ORIGINAL ARTICLE

Genetic structure, phylogeography and potential distribution modeling suggest a population expansion in the mesquite *Prosopis laevigata* since the last interglacial

Gonzalo Contreras-Negrete[1] · Luis Letelier[2] · Javier Piña-Torres[3] · Antonio González-Rodríguez[1]

Abstract

Patterns of genetic and phylogeographic structure and recent population history of plant species in the Mexican arid zones have been scarcely investigated. *Prosopis laevigata* is the most widely spread species of mesquite in Mexico, with extensive populations in the arid and semiarid zones of the central and northern plateaus and scattered presence in southern Mexico. We evaluated the genetic and phylogeographic structure of this species to infer its recent demographic history. We genotyped six nuclear microsatellite loci and sequenced the psbA3´-trnH chloroplast DNA (cpDNA) region in individuals from 21 populations covering the whole distribution of the species. Nuclear genetic diversity was moderately high ($H_E = 0.527$), and genetic differentiation was moderate ($F_{ST} = 0.16$). A positive correlation between genetic diversity and latitude was observed. The cpDNA analyses indicated a lack of phylogeographic structure in *P. laevigata* ($G_{ST} = 0.090$, $N_{ST} = 0.101$; $P = 0.497$). Historical demography statistics indicated a population expansion supported by a skyline plot analysis, the star-like shape of the haplotype network, and the unimodal shape of the mismatch distribution. Ecological niche modeling suggested a contracted distribution into west-central Mexico during the Last Interglacial (~ 140 Ka), followed by an expansion in both northwards and southwards directions in the Last Glacial Maximum (~ 22 Ka), which continued in the mid-Holocene (~ 6 Ka) and the present. Results are congruent with a recent population growth and colonization of newly opened arid zones by *P. laevigata* populations. This pattern is consistent with the high capacity of colonization of nutrient-poor areas, high germination rates and resistance to drought reported for *Prosopis* species.

Keywords Arid zones · Phylogeography · Population expansion · Population genetics · *Prosopis*

✉ Antonio González-Rodríguez
agrodrig@cieco.unam.mx

[1] Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México (UNAM), Antigua Carretera a Pátzcuaro 8701, Morelia, Michoacán 58190, México

[2] Instituto de Ciencias Biológicas y Núcleo Científico Multidisciplinario, Universidad de Talca, 1 Poniente 1141, Casilla 747- 721, Talca, Chile

[3] Escuela Nacional de Estudios Superiores Morelia, Universidad Nacional Autónoma de México (UNAM), Antigua Carretera a Pátzcuaro 8701, Morelia, Michoacán 58190, México

Introduction

During the Pleistocene and the beginning of the Holocene many territories experienced the effects of climatic and ecological instability produced by the advance and retreat of glaciers, triggering drastic changes in plant communities (Metcalfe et al. 2000; Metcalfe 2006). This transition has been recognized as a main factor that influenced the distribution, population dynamics, and the current patterns of genetic diversity of plant species (Ramírez-Barahona and Eguiarte 2013; Pedersen et al. 2015). For the Nearctic, paleoecological records and genetic patterns show the migration of temperate forests toward lower latitude and altitude in response to the advance of the glacial layers from north to south (Roberts and Hamann 2015; Napier et al. 2019). In contrast, in the subtropical and tropical regions within the Mexican territory, during glaciations, the establishment and geographic expansion of plants with

higher affinity to cooler climates have been described, while tropical species were restricted to refugia where conditions allowed them to persist (Gugger et al. 2011; Ornelas et al. 2013; Ramírez-Barahona and Eguiarte 2013, 2014).

