The role of historical and contemporary environmental factors in the distribution of genetic diversity in the teosinte in Mexico

Gasca-Pineda, Jaime[1*], Yocelyn T. Gutierrez-Guerrero[1], Erika Aguirre-Planter[1], Peter Tiffin[2], Luis E. Eguiarte[1*]

[1] Laboratorio de Evolución Molecular y Experimental, Instituto de Ecología, Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, México.

[2] Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108, USA

* Correspondence: Jaime Gasca-Pineda and Luis E. Eguiarte, Laboratorio de Evolución Molecular y Experimental, Instituto de Ecología, Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, AP 70-275, Coyoacán, 04510, México DF, México. fruns@unam.mx

Keywords: genetic structure, microsatellite markers, demographic inference, isolation by distance, isolation by environment

Abstract

The wild maize, commonly known as teosinte, has a wide distribution and is abundant in Mexico. Previous reports indicate that the environmental conditions are determinant in the

distribution and amount of genetic diversity. In this work, we used a set of neutral markers in a large number of samples and localities to explore the influence on genetic diversity of historical environmental shifts. We genotyped 527 individuals from 29 localities using a set of 22 microsatellites. We found high levels of genetic diversity and significant values of structure among localities. Also, significant values of heterozygote deficit and variable levels of selfing were recovered. The historical demographic analysis implemented concur with the distribution modeling analyses. We conclude that the genetic diversity of teosinte is a mixture of local events and historical factors.

Introduction

The *Zea mays* species include the cultivated (*Zea mays ssp. mays* L.), and wild maize populations, commonly known as teosintes. From the current recognized teosinte subspecies, *Z. m. parviglumis* and *Z. m. mexicana* are the most abundant and are widely distributed in central-west México (Aguirre-Liguori et al. 2016). *Z. m. parviglumis* grows in the lowlands (<1900m above sea level) in warm and moist conditions, while *Z. m. mexicana* is found in colder and dryer highlands (>1500m above sea level) (Aguirre-Liguori et al. 2016). Aside from the intrinsic interest due to their relation with the cultivated crops (Aguirre-Liguori et al. 2016), the wide distribution of teosintes and contrasting environmental conditions make them an attractive system to study microevolutionary process, including genetic divergence, local adaptation, and phylogeography.

Previous studies dealing with teosinte genetic diversity were focused on domestication processes (Fukunaga et al. 2005; van Heerwaarden et al. 2011), particular genes (Moeller et al. 2007; Ross-Ibarra et al. 2009), and complete genomes (Fustier et al. 2017). Due to their specific objectives, however, these studies used broad geographic scales and/or a large number of localities with few individuals. Conversely, other works implemented a high number of individuals, yet using a few molecular markers, either isozyme or microsatellites (Buckler et al. 2006; Fukunga et al. 2005), and more recently, some works implemented SNPs (Pyhäjärvi et al. 2013; Aguirre-Liguori et al. 2017). In general, these studies indicated that the diversity and distribution of teosintes are strongly affected by environmental factors (Aguirre-Liguori et al. 2017). Thus, it is expected that historical events like the Last Glacial Maximum (LGM) and the Younger dryas (Reference) influenced the distribution range of teosinte populations, with a concurrent influence on their genetic diversity. Nevertheless, most studies on teosintes left open questions about the fine details of microevolution, as genetic structure, past demographic events, selfing rates, and the geographical distribution of genetic diversity. Thus, relevant questions have remained: Are there high levels of genetic structure among teosinte populations, or they comprise a genetic continuum? The past climate changes influenced the amount and distribution of current genetic diversity? Are there local factors, like selfing and local adaptation shaping genetic diversity? Using neutral molecular data, is there evidence of gene flow between subspecies?

In this sense, neutral molecular markers, such as microsatellites, offer a wide variety of robust analyses (Selkoe and Toonen 2006; Guichoux et al. 2011; Kalia et al. 2011) including

demographics and divergence events. Even though the appealing analyses that could be performed using microsatellites, recent studies demonstrated that in order to achieve reliable conclusions from microsatellites it is recommended to use a high number of loci (>20) and a good representation of the species geographic distribution including a large number of populations and individuals (Aguirre et al. 2019 in press). Moreover, in the case of *Zea mays*, there are public resources that allows to select microsatellite markers (https://www.maizegdb.org) to avoid physical linkage, that in addition, were already validated in studies of genetic diversity in maize (Fukunaga et al. 2005; Orozco-Ramírez et al. 2016).

In this work, we used a set of 22 microsatellite loci to conduct thorough analyses of the amount and geographic distribution of genetic diversity of *Z. m. parviglumis* and *Z. m. mexicana* in Mexico. The studied localities encompassed a wide range of environmental conditions and most of the geographic distribution of both subspecies (Aguirre-Liguori et al. 2017). We estimated the genetic diversity, genetic structure, inbreeding and selfing levels, and identified genetic barriers among localities of 29 teosinte localities. Additionally, we used present and past environmental data to evaluate the influence of past environmental conditions in the demographic history of teosintes.

Materials and Methods

Sampling, DNA extraction and genotyping

We sampled 29 sites including 14 localities of *Zea mays parviglumis* and 15 localities *Zea mays mexicana* subspecies (Fig. 1, Table S1, Supplementary Materials) encompassing a broad altitudinal and environmental gradient of teosinte populations. Seeds were collected from 10-26 plants per site, and one seed per plant was germinated in greenhouse obtaining a total of 527 plants. DNA extraction was carried out using a modified version of the CTAB extraction method (Doyle and Doyle 1987).

To evaluate the genetic diversity of teosintes, we amplified a set of 22 microsatellite loci previously used for maize genetic diversity analyses (Table S2). To avoid physical linkage, loci were selected considering their chromosomal position and distance within chromosomes (>150cM). The microsatellites names and primer sequences are reported in Supplementary Materials (Table S2). The PCR products were visualized on 2% agarose gels stained with ethidium bromide. The fluorescent-labeled PCR products were genotyped in the UIUC Core Sequencing Facility (https://unicorn.biotech.illinois.edu/) using GeneScan LIZ 600 size standard (Applied Biosystems). Allele sizes were determined using Genemapper v 3.7 (Applied Biosystems). Presence of genotyping errors and null alleles were assessed using MICRO-CHECKER v 2.2.3 (Van Oosterhout et al. 2004) using the Bonferroni correction for multiple comparisons. To evaluate potential bias due to null alleles in genetic structure estimates, we use the FreeNa software (Chapuis and Estoup 2007) to implement a statistical test for differences between the standard FST versus the FST with the ENA (Excluding Null Alleles) correction.

