to be determinants in population differentiation. The difference between cytoplasmic and nuclear  $F_{st}$  is not as large as expected when compared to other allogamous plants, and even less than that seen for some autogamous grasses (Ennos, 1994). Data from the survey on farmer management corroborate the genetic data analysis and indicate a large amount of

seed-flow between maize populations within the Central Valleys of Oaxaca. Considerable seed exchange by farmers seems to be common in Mexico (Louette *et al*, 1997). Furthermore, while the proportion of farmers reporting seed flow is high, it could be underevaluated as farmers are keen to test new populations (Bellon *et al*, 2003), which often results in seed mixing as a consequence of storage practices. The maize ears in husks will frequently be kept after harvest in a single pile, regardless of whether they are from the tested population or a farmer's pre-existing population. We observed that populations from other villages are often tested, and, although the farmer may not adopt them, seed mixing may occur. Seed flow among farmers allows long distance gene flow within and among villages. These high levels of gene flow ensure the maintenance of high levels of genetic diversity. Levels of diversity are high when compared to those measured with maize accessions representing the entire maize genetic diversity in the Americas (Matsuoka *et al*, 2002). Therefore, a maize landrace should not be considered as a separate entity, but rather as an open genetic system. Furthermore, our results underline the importance of farmers' choices in determining gene flow among maize landrace populations, as seed flow results from farmers' decisions.

To the best of our knowledge, this population study in maize is the first to be conducted based on a small geographic area. In contrast to our research, other maize population studies (Sanou *et al*, 1996; Gauthier *et al*, 2002) looked at larger areas (Europe and Burkina-Faso) and showed much higher population differentiation. We believe more studies are needed on maize populations in Mexico and more generally the Americas, both at the regional and inter-regional scales, to investigate the patterns of population structure. Although distance does not seem to contribute to village isolation within small geographic areas, the Matsuoka *et al* (2002) results clearly suggest isolation by distance at a continental scale.

### Temporal heterogeneity in allelic frequencies

While there is little among-population differentiation, a large amount of homozygote excess is observed within these populations. Usually, homozygote excess is attributed to consanguineous mating, population substructure, or to an artifact due to factors like null alleles. Enjalbert and David (2000) have inferred the outcrossing rate in wheat using molecular data at various loci. In a more recent study, Overall and Nichols (2001) have shown that it is possible to distinguish consanguinity from population substructure using multilocus genotype data. However, none of the above applies to the situation described in this paper. The biology of the species rejects the simple explanation of inbreeding, and the significant variation of homozygote excess among loci ( $F_{is}$  value differing considerably from one locus to another) does not correspond to what would be expected for population admixture. Nonrandom mating has been previously described in open-pollinated maize populations (Brown and Allard, 1970; Bijlsma *et al*, 1986). Kahler *et al* (1989) have shown that selfing did not contribute significantly to the inbreeding that occurred in the studied population. Here we present an unusual case of Wahlund effect and inbreeding, which is in some ways similar to that described in mole crickets (Nevo *et al*, 2000), and which

corresponds to assortative mating. We show that in the populations with the most elevated homozygote excess, there is a significant correlation between relatedness and the mean family divergence in flowering time. It appears that overall  $F_{is}$  value depends on both the flowering range and on the anthesis-silking interval of a given population (FLOR and ASI together explain around 39% of the variation for homozygote excess). The existence of a large anthesis-silking interval has the consequence of preventing assortative mating because of the long delay between male and female flowering. A low anthesis-silking interval and large flowering range of the population will result in a temporal heterogeneity in allelic composition for maternal plants and the pollen pool. Work by Kahler *et al* (1989) had previously suggested that homozygote excess in an open-pollinated maize population could be the consequence of positive assortative mating resulting from an overlap in the flowering period of plants carrying alleles that are identical by descent from a recent common ancestor.

