A test of ecological and ethnolinguistic determinants of maize diversity in southern Mexico.

Quetzalcóatl Orozco-Ramírez (Corresponding Author). Centro de Investigaciones en Geografía Ambiental, UNAM. Antigua Carretera a Pátzcuaro, Col. San José de la Huerta 8701, Morelia, Michoacán, México CP 58190. Tel. +52 1 438 112 5488. qorozco@gmail.com

Amalio Santacruz-Varela. Colegio de Postgraduados. Montecillos, Texcoco, Estado de México, México.

Jeffrey Ross-Ibarra. Dept. of Plant Sciences, Center for Population Biology, and Genome Center, University of California. Davis, California, USA

Stephen Brush. Department of Human Ecology, University of California. Davis, California, USA

Running title: Maize population structure and ethnolinguistic variation

Total words: 3959

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1192v1 | CC-BY 4.0 Open Access | rec: 23 Jun 2015, publ: 23 Jun 2015

Abstract

While prevailing theories of crop evolution suggest that crop diversity and cultural diversity should be linked, empirical evidence for such a link remains inconclusive. In particular, few studies have investigated such patterns on a local scale. Here, we address this issue by examining the determinants of maize diversity in a local region of high cultural and biological richness in Southern Mexico. We collected maize samples from Mixtec and Chatino villages at low and mid elevations. Although morphological traits show few patterns of population structure, we see clear genetic differentiation among villages, with ethnicity explaining a larger proportion of the differentiation than altitude. Consistent with an important role of ethnicity in patterning seed exchange, metapopulation model-based estimates of differentiation match the genetic data within village and ethnic group, but dramatically underestimate differentiation when all four villages are taken together. Our research provides insights about the importance of social origin in structuring maize diversity at the local scale.

Key Words: Maize Diversity, Genetic Diversity, Population Structure, Mexico-Southern, Indigenous People, Crop Diversity

INTRODUCTION

The last decade has seen progress in research about crop diversity linked to cultural diversity and social factors, but there is still much to understand about this complex relationship. Crop evolution and diversity depend on selection mediated both by the environment and by farmers (Harlan, 1975). Previous research has shown that social factors such as ritual use, identification as indigenous or mestizo, or even simple aesthetics, can contribute to the maintenance of particular landraces (Bellon, 1996; Zimmerer, 1996; Brush and Perales, 2007). We might thus expect that culturally determined preferences and perceptions have molded crop populations, but demonstrating the role of cultural diversity in generating crop diversity has been more difficult. Differentiation of maize populations and other crop populations by cultural variation was suggested long ago (Anderson, 1946; Hernández, 1972), but only recently has empirical data been reported (Perales *et al.*, 2005, Labeyrie *et al.*, 2014).

Farmers rigorously select seed for use in each coming year, and this selection is sufficient to maintain distinctive traits in the face of abundant pollen flow and extensive seed movement (Louette and Smale, 2000; Ortega-Paczka, 2003). Cultural differences between groups may be expressed as preferences for colors, textures and uses for particular varieties (Ortega-Paczka, 2003), but also serve to erect barriers to the movement of seed (Hernández, 1972). Consistent with this, many workers have documented correlations between ethnolinguistic and biological diversity (Maffi, 2005), but the relationship between these factors is complex, in part because ethnolinguistic groups often inhabit different environments and ecological niches (Brush, 2004). In contrast to the popular assumption that there is a direct relationship between ethnolinguistic diversity and maize diversity, there is little research that has formally and systematically addressed that interaction.

Genetic research describes continuous variation among domesticated maize, although regional clusters or complexes are apparent (Matsuoka *et al.*, 2002; Vigouroux *et al.*, 2008). Clustering is most evident in the use of isozymes to measure genetic distance and construct phylogenetic trees that divide in eco-geographic regions by latitude, longitude and altitude (Sánchez *et al.*, 2000). Boege (2008) described agro-biodiversity in indigenous territories, but his ability to draw conclusions about specific races for specific ethnic groups was very limited. Using morphological characteristics that are under farmer selection, social scientists and plant biologists have shown that farmers maintain morphologically distinct maize populations at much smaller regional scales (Pressoir and Berthaud, 2004a; Perales *et al.*, 2005). Particularly, Perales *et al.* (2005) found that ethno-linguistic diversity in the same environment was linked to maize morphological diversity, but not to genetic differentiation based on isozymes. In contrast to the findings of Pressoir and Berthaud (2004b) and Perales *et al.* (2005), van Etten *et al.* (2008) found that maize populations from different villages within a small, culturally homogeneous region in Guatemala are both genetically and phenotypically separated. They confirmed, however, the central finding of the Mexican case studies that social origin plays a significant role in determining the patterns of maize in the region. Interestingly, maize diversity in NW Guatemala was more discernible between communities than between regions, a finding that van Etten *et al.* (2008) attribute to patterns of seed exchange at local and regional levels and to diffusion of innovations at the regional level.

This paper examines maize diversity and population structure at the local scale and their relation to both ethno-linguistic variation and environmental variation (elevation) in southern Mexico. We studied farmers who speak either Chatino or Mixtec, two languages of the Oto-Manguean family that have been separated for approximately 4700 years (Kaufman, 1990). We collected maize populations from two environments — low- and mid-elevation — in two neighboring, indigenous municipalities separated by language affiliation. We hypothesized that maize collections from the same municipality would be more similar than those from different municipalities even though comparable environmental variation occurs within each. By using paired villages, we are able to separate the effects of environmental and ethnolinguistic variation. We find the effect of ethnicity is stronger in structuring maize populations in terms of morphology and genetics. Application of a metapopulation model suggests that genetic differentiation is due to the lack seed flow between ethnolinguistic areas.