Population Genetics Analyses

We estimated the commonly used population genetics summary statistics: observed number of alleles (*Na*), the rarefied allelic richness (*Ar*, rarefied to the minimum sample size), the observed (*HO*) and expected heterozygosity (*HE*), using the package "adegenet" version 2.1.1 (Jombart 2008) implemented in R (R Core Team 2015). To detect deviations from Hardy-Weinberg proportions, we performed the test of excess or deficiency of heterozygotes by means the fixation index (FIS) (Wright 1965; Weir and Cockerham, 1984) using Genepop version 4.6 (Rousset 2008). Its significance was calculated using the Fisher exact test with 10,000 dememorisation steps, 1,000 batches, and 5,000 iterations per batch. To approximate the degree of selfing in teosinte localities, we calculated the sg^2 estimate based on multilocus comparison of linkage disequilibria using the software RMES (David et al. 2007).

To evaluate the genetic structure among localities, we use the estimated the R package DEMEtics (Gerlach et al. 2010) to estimate the Jost's Dest. We selected this measure because the standard FST may underestimate the population differentiation due to the high mutation rates of microsatellites (Balloux and Lugon-Moulin 2002), and large population sizes of teosintes. We calculated the 95% confidence intervals (C.I.) as a measure of significance; if Dest C.I. values did not reach zero, we took the value as significant. Additionally, we performed a Principal Component Analysis (PCA) to evaluate the allelic differences at the individual level, using localities as point classes with the "adegenet" R package.

To detect genetic grouping among individuals and localities, we run Structure v2.3 (Pritchard et al. 2000) using the admixture with correlated allele frequencies model (Falush et al. 2003). We performed 25 independent runs for each value of *K* from 1 to 30 using 1,000,000 MCMC steps followed by 500,000 steps as burn-in. To find the most likely number of clusters (*K*) we used the method proposed by Evanno et al. (2005) implemented in Structure Harvester (Earl 2012). Also, all lnP(D) values were plotted against k to choose the *K* value with higher likelihood and lower variance. We performed the analyses for the two subspecies as a whole, as well as each subspecies alone.

Gene flow was approximated by means the recent migration rates (m) estimated by BayesASS 3.0.3 (Wilson and Rannala, 2003). As BayesASS does not perform well with multiple gene pools within populations (Wilson and Rannala, 2003), we used the genetic clusters obtained by Structure. To ensure good chain-mixing chain, we performed several test runs varying the values of parameters to achieve acceptance rations of 0.25-60 as recommended by BayesASS authors. The software was run for 10x109 Markov chain Monte Carlo iterations sampling every 100,000 steps, and discarding the first 10% of the iterations as burn-in period. MCMC convergence was assessed by plotting the traces for all parameters using the R package “coda”. In order to detect genetic barriers among localities, we implemented the Monmonier algorithm (Monmonier 1973) using “adegenet". To perform this analysis, we used a Gabriel graph as connection network among localities and used the localities scores obtained from the previously PCA as distance matrix. The distance threshold between immediate neighbors was chosen as an abrupt decrease between connected points. We use the function “optimize.monmonier” to obtain the best

boundaries.

Isolation by Distance and Isolation by Environment

To approximate Isolation by Distance (IBD) patterns for each subspecies, we performed standard mantel tests (Mantel, 1967) using paired Nei genetic distances between localities versus the log10 of geographic distances in meters. The Mantel tests and their significance (10,000 random replications) were estimated by means the R package "vegan" (Oksanen et al. 2009), while the geographic distances using the "raster" package (Hijmans and van Etten 2016). We evaluated the relative effects of geographical distance and environment on genetic differentiation using "Sunder" (Botta et al. 2014). This method uses the decay in covariance in allele frequencies between populations as a function of the geographical and ecological (here, the environmental distance matrix) distances. We implemented the three models available, *G* where the geographic distance alone explains the decline in covariance in allele frequencies, *E* where the environmental distance explains the decline and *G*+*E* where both distances explain equally the decline in the allele frequencies. We ran the analysis using 10 million of iterations, sampling every 1,000 iterations. The initial state and upper bounds of the Dirichlet prior parameters distribution were implemented as suggested by the software authors. The convergence was assessed by plots of the parameters traces, and we used the 10 % of the data as a validation set for the cross-validation step. The environmental distance matrix was generated using the euclidean distance of the bioclim dataset obtained for the SDM analyses (see below).

Demographic analyses

We approximate changes in the effective size (N_e) of teosinte populations using the R package "VarEff" v1.2 (Nikolic and Chevalet 2014). This software estimates the effective sizes from present to an ancestral time by the simulation of demographic steps of constant size using a coalescent approach. The model calculates posterior probabilities by means the approximation of likelihood of each step, and plots the results in terms of effective size through time. We use previous reported mutation rate for microsatellites in teosinte 0.00077 (Vigouroux et al. 2002), to approximate the time in generations of change in effective size. The analyses considered the genetic clusters obtained by Structure. We run the program for 100,000 steps with a batch length of 10 steps and sampling every 10 steps with an acceptance ratio of 0.25. To avoid a strict stepwise mutation model, we used the Two-Phase model (Di Rienzo et al., 1994) with the parameter of proportion of mutational events (c) fixed at 0.15.

Spatial Distribution Modeling

We gathered a total of 375 georeferenced points including this study and previous published records of *Z. m. parviglumis* and *Z. m. mexicana* (conabio.org). In order to reduce bias due to autocorrelation caused by local overrepresentation, we generated a 1/10 degree grid and then selected one random point using the R package "raster". The final data set included 62 points for *Z. m. mexicana* and 69 points for *Z. m. parviglumis*. Therefore, we obtained the environmental layers from the Worldclim database (http://worldclim.org/bioclim, Hijmans et al., 2005). To

reduce variable redundancy, we conducted a paired Spearman's rank correlation test and excluded variables with values >|0.75|. The final set included the following ten variables: BIO2 (Mean Diurnal Range), BIO3 (Isothermality), BIO4 (Temperature Seasonality), BIO8 (Mean Temperature of Wettest Quarter), BIO12 (Annual Precipitation), BIO14 (Precipitation of Driest Month), BIO15 (Precipitation Seasonality), BIO17 (Precipitation of Driest Quarter), BIO18 (Precipitation of Warmest Quarter), BIO19 (Precipitation of Coldest Quarter).