#### Marker loci with elevated $F_{is}$ values map close to flowering genes

The three loci showing the most elevated  $F_{is}$  values are in regions known to be associated with flowering traits. Phi011 is within the interval between *dwarf8* and *indeterminate1*. While *dwarf8* has been shown to be associated with variation in flowering time (Thornsberry *et al*, 2001), *Indeterminate1* is a putative transcriptional regulator of floral transition that is thought to be a major player in controlling flowering time. Phi024 is in a region that is syntenic to the *dwarf8-indeterminate1* region and maps close to *dwarf9*, a possible duplication of *dwarf8* on chromosome 5 (Gale and Devos, 1998). Phi452693 maps close to a major anthesis-silking interval QTL (Veldboom *et al*, 1994; Ribaut *et al*, 1996). A large flowering range within a given population will lower the effective population size of a given class of flowering time and therefore lead to interclass differentiation.

#### Regarding phenotypic evolution

Understanding population structure and its pattern are crucial to understanding phenotypic evolution. It makes possible association studies with clear assumptions about population structure and its origin. The Central Valleys of Oaxaca offer a unique model for the study of the impact of farmer management and selection on the phenotypic diversification and evolution of maize because of the large amounts of phenotypic variation (Bellon *et al*, 2003), the unique position of these populations in maize evolution and diversification (Matsuoka *et al*, 2002), and the patterns of population structure in this region described in our results. The observed variation in flowering range and in anthesis-silking interval between populations suggest that the pattern of population structure for these traits could be very different from that described for molecular markers. In a future paper, we will describe the pattern of population structure for quantitative traits and the impact of farmer management and selection on population differentiation for these traits, in order to understand the basis of phenotypic evolution in maize.

#### Acknowledgements

We thank two anonymous reviewers for helpful comments in improving the manuscript and CIMMYT science editor David Poland for his editorial support.

#### References

Belkhir KP, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2001). *GENETIX, logiciel sous Windows TM pour la génétique des populations*, Release 4.02. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000. Université de Montpellier II, Montpellier, France.

Bellon MR, Berthaud J, Smale M, Aguirre JA, Taba S, Aragón F *et al* (2003). Participatory landrace selection for on-farm conservation: an example from the Central Valleys of Oaxaca, Mexico. *Genet Res Crop Evol* **50**: 401–416.

Bellon MR, Brush SB (1994). Keepers of maize in Chiapas, Mexico. *Economic Bot* **48**: 196–209.

Benz BF (2001). Archaeological evidence of teosinte domestication from Guila Naquitz, Oaxaca. *Proc Natl Acad Sci USA* **98**: 2104–2106.

Bijlsma R, Allard RW, Kahler AL (1986). Nonrandom mating in an open-pollinated maize population. *Genetics* **112**: 669–680.

Brown AHD, Allard RW (1970). Estimation of the mating system in open-pollinated maize populations using isozyme polymorphisms. *Genetics* **66**: 133–145.

Dewey SE, Heywood JS (1988). Spatial genetic structure in a population of *Psychotria nervosa* I. Distribution of genotypes. *Evolution* **42**: 834–838.

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.

Enjalbert J, David JL (2000). Inferring recent outcrossing rates using multilocus individual heterozygosity: application to evolving wheat populations. *Genetics* **156**: 1973–1982.

Ennos RA (1994). Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**: 250–259.

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

Garnier-Géry P (1992). *Contribution à l'étude de la variabilité inter- et intra-population chez le maïs (Zea mays L.): valorisation d'informations agromorphologiques et enzymatiques*, PhD. Institut National Agronomique Paris-Grignon.

Gauthier P, Gouesnard B, Dallard J, Redaelli R, Rebourg C, Charcosset A *et al* (2002). RFLP diversity and relationships among traditional European maize populations. *Theor Appl Genet* **105**: 91–99.

Hardy OJ, Vekemans X (1999). Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**: 145–154.