MATERIALS AND METHODS

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1192v1 | CC-BY 4.0 Open Access | rec: 23 Jun 2015, publ: 23 Jun 2015

Study site

Fieldwork was carried out in the Sierra Sur of Oaxaca (Fig. 1). This mountainous range extends along the Pacific Coast from southern Jalisco to the Isthmus of Tehuantepec in the state of Oaxaca. We worked in the Mixtec municipality of Santiago Amoltepec and the Chatino municipality of Santa Cruz Zenzontepec. These indigenous communities have an ancient and shared history: their townships are only 12 kilometers apart and they have been affected by the same regional and historical dynamics. The topography is abrupt, with mountains, canyons, and hills leading to elevation variation from 105 to 2150 masl. The climate is hot in the lowlands and temperate in the higher elevations. The mean annual temperature is 26 °C in the lowlands and 18°C in the upper elevations. The rainy season starts in May and ends in October, with an average from 1500 to 2000 millimeters precipitation per year (INEGI, 2013). Humidity differs strongly between the lowlands and highlands and is affected by the exposure in the hills. Soil diversity is high because of the complex geology; according to the most detailed available information, Litosol and Regosol eutrico are the most important soil types (INEGI, 2013). In general, soils present some level of erosion due to agricultural practices, runoff, and wind. Municipalities are integrated internally through local governments and markets. Beyond some seed in backyard gardens, no hybrid or improved seed has been planted in the area. Crop management is similar in both municipalities, the only difference being the use of fertilizer in the Mixtec municipality, Amoltepec, and not in the Chatino municipality, Zenzontepec.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1192v1 | CC-BY 4.0 Open Access | rec: 23 Jun 2015, publ: 23 Jun 2015

Maize collections and reciprocal common gardens

To test whether maize diversity and population structure are shaped by ethnicity and/or environmental factors, maize collections and common garden experiments with morphological characterization were performed. Four villages were selected: one Chatino and one Mixtec at middle elevation (1000-1300 masl) identified as Ch-M and M-M respectively; and one Chatino and one Mixtec in the lowlands (400-600 masl), identified as Ch-L and M-L. A detailed description of the sampling and common gardens are described elsewhere (Orozco-Ramirez *et al.,* 2014). We collected 135 maize samples from the four villages. Each maize sample consisted of 12 seed quality ears of each seed lot (farmer-identified variety) that the household planted in the previous year. Ecological information and management of each seed lot was recorded by a survey. Five maize samples from each village were selected to plant in the common gardens; these samples resembled the total variation of maize in that particular village. In this analysis, we used only the data from two common gardens under fertilization treatment, one in the Chatino village low elevation (Ch-L) and another in Mixtec middle elevation (M-M). These were the fields with the best soil conditions to perform morphological characterization. Morphological data recorded from common gardens included: days to anthesis, days to silking, tassel branches, tassel length, stem diameter, leaf length, leaves per plant, leaf width, ear height on the plant, plant height, ear diameter, ear length, grain rows, kernels per row, cob diameter. Twenty plants were measured from the two rows in the center. Flowering time was recorded when 50% of the plants had reached anthesis or were silking. For ear variables, all the plants of one row located in the center of the plot were harvested. We averaged each variable over plots, resulting in 60 experimental units in each common garden.

Molecular analysis

Molecular analysis was carried out at the Colegio de Postgraduados, Mexico. We utilized the same 20 maize seed lots (but not the same physical individuals) as used in common gardens for microsatellite genotyping. DNA was extracted from 10 individuals randomly selected for each population, using the standard protocol prescribed by the ChargeSwitch gDNA Plant Kit (Invotrogen ™). We used 100-150 mg from seedling tissue. DNA extraction was made by a King Fisher Flex (Thermo Scientific) automatic extractor. The DNA samples selected had a DNA concentration above of 50 ng/μl and an absorbance ration from 1.40 to 1.80 at 260/280 nm wavelength. The DNA was evaluated by a Nanodrop 2000 spectrophotometer (Thermo Scientific). Extracted samples were genotyped for 15 microsatellite loci, listed in Appendix Table A.1. Fluorescently labeled primers (ROX, 6-FAM, HEX) were obtained for these loci (Invotrogen™). Multiple Polymerase Chain Reactions (PCR) were performed in a 25 μl reaction volume, containing 4 pmol/μl of R and F primer (Invitrogen), 0.16 nM of dNTP mix (Promega), 1.2 nM of $MgCl_2$ (Promega), 0.8 of 5X GoTaq flexi buffer (Promega), 1 U of GoTaq flexi DNA polymerase (Promega) and 25 nM of DNA. The amplification program was: 95°C for 4 minutes; followed by 25 cycles of 95°C for 1 minute, 55°C for 2 minutes, and 72°C for 2 minutes;

followed by extension at 72°C for 60 minutes. PCR was performed in a GeneAmp PCR System 9700 (Applied Biosystems). PCR products were analyzed by capillary electrophoresis in a 3130 Genetic Analyzer (Applied Biosystems). Fragment sizes were scored using GeneMapper v4.0 (Applied Biosystems).

Statistical analysis

Maize population structure based on morphological traits from the common gardens was analyzed by principal components analysis (PCA). We excluded days to silking and height of ear in the plant because these variables were highly correlated to days to tasseling and plant height, respectively. Data were standardized by subtracting the mean from each observation and dividing by the standard deviation. The unit of analysis was seed lot sample at each plot in the common gardens. In order to define the effect of elevation and ethnolingusitic group in structuring populations using morphological traits, we did a permutational multivariate analysis of variance using distance matrices (Anderson 2001). This is a non-parametric method that partitions a distance matrix among sources of variation. For this analysis we used the same variables as for the PCA. The distance matrix was calculated using the *Euclidian* method on the standardized data and we allowed for 9999 permutations to calculate F statistics and to have an ample margin to reject the null hypothesis at an α-level of 0.05. For this analysis we applied the function *adonis* in the package *vegan* (Oksanen *et al.*, 2015) for R (R Core Team, 2014).

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1192v1 | CC-BY 4.0 Open Access | rec: 23 Jun 2015, publ: 23 Jun 2015

We used model-based clustering to evaluate population structure as implemented in the software STRUCTURE 2.3.4, (Pritchard *et al.*, 2000), using the admixture model with correlated allele frequencies and allowing the model to use location information for the samples to assist the clustering. The estimated proportion of each cluster forming an individual genome (q) was calculated for K ranging from 1 to 10 populations, with ten runs for each K value. We used a burn-in period of 100,000 and 100,000 iterations for estimating the parameters. The criterion suggested by Evanno *et al*. (2005), based on the second order rate of change in the log probability of data between successive K values was used to determine the most likely number of clusters (K).