To generate the Species Distribution Models (SDMs) we used the platform "biomod2" v 3.1 (Thullier et al. 2014) in R. The SDMs were constructed by means ensemble forecasting models (Araújo and New, 2007) using the committee-averaging criteria. We implemented four modelling algorithms in our analysis: Generalised Linear Model (GLM, McCullagh and Nelder, 1989), Generalized Boosted Model (GBM, Friedman 2001), MaxEnt (Phillips et al., 2006) and random forest (RF, Breiman 2001). The presence data set was split into 70% for model calibration and 30% for model evaluation. Using this 70%-30% criteria, five random-sampled data sets were run for each algorithm. Therefore, we generated five pseudo-absence datasets of 5,000 points selected at random. Model performance was assessed using the true skill statistic (TSS) and the area under the receiver operating characteristic curve (AUC), models that did not meet thresholds of TSS>0.8 and 0.95>AUC, were discarded. Overall, we computed eighty models (two presence-absence data sets, five evaluation runs and four algorithms, two subespecies). The final ensembles were transformed to binary points selecting the threshold that maximized the TSS score.

Results

Overall, we detected a total of 355 alleles for the 22 microsatellite analyzed loci, with an allelic range of 6 to 30 alleles per locus (all values are reported in Table S3). The overall genetic diversity was HO=0.565 (s.d.=0.104), and HE=0.818 (s.d.=0.077) (Fig. 2a). At subspecies level, *Z. m. mexicana* showed HE=0.798 (s.d.=0.083), HO=0.563 (s.d.=0.120) and Ar=4.198 (s.d.=0.760), while *Z. m. parviglumis* HE=0.818 (s.d.=0.095), HO=0.538 (s.d.=0.095), Ar= 3.948 (s.d.=0.683). The HE estimates were statistically different ($p< 0.05$) between subspecies, suggesting that *Z. m. parviglumis* had higher genetic diversity. Furthermore, we found a wide range of genetic diversity among localities, ranging from HE=0.364 (s.d.=0.172) to HE=0.769 (s.d.=0.115) and an allelic richness from Ar=2.335 (s.d.=1.172) to Ar=4.749 (s.d.=1.218) for pVP and pCH respectively (Fig. 2a, Table S3). Worth to say that we did not find a statistical correlation between diversity and locality sample size ($r2<0.01$, $p>0.2$).

We found an overall F_{IS} value of 0.186 ($p<0.001$), while Z. *m. parviglumis* displayed F_{IS}=0.3374 and *Z. m. mexicana* F_{IS}=0.2621 (Fig. 2b) respectively. In general, we found high F_{IS} values among localities (Fig. 2b), ranging from 0.0863 (pTC) to 0.3025 (pHU), with higher fluctuation for *Z. m. parviglumis*. Only the Z. *m. parviglumis* localities pGU and pTC were in the H-W proportions. According to the F_{IS} values, Micro-Checker identified the presence of at least one null allele in each locality. However, we did not find any locus with a consistent null-allele signal across populations. Moreover, the global F_{ST} obtained with the ENA correction was not significantly different from the standard (not corrected) F_{ST} value, so we keep all loci for all

posterior analyses. The selfing estimation by means RMES software displayed fluctuating sg^2 values for both teosinte subespecies (Fig. 3); however *Z. m. parviglumis* localities displayed higher sg^2 values ranging from zero (no selfing) to 0.309, while *Z. m. parviglumis* displayed a range from zero to 0.133. In both subespecies, only one third of selfing estimates were significant. Also, we do not observed geographical pattern of the significant values of selfing,

We obtained an overall Dest value of Dest=0.433, and of Dest=0.147 between subspecies, and found contrasting values of genetic structure among teosinte localities (Fig. 4; Table S3). The Western-Jalisco localities (pVP and pTP) along with Oaxaca (pOX) were the most differentiated from the rest of localities. Therefore, we observed the highest values between the *Z. m. parviglumis* pHU and pVP (Dest=0.796), while the lower values among the Michoacan localities (mAC1, mAC2, mAC3, mSAM and mHU30; mean Dest=0.136) of *Z. m. mexicana*.

The Structure analysis implementing the Evanno test and the ln(P) plots indicated the presence of four genetic clusters (Fig. 5a). The analysis divided the localities in subspecies and each subspecies in two groups. *Z. m. parviglumis* displayed a subdivision between the Western-Jalisco localities (pVP and pTP) and the rest of the localities, while in *Z. m. mexicana*, the subdivision was between the localities of Michoacan and the rest of *Z. m. mexicana*. Moreover, the analysis showed admixed individuals between subspecies (Fig. 5a). It is worth saying that we obtained similar results when analyzing each subspecies in separated way. The recent-migration analysis conducted with BayesASS yield very low migration rates between the clusters obtained with Structure (Fig. S3). The higher rates (0.009) were observed between Western-Jalisco and and *Z.*

m. parviglumis and Michoacan clusters. According to the Dest and Structure results, the PCA displayed that the Western-Jalisco localities (pVP and pTP) had the most genetically divergent individuals (Fig. 5b), but also displayed genetic closeness among some individuals. The recent migration rates analysis performed with BayesAss, displayed low migration rates among Structure clusters (Fig. S3). On the other hand, the Monmonier analyses revealed significant boundaries between the Western-Jalisco localities and the rest of the localities (data not shown). However, due to the possible bias generated by these genetic divergent localities (as displayed by Dest and PCA), we rerun the Monmonier analysis excluding pVP and pTP, and we found significant boundaries among the Michoacan localities (Fig. 5c); moreover, the analysis did not detect significant boundaries within localities of subspecies, nor between subspecies among the eastern localities.

The Mantel tests among genetic distance against geographic and ecological distances showed significant and positive correlations in for both distance matrices in both subspecies ($p<0.01$, $r2>0.25$; Fig. S3). Nevertheless, the Sunder analysis showed the higher probability for the environmental (*E*) model, than the geographic (*G*) and the geographic+environment (*G*+*E*) models in both subespecies, indicating that the environmental conditions have a stronger influence on the change of allelic frequencies than the geographic distance.