In recent years, efforts also have been made to understand the effect of glacial cycles on the genetic diversity and structure of plant species in Mexico, mainly in temperate and mountain cloud forests (e.g., González-Rodríguez et al. 2004; Gugger et al. 2011; Gutiérrez-Rodríguez et al. 2011; Ornelas et al. 2013; Ruiz-Sanchez and Ornelas 2014; Pérez-Crespo et al. 2017; Rodríguez-Correa et al. 2017). The impact of these climatic fluctuations on the population history of plant species in arid zones has been less studied but was apparently dramatic (Nason et al. 2002; Garrick et al. 2009; Ruiz-Sánchez et al. 2012; Vásquez-Cruz and Sosa 2016; Angulo et al. 2017; Cornejo-Romero et al. 2017; Loera et al. 2017; Scheinvar et al. 2017; Ornelas et al. 2018, 2019). Available evidence indicates that, because of climate cycles, the distribution ranges of xerophytic species expanded and contracted/fragmented recurrently, with concomitant lineage divergence and historical demography changes. Areas where populations of such species survived during climatically adverse periods have been called 'xerophilous refugia'. However, contrasting results have been obtained on whether the maximum range contraction occurred during the periods of glacial maxima or during interglacial periods. The first pattern has been considered consistent with the glacial refugia hypothesis (GRH; Hewitt 2000, 2004), while the second has been named the interglacial refugia hypothesis (IRH; Hewitt 2004; Ornelas et al. 2018). In general, the IRH has been supported by results of ecological niche modeling and phylogeographic analysis for several species in the Chihuahuan and Sonoran deserts, the Tehuacán-Cuicatlán Valley and the Oaxaca Central Valleys, which have shown marked southwards range retractions or shifts during the last interglacial (LIG, ~ 120–140 Ka), followed by northwards expansions during the last glacial maximum (LGM, ~ 22 Ka) and the Holocene. Genetic signatures of these processes include historical demography evidence of population expansions and latitudinal gradients of genetic diversity, with northern populations showing genetic depauperation in comparison to southern populations (Ruiz-Sánchez et al. 2012; Angulo et al. 2017; Scheinvar et al. 2017). However, the opposite pattern (i.e., expansion during the LIG and contraction during the LGM), consistent with the GRH has also been found (Cornejo-Romero et al. 2017). Therefore, extensive research is still needed to better understand the biogeographic history of the arid zones of Mexico, considering that they constitute nearly 50% of the country surface (Challenger 1998) and have a floristic diversity of around 6,000 plant species

(Rzedowski et al. 1993; Cervantes 2005) with a large number of endemisms, representing a set of natural resources that offer multiple alternatives for appropriation of timber, food and other goods (Cervantes 2005).

Mesquites (genus *Prosopis* L.) are among the most important tree groups in arid and semiarid areas across Mexico. The genus probably originated in tropical Africa (Burkart 1976), given the existence of *Prosopis africana* (Guill. & Perr.) Taub. 1893 the most basal of the species in the phylogeny of *Prosopis* (Pohill and Raven 1981). For the Americas, the presence of 43 species of *Prosopis* has been documented, with South America, and particularly Argentina, representing the main diversity center with 34 and 29 species, respectively. In North America, 10 species are known, most of them part of the "Mexico-Texas" complex (Rzedowski 1988; Palacios 2006). In México, *Prosopis laevigata* (Humb. & Bonpl. ex Willd.) M.C.Johnst is the most widespread species, and the "typical" mesquite of the central region of the country (Burkart 1976; Rzedowski 1988). This species represents an excellent study system to analyze the population history of arid-adapted plant species in Mexico, given its wide distribution and high abundance. The geographic distribution of *P. laevigata* includes several physiographic regions in Mexico characterized by arid or semiarid climate, such as the Tamaulipas Plains, the Mexican Altiplano, parts of the Trans-Mexican Volcanic Belt, the Balsas Depression, the Tehuacán-Cuicatlán Valley and the Oaxaca Central Valleys (Rzedowski 1988; Palacios 2006). These physiographic regions are characterized by geological elements that could have acted as barriers or corridors for *P. laevigata* at different historical periods but, at present, the distribution of the species is partially interrupted by areas with temperate vegetation characteristic of the highlands of the Trans-Mexican Volcanic Belt, and by the Sierra Madre Oriental, a mountain chain that separates the populations from the Mexican Altiplano from those of the Tamaulipas Plains. However, there are no morphological peculiarities in the populations associated with the distribution areas, and therefore, they are considered as the same species (Rzedowski 1988).