Due to low genetic differentiation among maize samples within village in the STRUCTURE, we treat each village as a single population to then perform a locus by locus AMOVA (Excoffier *et al.*, 1992) grouping villages by ethnicity and then grouping by ethnic groups. We preferred locus by locus AMOVA because there were some missing data and we included individual level in the calculations. Significance was calculated using 16000 permutations and estimates of the proportion of variation at different levels were calculated as a weighted average across loci. Because each village of the same ethnicity is located at different elevation, when testing within ethnicity in each group we are testing for the effect of environment in structuring the population. We also calculated a matrix of genetic distance (F_{ST})

(Wright, 1951) among villages based on the number of different alleles using Arlequin 3.5 (Excoffier *et al.*, 2005), with significance (α=0.05) calculated after 1000 permutations.

Finally, we used the model proposed by van Heerwaarden *et al.* (2010) to estimate genetic differentiation due to seed management and seed flow. This model approaches maize fields in a village as a metapopulation and uses parameters from maize farmers' practices in traditional agricultural systems to estimate F_{ST} following Slatkin (1991). We compared model-based estimates of F_{ST} to F_{ST} calculated from our SSR genotyping. F_{ST} was calculated for each village and globally using Arlequin 3.5 (Excoffier *et al.*, 2005). The model uses these parameters: number of demes (n), number of ears planted per deme (N_f), total number of plants per deme (N), number of migrating ears (N_{fm}), replacement probability (e), migration (mixture proportion) proportion ($m=N_{fm}/N_f$), proportion of seed mixture (p_m), and proportion pollen migration (m_g). Values used for each parameter are in Table 1; unless otherwise stated data used came from our field surveys (Orozco-Ramirez *et al*. 2014). Number of demes (n) were calculated by multiplying number of households by mean number of seed lots for each village. Number of ears planted per deme (N_f) was calculated from average planted area, seed sown per hectare, average kernel weight in the region (Aragon-Cuevas *et al.,* 2012) and the average number of kernels used as seed from each ear. The total number of plants per deme (N) was estimated by multiplying the average kernels per ear by N_f. The number of migrating ears (N_{fm}) was calculated from seed exchange averaged over farmers in each village (kg). First we obtained the proportion of seed exchanged respect to total seed planted and then multiply that proportion by N_f. Replacement probability (e) was calculated as the proportion of new seed lots with respect to the total reported for the previous season. Initial values of the proportion of pollen migration (m_g) were taken from the literature (Messeguer *et al.*, 2006), but these were later fitted to the observed F_{ST} data.

RESULTS

Maize diversity in the area of study

On average, the number of landraces cultivated by a single farmer ranges from 1.33 in M-L to 1.89 in M-M. In M-M, the majority of farmers (67%) have two landraces. In the other three villages, most farmers had only one landrace. In M-L, fewer farmers had two landraces than in the other three villages. Only in Ch-M did farmers have four landraces, but the percentage of farmers with more than three landraces was very low, and most in that village have only one landrace. In total we found seven racial groups in the four villages (*Tuxpeño, Olotillo, Conejo, Tepecintle, Pepitilla, Elotes occidentales,* and *Zapalote grande*). A test for independence of race frequency by village finds little support for a dependence on elevation (p-value = 0.06793, 1000 Monte Carlo simulations), but strong support for dependence on ethnolinguistic group (p value = 0.0001, 1000 Monte Carlo simulations). Mixtec communities have more of the *Conejo* and *Tepecintle* races, and Chatino villages have more of the *Olotillo* and *Tuxpeño* races (Fig. 1).

Maize population structure based on morphological traits

Principle component analysis reveals a continuum of maize morphological diversity across the region. The plot of the first two principal components does not show clearly separate groups in either common garden (Fig 2). However, in both common gardens it is possible to see greater clustering of samples when labeled by ethnicity (Fig. 2A and 2C) than by elevation (Fig. 2B and 2D).

According to permutational multivariate analyses of variance the main effect of ethnicity was significant in structuring morphological variation, but neither elevation nor the interaction between elevation and ethnicity were significant (Table 3). Comparison of the mean sum of squares (Anderson, 2001) suggests that ethnicity has a stronger effect than elevation in structuring maize populations (Table 3). Nonetheless, there are some differences between common gardens. The effect of ethnicity was weaker in the low elevation garden, perhaps due to the lower overall morphological variation observed.

Maize population structure based on molecular markers

Results from STRUCTURE software suggested the existence of two clusters defined by ethnicity (Fig. 3). The highest value of ΔK was found at K=2, but at higher values of K the Mixtec area shows separation between villages. Chatino villages are not similarly separated. Most individuals show evidence of admixture, and there is no correspondence between maize race

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1192v1 | CC-BY 4.0 Open Access | rec: 23 Jun 2015, publ: 23 Jun 2015

name and STRUCTURE. The only exception is a sample from M-M of the *Conejo* race, which at higher values of K (≥4) forms a stable cluster with low admixture. Samples never cluster by elevation.

Our AMOVA finds relatively strong population structure (Table 4). Most of the genetic variation (73%) was found within populations, with less variation assigned to ethnic group (4%) and elevation (1.75%). The results of the AMOVA are confirmed by the matrix of pairwise genetic distance (F_{ST}) among villages, although all values are low. Fst between villages of the same ethnic group (0.019 Mixtec, 0.021 Chatino) are lower than between populations of the same elevation but different ethnic groups (0.041 middle, 0.066 low). F_{ST} between villages from different ethnic groups and different elevations are also large (M-M vs Ch-L is 0.059, M-L vs Ch-M 0.045), but the largest difference is between Mixtec Lowlands and Chatino Lowlands (0.066). All F_{ST} values are significant (α=0.05) calculated after 1000 permutations.