The VarEff tests for historical demographic change displayed shifts in the past effective size in the four Structure clusters. In all cases, we observed an ancient and lower effective size followed by a demographic growth and a posterior reduction (Fig. 7). Furthermore, we observed that *Z. m.*

mexicana had a larger past effective size than *Z. m. parviglumis*. In addition, the current effective sizes were rather small.

The SDM displayed that subespecies had different geographic distribution with a ~3.5 % of overlapping distribution on the total area. This overlapping areas were located in the boundaries of the distribution of subspecies (Fig. 6). Furthermore, both subspecies displayed fluctuating areas in the four different time-periods. In the LIG (~120kys) period we observed a very reduced distribution area, while in contrast, in LGM (~21kys) we observed an increase in the area of potential distribution. In the MID (~6kys) we observed an acute reduction in the potential distribution, followed by a drastic increment (Fig. 6) in the potential distribution. The severe reduction in the predicted suitable area in MID is related to the contrasting environmental values at this period (Fig. S1).

Discussion

Our analyses of the genetic diversity on current populations of Z. m. parviglumis and *Z. m. mexicana* revealed the influence of historical processes. The range of genetic diversity of localities, as well as the different values of the genetic structure, showed that the wild maize had been under global and local pressures that shaped the genetic diversity. At the local scale, heterozygosity and the number of alleles are consistent with previous reports on teosinte. However, the low diversity detected in Western-Jalisco populations could be related to genetic drift due to genetic isolation; this was confirmed by the low effective size observed for these localities (see below). Besides, the FIS analyses showed that most teosinte localities had from

moderate to high levels of heterozygote deficiency, and only two localities were in the H-W proportions. This result suggests an ongoing evolutionary process that impedes stable allele frequencies. The significant and negative FIS values would indicate inbreeding within localities, but also a Wahlund effect due to an incomplete gene pool sampling. In this sense, the selfing analysis values displayed significant values indicating the presence of selfing associated with non-random mating.

The genetic structure observed between localities suggested a degree of isolation among localities that lead to genetic clustering within subspecies, as shown by the Western-Jalisco and Michoacan clusters. In this sense, the Western-Jalisco localities (pVP and pTP) were the most differentiated even with relatively close localities (pGU, pEJ, and pTO). Also, the Oaxaca (pOX) locality was highly structured; however, this was expected as this locality the most geographically distant. Besides, Structure revealed four main clusters that were not strictly correlated with the geographic closeness of localities, and on the other hand, the PCA confirmed the separation of the genetic clusters. The genetic barriers among subspecies (previous exclusion of Western-Jalisco localities). Moreover, the contrast of the admixed individuals of Structure clusters with the low migration rates estimated could be explained by the presence shared ancestral alleles, more than an effective migration among subspecies.

The Sunder analysis showed that the environment is a determinant factor for genetic differentiation in teosinte. This result suggests possible local adaptation that promotes the genetic divergence between populations, even in the presence of genetic geneflow. Moreover, the

environmental conditions were notably different among genetic clusters (even within subspecies) (Fig. S4). This was also confirmed by the current SDM projections, as showed by the reduced proportion of suitable area shared by subspecies. In this context, all results indicated that the environmental requirements of teosintes are a significant determinant of their genetic divergence. Also, the results suggest that even in the presence of gene flow, the habitat selection would keep the subspecies as separate genetic lineages (and possibly within subspecies).

The demographic results agreed with the different evaluated periods in SDM analyses. reflecting the distinct environmental preferences The SDM showed as previous reports, the environmental conditions of teosintes can be restricted for both drastic changes in the distribution suggesting that the past environmental conditions were not suitable for teosintes populations.

Acknowledgments

We thank L. Espinosa-Asuar and M. Rosas-Barrera for their technical support, and Alberto Villasante Barahona for his support in generating the microsatellite datasets. We thank Sivia Barrientos for her support during the elaboration of this investigation. This work was funded by grants CB2011/167826 (CONACYT Investigación Científica Básica), CN-10-393 (UC MEXUS-CONACYT), and M12-A03 ECOS Nord France CONACYT-ANUIES.

References

Aguirre-Liguori, J. A., Aguirre-Planter, E., & Eguiarte, L. E. (2016). Genetics and ecology of wild and cultivated maize: domestication and introgression. In *Ethnobotany of Mexico* (pp. 403-416). Springer, New York, NY.

Aguirre-Liguori, J. A., Tenaillon, M. I., Vázquez-Lobo, A., Gaut, B. S., Jaramillo-Correa, J. P., Montes-Hernandez, S., ... & Eguiarte, L. E. (2017). Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Molecular ecology*, *26*(16), 4226-4240.

Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in ecology & evolution*, *22*(1), 42-47.

Balloux, F., & Lugon-Moulin, N. (2002). The estimation of population differentiation with microsatellite markers. *Molecular ecology*, *11*(2), 155-165.

Botta, F., Eriksen, C., Fontaine, M. C., & Guillot, G. (2015). Enhanced computational methods for quantifying the effect of geographic and environmental isolation on genetic differentiation. *Methods in Ecology and Evolution*, *6*(11), 1270-1277.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Buckler, E. S., Gaut, B. S., & McMullen, M. D. (2006). Molecular and functional diversity of maize. *Current opinion in plant biology*, *9*(2), 172-176.

Chapuis, M. P., & Estoup, A. (2006). Microsatellite null alleles and estimation of population differentiation. *Molecular biology and evolution*, *24*(3), 621-631.

David, P., Pujol, B., Viard, F., Castella, V., & Goudet, J. (2007). Reliable selfing rate estimates from imperfect population genetic data. *Molecular ecology*, *16*(12), 2474-2487.

Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M., Slatkin, M., & Freimer, N. B. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences*, *91*(8), 3166-3170.

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin, 19, 11–1.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, *19*(1), 1-67.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources*, *4*(2), 359-361.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, *14*(8), 2611-2620.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*(4), 1567-1587.

Fukunaga, K., Hill, J., Vigouroux, Y., Matsuoka, Y., Sanchez, J., Liu, K., ... & Doebley, J. (2005). Genetic diversity and population structure of teosinte. *Genetics*, *169*(4), 2241-2254.

Fustier, M. A., Brandenburg, J. T., Boitard, S., Lapeyronnie, J., Eguiarte, L. E., Vigouroux, Y., ... & Tenaillon, M. I. (2017). Signatures of local adaptation in lowland and highland teosintes from whole‐genome sequencing of pooled samples. *Molecular ecology*, *26*(10), 2738-2756.