In this study, our main goal was to analyze the genetic diversity and structure of *P. laevigata* throughout its distribution in Mexico to gain insight into the evolutionary history of the species and contribute to the understanding of the historical biogeography of the Mexican arid zones. In particular, we aimed at i) determining if population differentiation in *P. laevigata* is associated with its distribution in different physiographic regions, and ii) testing if the historical demography and geographic distribution of the species during the LIG, LGM and middle-Holocene (MH) periods conforms to the GRH or the IRH.

Materials and methods

Study system

Prosopis laevigata (Fabaceae, Mimosoideae) is a tree or bush up to 12 m in height, trunk 0.3 to 1 m in diameter, with thick bark of blackish brown color (Calderón and Rzedowski 2001). *Prosopis* species have been characterized as self-compatible (Simpson 1977; Masuelli and Balboa 1989; Genise et al. 1990) with percentages of self-fertilization between 65 and 85% (Galindo-Almanza 1992). The main pollination syndrome of the genus is entomophily, with hymenopterans such as *Ahsmaediella*, *Calicodoma* (Megachillidae), *Coletes* (Colletidae) and *Apis mellifera* (Apidae) being the most frequent pollinators (Galindo-Almanza 1992). Mammalochory and hydrochory have been reported as the main seed dispersal syndromes in *Prosopis* species (Campos and Ojeda 1997; de Noir et al.2002; Pasiecznik et al. 2002).

Sampling methods

We collected fresh leaves of 220 individuals from 21 populations of *P. laevigata* covering the whole distribution of the species (Table 1), including the Oaxaca Central Valleys (OCV), the Tehuacán-Cuicatlán Valley (TCV), the Balsas Depression (BD), the Trans-Mexican Volcanic Belt (TMBV), the Mexican Altiplano (MA) and the Tamaulipas Plains (TP). In each population, seven to 12 adult trees were collected, with a distance of at least 50 m between individuals. The fresh leaves were kept on ice for subsequent storage at -70 °C. Additionally, voucher samples were collected from each of the populations, that were processed for later storage at the IEB Herbarium of the Instituto de Ecología A. C. (Pátzcuaro, Michoacán, México).

Laboratory procedures

Genomic DNA was extracted using a modified CTAB protocol (Otero-Arnaiz et al. 2005). We amplified six polymorphic microsatellite loci (Mo05, Mo07, Mo08, Mo09, Mo13 and Mo016) originally designed for *Prosopis chilensis* (Molina) Stuntz and *Prosopis flexuosa* DC. (Mottura et al. 2005). The PCR reactions were carried out using Platinum

Table 1 Geographic data of the populations collected of *Prosopis laevigata.*

Number	Physiographic region	State	Sample size	Latitude N	Longitude W	Elevation (m a.s.l.)
1	BD	Guerrero	10	17° 55′ 41.45″	99° 35′ 28.25″	515
2	BD	Guerrero	9	18° 03′ 22.1″	100° 49′ 41.85″	1373
3	OCV	Oaxaca	11	18° 06′ 10.59″	97° 20′ 8.50″	1708
4	TCV	Puebla	7	18° 35′ 18.19″	97° 24′ 30.99″	2075
5	TCV	Puebla	8	18° 33′ 41.69″	97° 55′ 29.79″	1778
6	TMBV	Hidalgo	11	20° 34′ 16.65″	98° 36′ 12.54″	2280
7	TMBV	Hidalgo	11	20° 04′ 38.69″	99° 20′ 58.78″	2293
8	TMBV	Querétaro	11	20° 41′ 54.91″	99° 47′ 36.06″	1921
9	TMBV	Michoacán	12	19° 48′ 36.77″	100° 57′ 23″	2249
10	TMBV	Guanajuato	12	20° 24′ 0.18″	101° 34′ 33″	1733
11	TMBV	Jalisco	12	21° 04′ 53.82″	102° 24′ 43.21″	1902
12	TMBV	Michoacán	12	20° 13′ 53.2″	102° 25′ 48.17″	1693
13	TMBV	Jalisco	10	19° 56′ 53.06″	103° 29′ 15.5″	1419
14	TMBV	Aguascalientes	9	21° 53′ 13″	102° 24′ 50.4″	1870
15	MA	San Luis P	11	21° 44′ 56.7″	100° 41′ 47.52″	1852
16	MA	San Luis P	12	22° 37′ 54″	101° 44′ 22.2″	2084
17	MA	San Luis P	11	23° 30′ 23.24″	100° 37′ 37.48″	2168
18	TP	Veracruz	12	22° 09′ 25.62″	98° 04′ 39.06″	32
19	TP	Nuevo León	12	24° 53′ 06.76″	99° 35′ 54.29″	888
20	TP	Tamaulipas	10	23° 44′ 56.32″	99° 02′ 29.58″	900
21	TP	Tamaulipas	11	23° 27′ 10.14″	98° 03′ 0.40″	54