Metapopulation model

We used the metapopulation model of van Heerwaarden *et al.* (2010) to estimate F_{ST} and theoretical pollen migration based on seed management and exchange from our surveys. Within villages, in order to obtain similar modeled values to F_{ST} to those calculated by SSR´s we had to assume similar pollen migration rates to those in the literature (Messeguer *et al*., 2006), with the exception of the M-M village where a very low pollen migration proportion (0.0089) was needed, the range in the other three villages was from 0.010 to 0.018 (Table 2). We next asked whether the model can be extended to villages within an ethnolinguistic group, and were able to find a reasonable fit of the model using pollen flow values within the range used for within-village comparisons. We found the model fitted very well using exactly the same pollen migration value for both ethnolinguistic areas, moreover this is in the within village range. Supporting the idea that gene flow between villages of the same ethnolinguistic group is similar than within villages. In contrast, fitting the model to elevation groupings or the entire study region, required much lower pollen flow estimates: 0.0087 for lowlands, 0.0085 for middle elevation lands and 0.0083 for the region (Table 2).

DISCUSSION

Results from our comparison of genetic and morphological variation among maize varieties cultivated by Mixtec and Chatino farmers support the hypothesis that cultural (ethnolinguistic) differences can shape diversity as much or more than the environment. Plant morphological characteristics measured in the common gardens display considerable variation without strong clustering. Nevertheless, it is possible to see the effect of ethnicity but not elevation in structuring the population (Table 3). All methods used to analyze molecular markers show greater support for structure due to ethnicity than environment (elevation).

Our results contrast with previous findings in the Oaxaca Valley (Pressoir and Berthaud 2003a) and Chiapas (Perales *et al.*, 2005) that suggest social origin only impacts morphological variation directly selected by farmers. Pressoir and Berthaud (2003a) argued that cultivation in different villages and farmer´s selection contribute to morphological differentiation, but that pollen migration among populations reduces genetic separation. Similarly, Perales *et al.* (2005) found morphological but no genetic differentiation between neighboring ethnolinguistic groups. Their surveys found that a large majority (>70%) of farmers were interested in receiving seed from villages of a different ethnic group, suggesting that seed movement may explain the extremely low F_{ST} values they observed. Comparing highland and lowland maize samples from four states in east-central Mexico, Van Heerwaarden (2007) reports genetic differentiation according to altitude but not according to social origin within altitudes. Most research, therefore, suggests gene flow is important among maize populations from different villages and that farmers' selection is important to maintaining morphological differentiation.   In contrast, we find modest morphological and genetic differentiation between ethnolinguistic areas (F_{CT} = 0.040) (Table 4) that are geographically quite close, showing greater global differentiation (F_{ST} = 0.111) (Table 2) than values reported by Pressoir and Berthaud (2004b) (0.003) or by van Heerwaarden (2007) (0.027). Our study was carried out in a region with no roads crossing it and with no common local or regional markets. This contrasts to the situation in the Central Valleys of Oaxaca, where roads and markets unify the region and exchange commonly occurs between villages. In contrast, our results suggest that ethnolinguistic differences effectively isolate maize populations in this region that are otherwise under similar natural and artificial selection pressures. We conclude that, at least in some cases, ethnolinguistic affiliation can reduce gene flow more than the environmental obstacles posed by altitude differences. We posit that ethnically based seed networks foster both morphological and genetic separation, an idea similar to that of Hernández (1972) who suggested that indigenous groups isolate maize populations in a way similar to geographic barriers.

Previous work has found that both genetic and morphological variation are strongly structured by elevation (Doebley *et al.*, 1985; Benz, 1986; Bretting and Goodman, 1989; Vigouroux *et al*., 2008; van Heerwaarden *et al*., 2011). Research on the distribution of maize races in central Mexico (Perales *et al.,* 2003) and Chiapas (Brush and Perales, 2007), found that maize races are distributed according to elevation, and common garden experiments suggest local adaptation to elevation (Mercer et al. 2008). Van Heerwaarden (2007) showed close association between maize genetic structure and elevation at a regional scale in east-central Mexico, and genetic analyses find a significant impact of elevation on genome-wide diversity in both maize and its wild relative teosinte (Pyhäjärvi *et al*., 2013; Bradburd *et al.,* 2014; Takuno *et al*., 2015). Contrary to these findings, we found no differentiation of maize populations by elevation (races, morphological traits and molecular markers), likely due to the much smaller geographic scale of our population sampling.

Models of metapopulation structure based on our survey data support a role for ethnicity in patterning genetic diversity in our study area. The metapopulation model of van Heerwaarden *et al.* (2010) model is able to fit F_{ST} values within most villages areas using pollen migration values similar to direct estimates reported in the literature (Messeguer *et al*., 2006) (Table 2). The strikingly high F_{ST} and correspondingly low pollen migration required to fit the model in M-M was due to the presence of a highly distinct early maturing landrace in part of the village (Figure 3); because of differences in flowering time we hypothesize there is likely very little pollen migration between this landrace and others. The model fits observed F_{ST} between ethnolinguistic groups well using pollen migration parameters within the range seen in individual villages (Table 2). The model was only able to fit observed F_{ST} values within elevation regimes (between ethnolinguistic groups) or among all villages of both groups, by a substantial reduction in the pollen migration parameter or setting seed migration frequency to 0 and a lesser reduction in pollen migration. Because of the good fit of the model within villages and between villages within an ethnic group, the decrease in migration (pollen or seed) required to fit the model to elevation groups or the entire data is consistent with the idea that ethnolinguistic group is a limitation to maize gene flow in this region.

Previous studies that have found morphological differences among maize from different populations have not found much differentiation at the genetic level, suggesting that selection for a particular maize ideotype cannot explain the genetic differentiation observed in our villages.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1192v1 | CC-BY 4.0 Open Access | rec: 23 Jun 2015, publ: 23 Jun 2015

Instead, we suggest that a reduction of gene flow by limited seed and pollen migration among villages of different ethnolinguistic groups has effected genetic structure both in morphological traits and in genome-wide markers. We propose that detailed investigation of seed networks is an important next step to understanding the processes that pattern genetic diversity in maize.