Hardy OJ, Vekemans X (2001). Patterns of allozyme variation in diploid and tetraploid *Centaurea jacea* at different spatial scales. *Evolution* **55**: 943–954.

Hardy OJ, Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* **2**: 618–620.

Kahler AL, Hallauer AR, Gardner CO (1986). Allozyme polymorphisms within and among open-pollinated and adapted exotic populations of maize. *Theor Appl Genet* **72**: 592–601.

Kahler AL, Shaw DV, Allard RW (1989). Nonrandom mating on tasseled and detasseled plants in an open pollinated population of maize. *Maydica* **34**: 15–21.

Lefort-Buson M, Lavergne V, Daudin JJ, Charcosset A, Sampoux JP, Gallais A (1991). Genetic variability among populations of maize germplasm. 2. Enzymatic polymorphism and its

relationship to quantitative trait diversity. *Maydica* **36**: 237–246.

Lewis PO, Zaykin D (2002). *GDA for Windows*, Release 1.1. Department of Ecology and Evolutionary Biology, The University of Connecticut, Storrs, CT.

Louette D, Charrier A, Berthaud J (1997). *In situ* conservation of maize in Mexico: genetic diversity and maize seed management in a traditional community. *Economic Bot* **51**: 20–38.

Louette D, Smale M (2000). Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. *Euphytica* **113**: 25–41.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci USA* **99**: 6080–6084.

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

Nevo E, Beiles A, Korol AB, Robin YI, Pavlicek T, Hamilton W (2000). Extraordinary multilocus genetic organization in mole crickets, Gryllotalpidae. *Evolution* **54**: 586–605.

Overall AD, Nichols RA (2001). A method for distinguishing consanguinity and population substructure using multilocus genotype data. *Mol Biol Evol* **18**: 2048–2056.

Piperno DR, Flannery KV (2001). The earliest archaeological maize (*Zea mays* L.) from highland Mexico: new accelerator mass spectrometry dates and their implications. *Proc Natl Acad Sci USA* **98**: 2101–2103.

Provan J, Lawrence P, Young G, Wright F, Bird R, Paglia G et al (1999). Analysis of the genus *Zea* (Poaceae) using polymorphic chloroplast simple sequence repeats. *Plant Syst Evol* **218**: 245–256.

Ribaut J-M, Hoisington DA, Deutsch JA, Jiang C, Gonzalez-de Leon D (1996). Identification of quantitative trait loci under drought conditions in tropical maize. 1. Flowering para-  
meters and the anthesis-silking interval. *Theor Appl Genet* **92**: 905–914.

Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**: 1219–1228.

Salanoubat M, Pernes J (1986). Enzyme polymorphisms within and between European maize populations. *Maydica* **31**: 269–278.

Sanou J, Gouesnard B, Charrier A (1996). Isozyme variability in West African maize cultivars (*Zea mays* L.). *Maydica* **42**: 1–11.

Smale M, Aguirre JA, Bellon MR, Mendoza J, Manuel Rosas I (1999). Farmer management of maize diversity in the central valleys of Oaxaca, Mexico: CIMMYT/INIFAP 1998 baseline socioeconomic survey. CIMMYT Economics working paper 99–09.. International Maize and Wheat Improvement Center (CIMMYT), Mexico, DF.

Sokal RR, Rohlf FJ (1995). *Biometry*, WH Freeman and Company, New York.

Thornberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler IV ES (2001). Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet* **28**: 286–289.

Veldboom LR, Lee M, Woodman WL (1994). Molecular marker-facilitated studies in an elite maize population: I Linkage analysis and determination of QTL of morphological traits. *Theor Appl Genet* **88**: 7–16.

Weir BS (1979). Inferences about linkage disequilibrium. *Biometrics* **25**: 235–254.

Weir BS (1996). *Genetic Data Analysis II*, Sinauer Associates, Inc., Sunderland.

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

Zeven AC (1998). Landraces: a review of definitions and classifications. *Euphytica* **104**: 127–139.