Gerlach, G., Jueterbock, A., Kraemer, P., Deppermann, J., & Harmand, P. (2010). Calculations of population differentiation based on GST and D: forget GST but not all of statistics!. *Molecular ecology*, *19*(18), 3845-3852.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., ... & Petit, R. J. (2011). Current trends in microsatellite genotyping. *Molecular ecology resources*, *11*(4), 591-611.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405.

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., & Dhawan, A. K. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica*, *177*(3), 309-334.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *25*(15), 1965-1978.

Hijmans, R. J., & van Etten, J. (2016). Raster: geographic data analysis and modeling. R package ver. 2.5–8.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, *27*(2 Part 1), 209-220.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, second ed. Chapman and Hall/CRC.

Moeller, D. A., Tenaillon, M. I., & Tiffin, P. (2007). Population structure and its effects on patterns of nucleotide polymorphism in teosinte (Zea mays ssp. parviglumis). *Genetics*, *176*(3), 1799-1809.

Monmonier, M. S. (1973). Maximum‐difference barriers: an alternative numerical regionalization method. *Geographical analysis*, *5*(3), 245-261.

Natacha Nikolic and Claude Chevalet (2014). VarEff: Estimation of effective population size and variation in the time scale. R package version 1.2. https://qgp.jouy.inra.fr

Oksanen Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2019). vegan: Community Ecology Package. R package version 2.5-5. https://CRAN.R-project.org/package=vegan

Orozco-Ramírez, Q., Ross-Ibarra, J., Santacruz-Varela, A., & Brush, S. (2016). Maize diversity associated with social origin and environmental variation in Southern Mexico. *Heredity*, *116*(5), 477.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling*, *190*(3-4), 231-259.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Pyhäjärvi, T., Hufford, M. B., Mezmouk, S., & Ross-Ibarra, J. (2013). Complex patterns of local adaptation in teosinte. *Genome biology and evolution*, *5*(9), 1594-1609.

Ross-Ibarra, J., Tenaillon, M., & Gaut, B. S. (2009). Historical divergence and gene flow in the genus Zea. *Genetics*, *181*(4), 1399-1413.

Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular ecology resources*, *8*(1), 103-106.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology letters*, *9*(5), 615-629.

Thuiller, W., Georges, D., Engler, R., & Breiner, F. (2014). Ensemble platform for species distribution modeling. *R Package Version*, 3-1.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535-538.

Vigouroux, Y., Jaqueth, J. S., Matsuoka, Y., Smith, O. S., Beavis, W. D., Smith, J. S. C., & Doebley, J. (2002). Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution*, *19*(8), 1251-1260.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, *38*(6), 1358-1370.

Wilson, G. A. and Rannala, B. (2003). Bayesian inference of recent migration rates using multi-

locus genotypes. Genetics 163: 1177–1191.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, *19*(3), 395-420.