ACKNOWLEDGEMENTS

We thank CONACYT and UC MEXUS for funding this research through a doctoral scholarship and a dissertation grant. We thank Msc. Flavio Aragon-Cuevas (INAFAP, Mexico) for maize racial classification; Cinthia Guzman, Laura Carrillo, and Juan Sánchez (Colegio de Postgraduados, Mexico) for genotyping work; Dr. Mark Grote (UC Davis) and Jonathan Fresnedo (UC Davis) for statistical advice; Joost van Heerwaarden for sharing an R script to run the meta-population model; and farmers and authorities from Santiago Amoltepec and Santa Cruz Zenzontepec for their support and for allowing to carry out this research. We also thank CIGA-UNAM for a postdoctoral scholarship to improve the writing of this paper.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

Sequence data are available at http://www.datadryad.org/

REFERENCES

Anderson E (1946). Maize in Mexico: A preliminary survey. *Ann Missouri Bot Gard* 33:147-247.

Anderson MJ (2001). A new method for non−parametric multivariate analysis of variance. *Austral Ecol* 26:32-46.

Aragón-Cuevas F, Figueroa-Cárdenas J, Flores-Zarate M, Gaytán-Martínez M, Véles Medina J (2012). *Calidad industrial de maíces nativos de la sierra sur de Oaxaca*. INIFAP: Oaxaca, Mexico.

Bellon MR (1996). The dynamics of crop infraspecific diversity: A conceptual framework at the farmer level. *Economic Bot* 50: 26-39.

Benz BF (1986). *Taxonomy and Evolution of Mexican Maize*. PhD. University of Wisconsin, Madison.

Boege E (2008). *El patrimonio biocultural de los pueblos indígenas de México: hacia la conservación in situ de la biodiversidad y agrodiversidad en los territorios indígenas*. Instituto Nacional de Antropologia e Historia; Comisión Nacional para el Desarrollo de los Pueblos Indígenas: México, D.F.

Bradburd GS, Ralph PL, Coop GM (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution* 67: 3258-3273.

Bretting P, Goodman M (1989). Karyotypic variation in Mesoamerican races of maize and its systematic significance. *Economic Bot* 43:107-124.

Brush S (2004). Farmers' bounty: locating crop diversity in the contemporary world. New Haven Yale Univ. Press.

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

Doebley JF, Goodman M, Stuber CW (1985). Isozyme variation in the races of maize from Mexico. *Amer J Bot* 72: 629-639.

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

Excoffier L, Laval G, Schneider S (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinformatics Online* 1:47.

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

Harlan J (1975). *Crops and Man*. American Society of Agronomy: Madison, WI.

Hernández XE (1972). Exploración etnobotánica en maíz. *Fitotecnia latinoamericana* 8:46-51.

INEGI (2013). *Conjunto de datos vectoriales de la serie topográfica y de recursos naturales escala 1:1 000 000*. INEGI: Mexico.

Kaufman T (1990). Early Otomanguean Homelands and Cultures: Some Premature Hypotheses. *University of Pittsburgh Working Papers in Linguistics* 1:91-136.

Labeyrie, V, et al. (2014). Influence of Ethnolinguistic Diversity on the Sorghum Genetic Patterns in Subsistence Farming Systems in Eastern Kenya. *PLoS One* 9(3):e92178.

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

Maffi L (2005). Linguistic, cultural, and biological diversity. *Annu Rev Anthropol* 34:599-617.

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1192v1 | CC-BY 4.0 Open Access | rec: 23 Jun 2015, publ: 23 Jun 2015

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

Mercer K, Martínez−Vásquez A, Perales HR (2008). Asymmetrical local adaptation of maize landraces along an altitudinal gradient. *Evol Applic* 1:489-500.

Messeguer J, Peñas G, Ballester J, Bas M, Serra J, Salvia J, Palaudelmàs M, Melé E (2006). Pollen−mediated gene flow in maize in real situations of coexistence. *Plant Biotech J* 4:633-645.

Oksanen J et al. (2015). *vegan: Community Ecology Package*. R package version 2.1-1. http://CRAN.R-project.org/package=vegan.

Orozco–Ramírez Q, Brush SB, Grote MN, Perales H (2014). A Minor Role for Environmental Adaptation in Local–Scale Maize Landrace Distribution: Results from a Common Garden Experiment in Oaxaca, Mexico. *Economic Bot* 68:383-396.

Ortega-Paczka R (2003). La diversidad del maíz en México. In *Sin maíz no hay pais,* eds. G. Esteva & C. Marielle, 123-154. Consejo Nacional para la Cultura y las Artes, Dirección General de Culturas Populares: Mexico City

Perales H, Benz B, Brush S (2005). Maize diversity and ethnolinguistic diversity in Chiapas, Mexico. *Proc Natl Acad Sci USA* 102:949-954.

Perales H, Brush S, Qualset CO (2003). Landraces of maize in Central Mexico: an altitudinal transect. *Economic Bot* 57:7-20.

Pressoir G, Berthaud J (2004a). Population structure and strong divergent selection shape phenotypic diversification in maize landraces. *Heredity* 92:95-101.

Pressoir G, Berthaud J (2004b). Patterns of population structure in maize landraces from the Central Valleys of Oaxaca in Mexico. *Heredity* 92:88-94.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.

Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J (2013). Complex patterns of local adaptation in teosinte. *Genome Biol Evol* 5:1594-1609.

R Core Team (2014). *R: A language and environment for statistical computing.* ISBN 3-900051-07-0. R Foundation for Statistical Computing: Vienna, Austria. url: http://www. R-project.org.

Sanchez J, Goodman M, Stuber C (2000). Isozymatic and morphological diversity in the races of maize of Mexico. *Economic Bot* 54:43-59.

Slatkin M (1991). Inbreeding coefficients and coalescence times. *Genetical Research*, 58:167-175.

Takuno S, Ralph P, Swarts K, Elshire RJ, Glaubitz JC, Buckler ES, Hufford MB, Ross-Ibarra J (2015). Independent molecular basis of convergent highland adaptation in maize. *Genetics* (In Press).

van Etten J, Fuentes López M, Molina Monterroso L, Ponciano Samayoa K (2008). Genetic diversity of maize (Zea mays L. ssp. mays) in communities of the western highlands of Guatemala: geographical patterns and processes. *Genet Resour Crop Ev* 55:303-317.

van Heerwaarden J (2007). *Population genetics of traditionally managed maize: farming practice as a determinant of genetic structure and identity of maize landraces in Mexico.* PhD. Wageningen University.

van Heerwaarden J, van Eeuwijk F, Ross-Ibarra J (2010). Genetic diversity in a crop metapopulation. *Heredity* 104:28-39.

van Heerwaarden J., et al. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc Natl Acad Sci USA* 108: 1088-1092.

Vigouroux Y, Glaubitz J, Matsuoka Y, Goodman M, Sanchez J (2008). Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Amer J Bot* 95:1240-1253

Wright, S (1951). The genetical structure of populations. *Ann. Eugen*. 15: 323-354.

Zimmerer K (1996). *Changing fortunes: biodiversity and peasant livelihood in the Peruvian Andes*. University of California Press: Berkeley

Table 1. Seed management parameters for F_{ST} estimation using van Heerwaarden *et al.* (2010) metapopulation model to estimate genetic structure.

Parameter	M-M	M-L	Ch-M	Ch-L	Total/Mean
Data from surveys					
Number of households	152	49	206	219	626
Mean of seed lots	1.89	1.33	1.57	1.4	1.55
Average planting area (ha)	0.52	0.61	0.56	0.78	0.62
Seed sow per hectare (kg)	16	16	16	16	16
Total seed used per deme (kg)	8.32	9.76	8.96	12.48	9.92
Ear weight (kg)[1]	0.087	0.087	0.087	0.087	0.087
Mean kernels per ear[1]	262	262	262	262	262
Mean of quantity seed exchange (kg)	3.3	4	5	4.9	4.2
Proportion of exchange seed	0.40	0.41	0.56	0.39	0.42
Parameters for the model					
Number of demes (n)	287	65	323	307	970
Number of ears planted per deme (N_f)	96	112	103	143	114
Total number of plants per deme (N)	25056	29392	26983	37583	29874
Number of migrating ears (N_{fm})	38	46	57	56	48
Seed lot replacement probability (e)	0.097	0.15	0.186	0.061	0.13
migration proportion (m=N_{fm}/N_f)	0.40	0.41	0.56	0.39	0.42
Proportion of seed mixture (p_m)	0.00125	0.00106	0	0	0.00058

[1] Estimated according to mean row and grain number per ear used for seed and using the average kernel weight estimated by Aragon-Cuevas *et al*. (2012) for landraces from the region.

Table 2. Results of van Heerwaarden *et al.* (2010) metapopulation model to estimate genetic structure based on seed management.

Results of the model	**M-M**	**M-L**	**Ch-M**	**Ch-L**	**Total/Mean**
Estimated F_{ST} by SSR	0.114	0.045	0.089	0.077	0.111
Proportion pollen migration (m_g) assumed to fit model results to F_{ST} by SSR	0.0089	0.0179	0.0114	0.0101	0.0083
	Mixtec villages		**Chatino villages**		
Estimated F_{ST} by SSR	0.089		0.081		
Proportion pollen migration (m_g) assumed to fit model results to F_{ST} by SSR	0.0112		0.0112		
	Middle elevation villages		**Low elevation villages**		
Estimated F_{ST} by SSR	0.120		0.094		0.111
Proportion pollen migration (m_g) assumed to fit model results to F_{ST} by SSR	0.0085		0.0087		0.083
Proportion pollen migration (m_g) assumed to fit model results to F_{ST} by SSR setting seed migration frequency (m) to approximately 0[1] and seed mixture (p_m) to 0	0.0089		0.0089		0.086

[1] the model does not accept zero migrating ears, we used 0.000001

Table 3. Permutational multivariate analysis of variance on Euclidean distances matrices for plant morphology traits for each common garden (M-M, Ch-L), using 9999 permutations, two levels for ethnicity (Mixtec and Chatino) and two levels for elevation (low and middle elevation)

	Factor	Df	Sums Of Sqs	Mean Sqs	F. Model	Pr(>F)
M-M Common garden	Elevation	1	17.8	17.8	1.5	0.200
	Ethnicity	1	70.1	70.1	5.9	**0.000**
	Residuals	56	666.0	11.9	0.9	
	Total	58	754.0			
Ch-L Commom garden	Elevation	1	16.8	16.8	1.4	0.235
	Ethnicity	1	46.0	46.0	3.7	**0.010**
	Residuals	56	691.2	12.3	0.9	
	Total	58	754.0			

Table 4. Genetic structure as revealed by AMOVA and F_{ST}.

Source of variation	Sum of squares	Variance components	Percentage variation	Fixation indices	p-value
Between ethnic groups	48.40	0.21	4.04	$F_{CT} = 0.040$	0.000
Among villages within ethnic groups (elevation effect)	27.85	0.09	1.75	$F_{SC} = 0.018$	0.000
Among individuals within villages	997.69	1.13	21.28	$F_{IS} = 0.226$	0.000
Within individuals	648.00	3.85	72.92	$F_{IT} = 0.271$	0.000
Total	1721.94	5.29			

Notes: p-value calculated after 16,000 permutations.

Titles and legends to figures

Figure 1. Map of villages and collections

Figure 2. Principal components analysis plot (PC1 vs PC2) for morphological traits sorted by ethnicity, and elevation, data from middle and lowlands common gardens.

Figure 3. Structure graphical results assuming two (a) and four (b) groups, after a burning period of 30,000 iterations and 1,000,000 replications for estimations. Each individual plant is represented by a vertical line. Each color represents the membership to each cluster (k). Labels in the x axis show the village of origin.