Figures

Figure 1

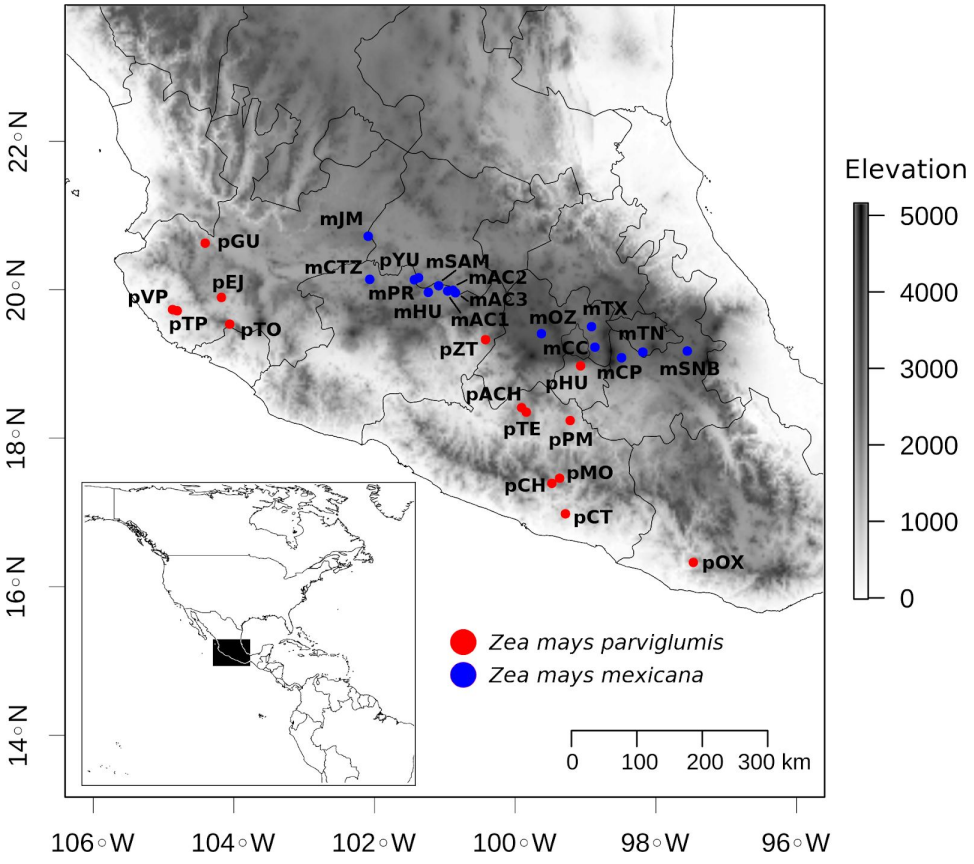


Figure 2

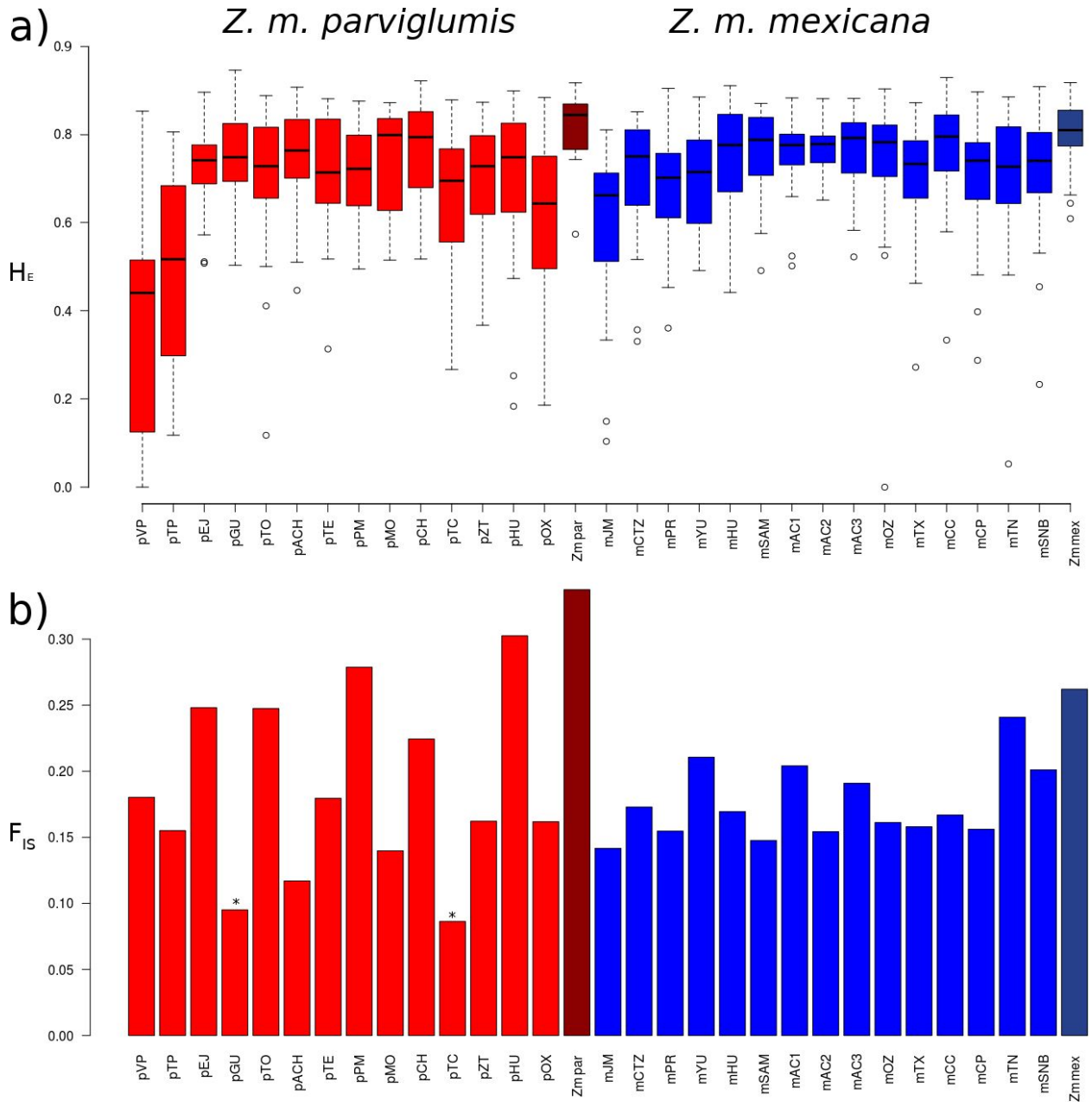

a)
Z. m. parviglumis
Z. m. mexicana
H_E
0.0
0.2
0.4
0.6
0.8
0.9
pVP
pTP
pEJ
pGU
pTO
pACH
pTE
pPM
pMO
pCH
pTC
pZT
pHU
pOX
Zm par
mJM
mCTZ
mPR
mYU
mHU
mSAM
mAC1
mAC2
mAC3
mOZ
mTX
mCC
mCP
mTN
mSNB
Zm mex
b)
F_IS
0.00
0.05
0.10
0.15
0.20
0.25
0.30
*
*

Figure 3

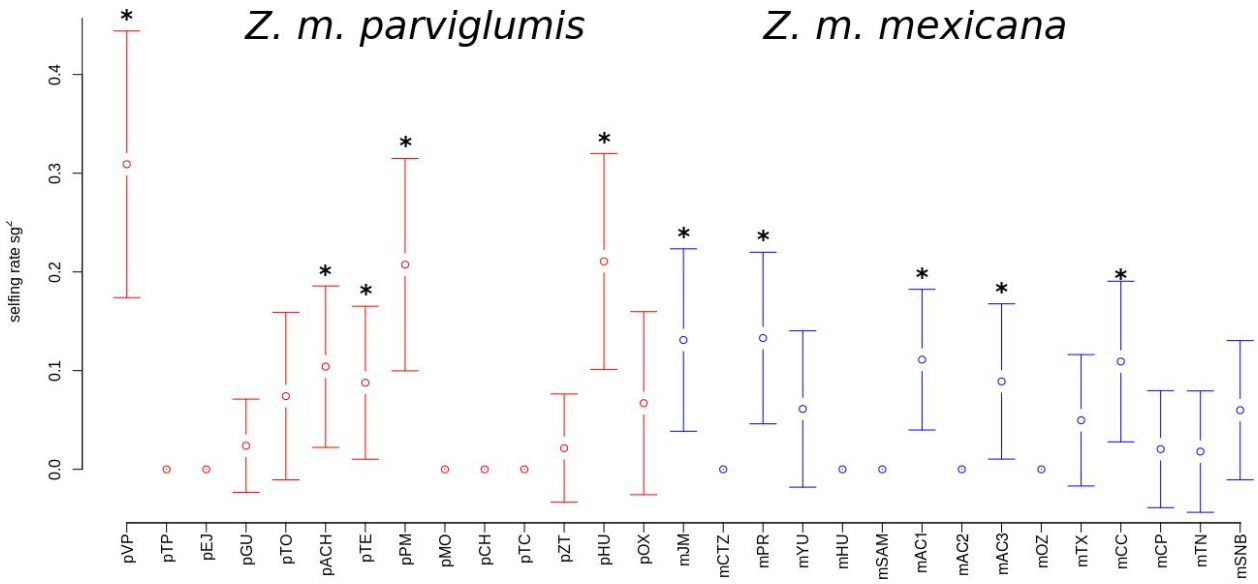

Z. m. parviglumis
Z. m. mexicana
selfing rate sg²
0.0
0.1
0.2
0.3
0.4
pVP
pTP
pEJ
pGU
pTO
pACH
pTE
pPM
pMO
pCH
pTC
pZT
pHU
pOX
mJM
mCTZ
mPR
mYU
mHU
mSAM
mAC1
mAC2
mAC3
mOZ
mTX
mCC
mCP
mTN
mSNB

Figure 4

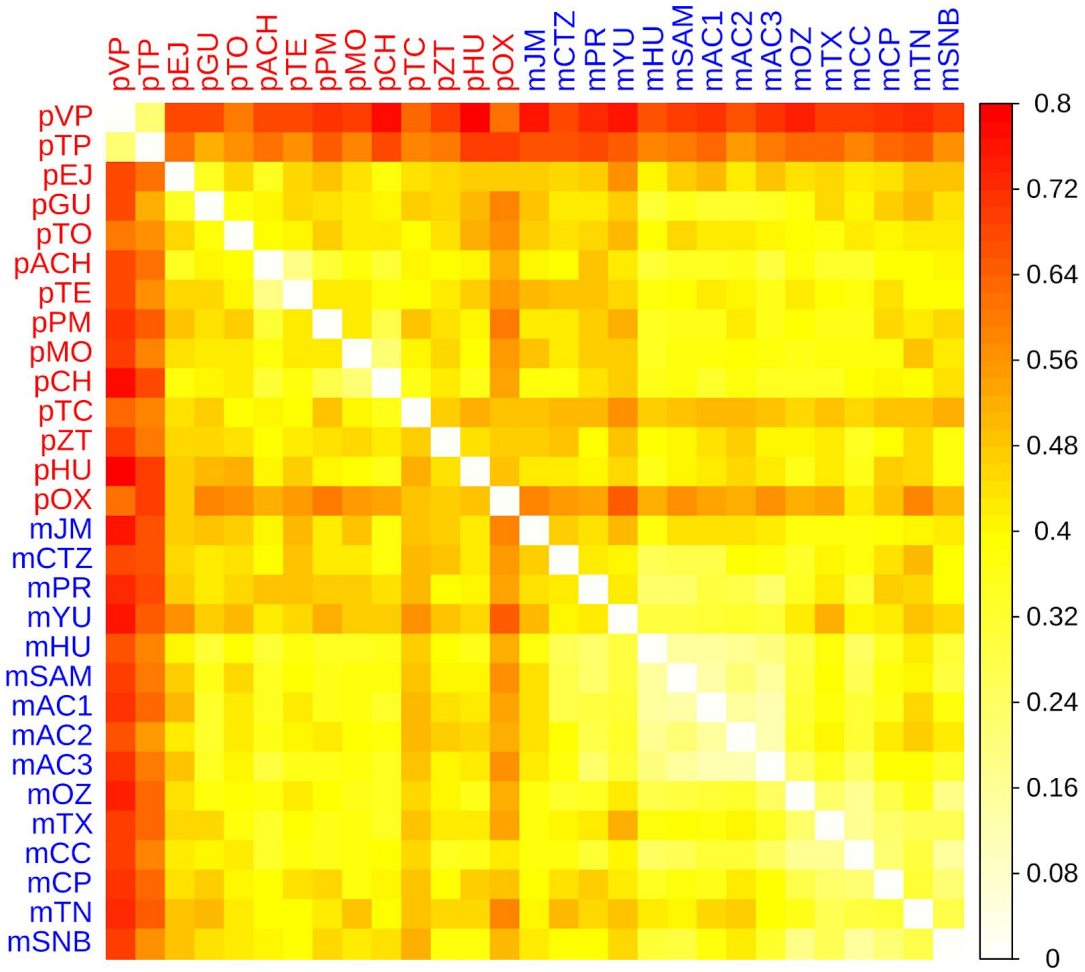

pVP
pTP
pEJ
pGU
pTO
pACH
pTE
pPM
pMO
pCH
pTC
pZT
pHU
pOX
mJM
mCTZ
mPR
mYU
mHU
mSAM
mAC1
mAC2
mAC3
mOZ
mTX
mCC
mCP
mTN
mSNB
0.8
0.72
0.64
0.56
0.48
0.4
0.32
0.24
0.16
0.08
0