Figure 1. Map of villages and collections

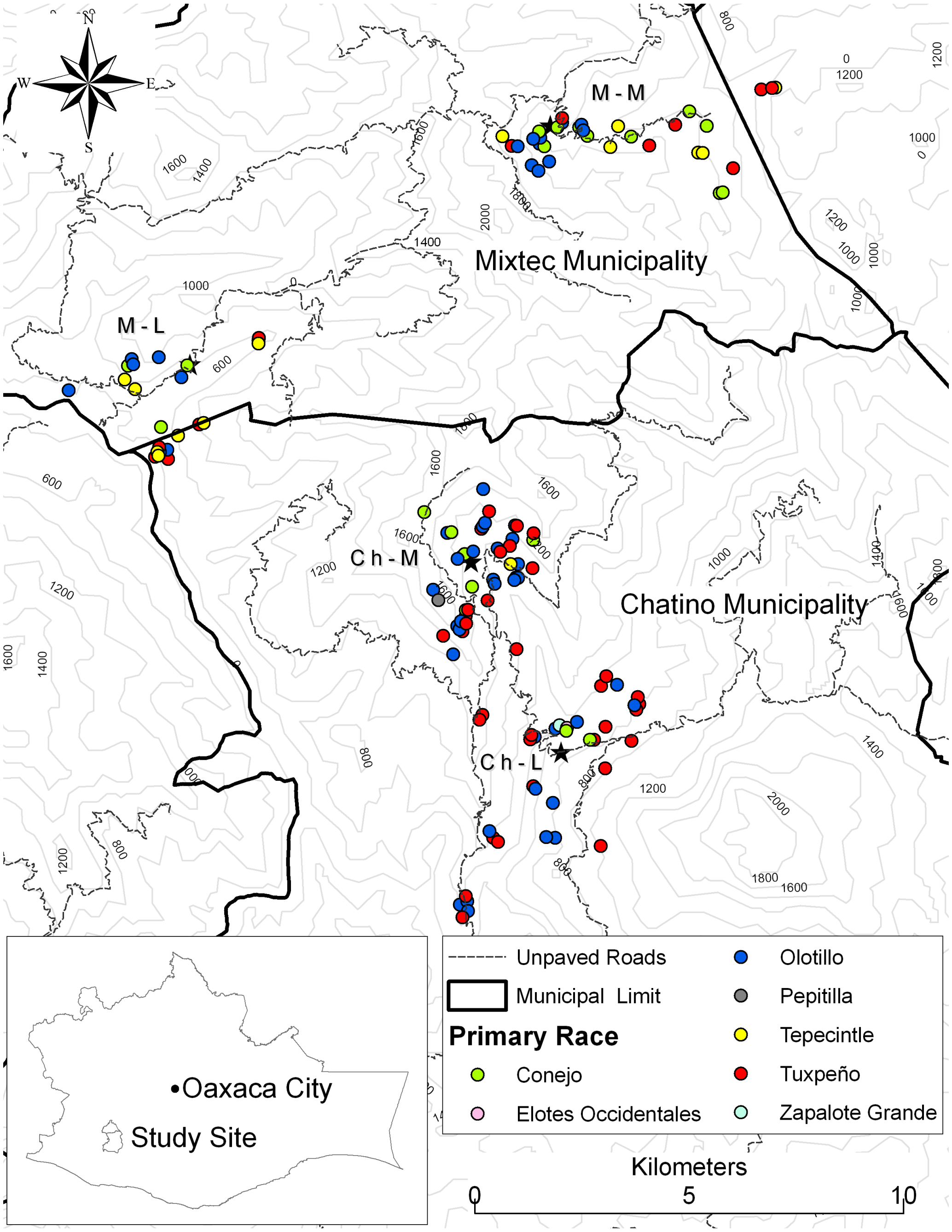


Figure 2. Principal components analysis scores plot (PC1 vs PC2) for morphological traits sorted by ethnicity, and elevation, data from middle and lowlands common gardens.

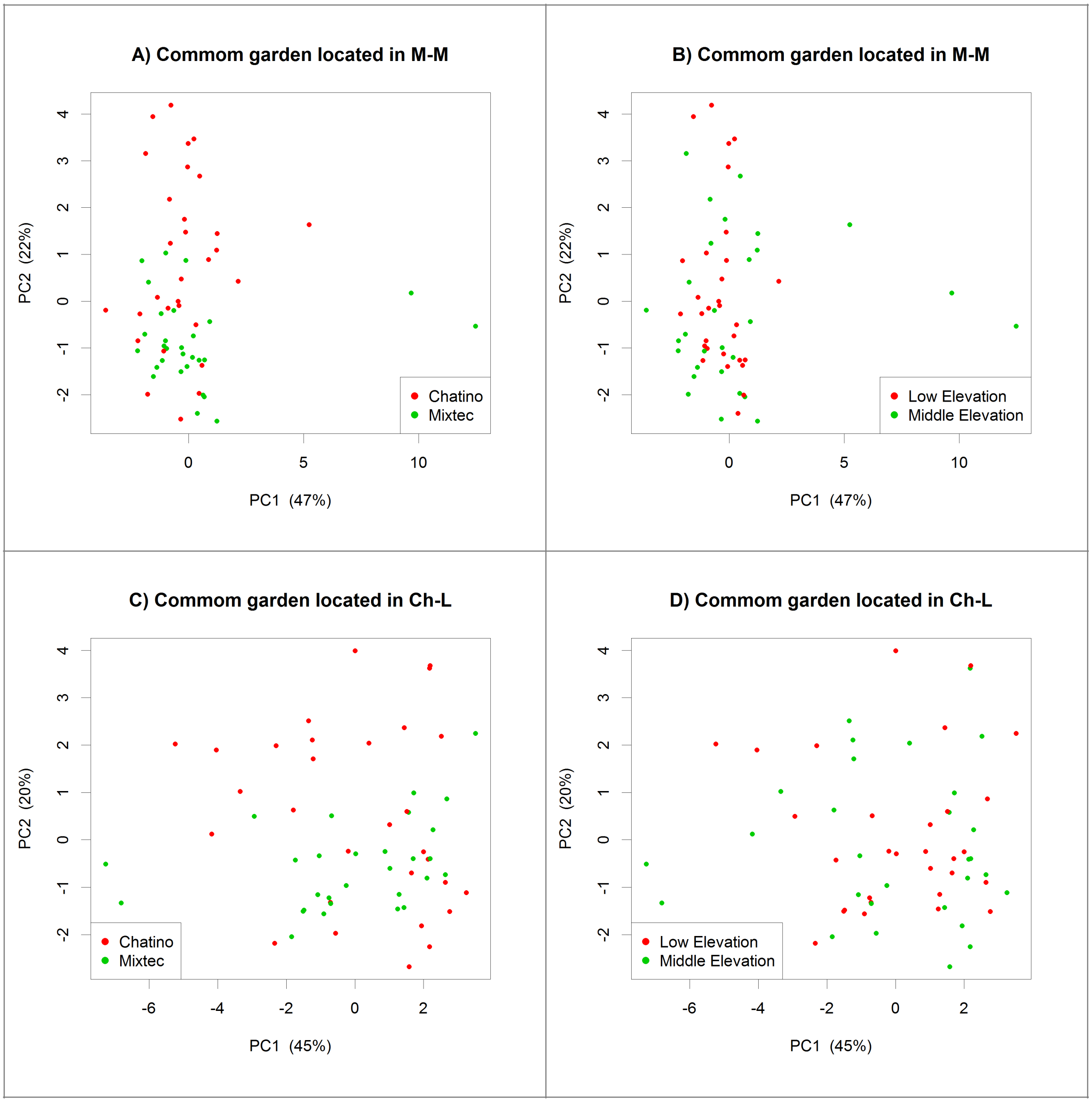
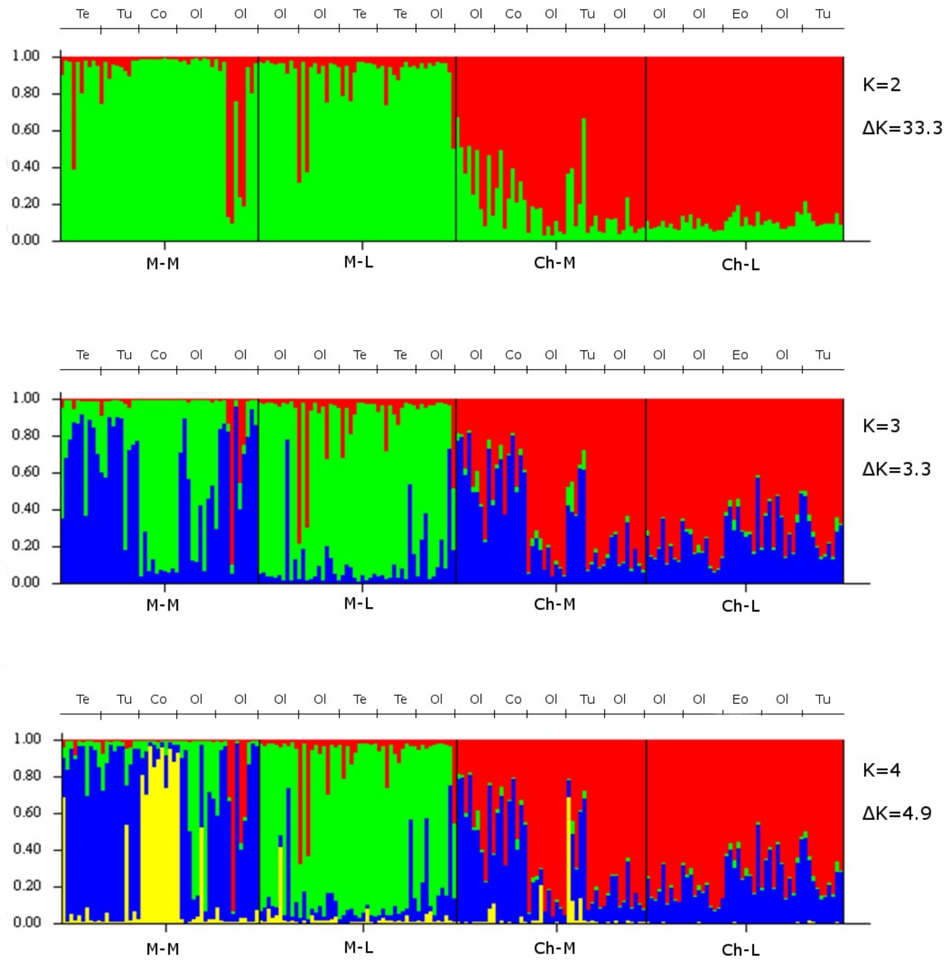

A) Commom garden located in M-M
B) Commom garden located in M-M
C) Commom garden located in Ch-L
D) Commom garden located in Ch-L
PC2 (22%)
PC1 (47%)
PC2 (20%)
PC1 (45%)
4
3
2
1
0
-1
-2
0
5
10
-6
-4
-2
0
2
● Chatino
● Mixtec
● Low Elevation
● Middle Elevation

Figure 3. STRUCTURE results assuming from two to seven groups (K), after a burning period of 100,000,000 iterations and 100,000 replications for estimations. Each individual plant is represented by a vertical line. Each color represents the membership to each cluster (K). Labels in the bottom axis show the village of origin and labels in the top axis show the race, ΔK was calculated after 10 runs.



Maize races; Te: Tepecintle, Tu: Tuxpeño, Co: Conejo, Ol: Olotillo, Eo: Elotes occidentales

APPENDIX

Table A.1. Simple Sequence Repeat loci used in molecular analysis.

Locus	BIN #	Repeat	Allelic Range (IBPGR)
phi051	7.06	AGG	136-154
phi115	8.03	ATAC	291-312
phi015	8.08	TTTG	76-113
phi033	9.02	CCT	224-270
phi053	3.05	ATGT	169-213
phi072	4.01	CAAA	134-163
phi093	4.08	CTAG	272-296
phi024	5.00	CCT	354-376
phi085	5.06	GCGTT	233-266
phi034	7.02	CCT	118-160
phi121	8.04	CCG	93-105
phi029	3.04	AG-AGGG	139-176
Phi073	3.05	AGC	184-203
phi96342	10.X	ATCC	223-256
phi109275	?	AGCT	119-144

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1192v1 | CC-BY 4.0 Open Access | rec: 23 Jun 2015, publ: 23 Jun 2015