Figure 5

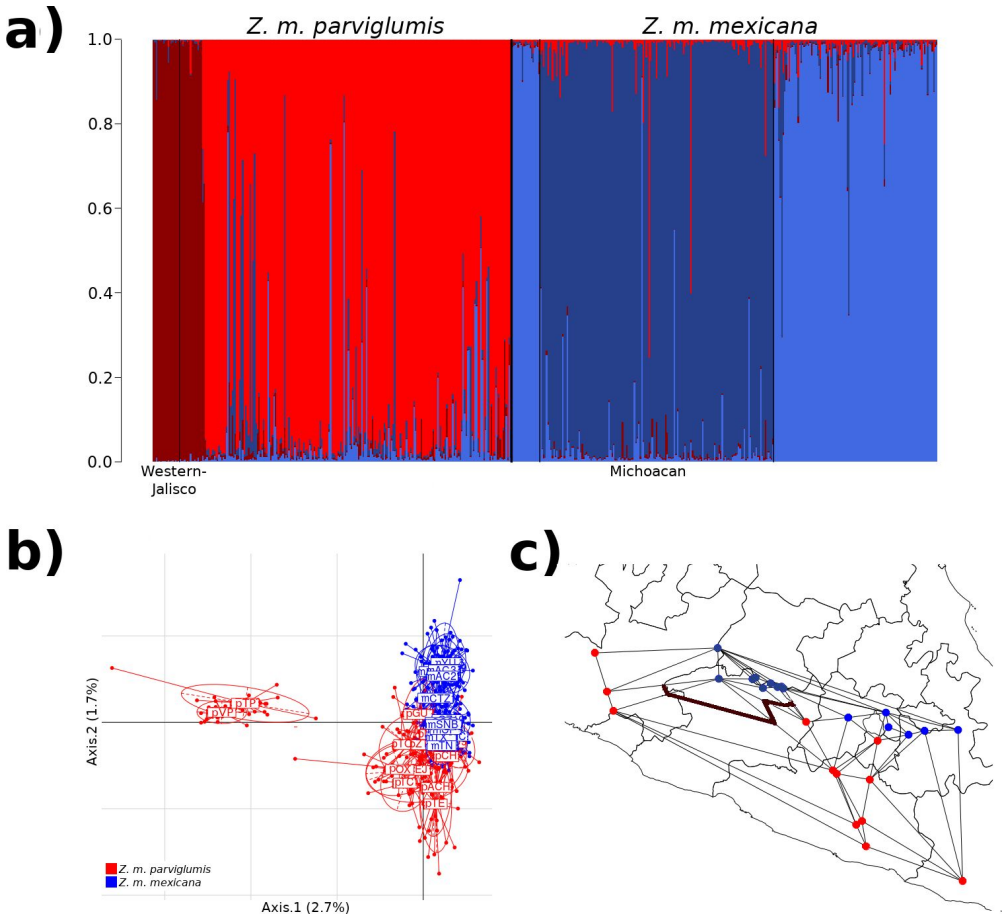


Figure 6

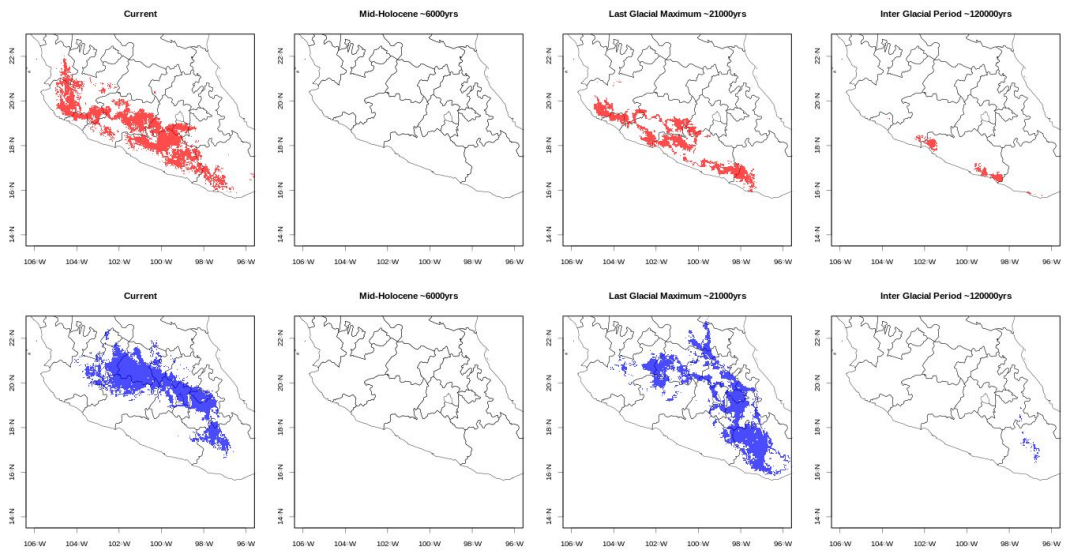
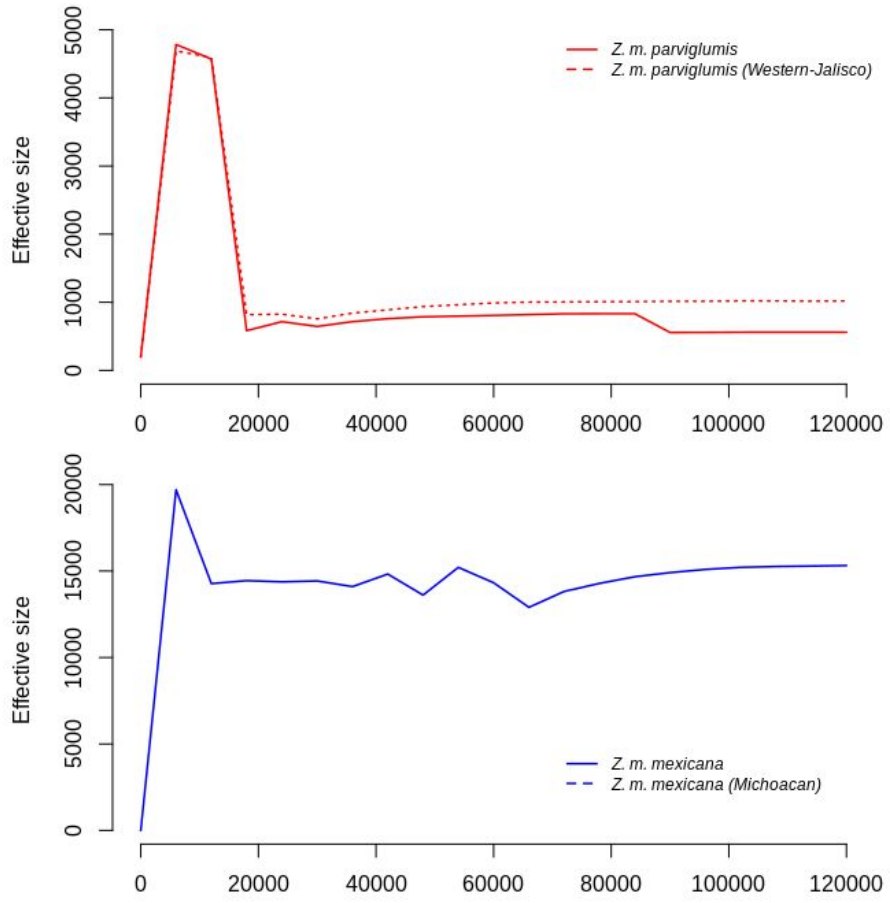


Figure 7



Figures Captions

Figure 1. Distribution map of teosinte samples. In red *Zea mays parviglumis*, in blue *Zea mays mexicana*.

Figure 2. a) Expected heterozygosity per locality for *Zea mays parviglumis* (red), and *Zea mays mexicana* (blue). Boxplots represents the dispersion of values among microsatellite loci. b) FIS values per locality, asterisks denote not-significant FIS values at $p<0.05$.

Figure 3. Selfing rates estimated by RMES. Bars represent standard deviation, asterisks denote values significantly different from zero.

Figure 4. Paired Dest values between all localities. Temperature color represent values.

Figure 5. a) Assignment plots obtained with Structure for K=3. b) Principal Component Analysis of individuals using localities as grouping factors. c) Monmonier analysis among localities excluding the Western-Jalisco localities (pTP and pVp).

Figure 6. Distribution Ensemble Models for *Zea mays parviglumis* (red) and *Zea mays mexicana* (blue) in four periods. Current, mid-holocene (MID, 6,000 years), last glacial maximum (LGM, 21000 years) and last interglacial period (LIG 120000, years).

Figure 7. Changes in effective sizes in the past estimated with VarEff. Time is represented in generations.