BD Balsas Depression; *OCV* Oaxaca Central Valleys; *TCV* Tehuacán-Cuicatlán Valley; *TMBV* Trans-Mexican Volcanic Belt; *MA* Mexican Altiplano; *TP* Tamaulipas Plains

Master Mix (Thermo Fisher Scientific). A final volume of 10 μL was used, including 5 μL of Platinum Master Mix, 0.5 μL of each primer (10 mM), 2.8 μL of distilled water, 0.2 μL of $MgCl_2$ (2 mM) and 1 μL of DNA (20 ng/μL). The amplification program was carried out as follows: five min at 94 °C as initial step, followed by 35 cycles with a denaturalization step at 94 °C for 90 s, annealing at 49 °C (Mo09), and 59 °C (Mo05, Mo07, Mo08, Mo13 and Mo016) for 90 s, extension at 72 °C for 90 s and a final step at 72° for 10 min. The forward primer of each pair was fluorescently labeled. The PCR products were mixed with formamide and Gen Scan Liz-600 as size standard (Applied Biosystems). The samples were denaturalized at 95 °C for two minutes and then analyzed using an ABI-PRISM 3100-Avant (Applied Biosystems) sequencer. The fragments obtained were sized and scored using the PeakScanner software (Applied Biosystems).

For the phylogeographic analysis, we initially tested seven universal chloroplast DNA (cpDNA) microsatellite loci (Weising and Gardner 1999), and the cpDNA intergenic region trnS-trnT, without finding any variation in these markers. Finally, we tested the psbA3′-trnH region (Shaw et al. 2007), which showed polymorphism. The amplification of this region was carried out in reactions with a final volume of 25 μL, including 12.5 μL of Master Mix (Promega, Woods Hollow Road Madison, WI, USA), 0.8 μL of forward and reverse primer (10 μM), 6 μL of distilled water, 1.5 μL of Bovine Serum Albumin (BSA), 1 μL of $MgCl_2$ (2 mM) and 2.4 μL of DNA (20 ng/μL). The amplification protocol was carried out as follows: five min at 94 °C as initial step, followed by 35 cycles with a denaturalization step at 94 °C for 90 s, an annealing step at 72 °C for 90 s, an extension step at 72 °C for 90 s, and a final step at 72 °C for 10 min. The PCR products were purified using the Qiaquick kit (QIAGEN, Valencia, CA, USA), following the manufacturer's instructions. Between three and nine individuals per population were analyzed. We failed to obtain sequences from populations 11 and 14 due to lack of amplification. The purified products were sent to MACROGEN USA for sequencing. The sequences reported in the present study are available from the GenBank database (accession numbers MK618446-MK618463).

Population genetics analyses

In order to assess the presence of null alleles, we used the software MICROCHECKER (Van Oosterhout et al. 2004). Given the high number of alleles found for the microsatellite markers as well as the relatively small sample size per population, we evaluated the statistical power of our data for detecting genetic differentiation with POWSIM (Ryman and Palm 2006). Different values of effective population size were assayed (500, 2000 and 5000) to evaluate statistical power at F_{ST} values of 0.001, 0.0025, 0.005, 0.01 and 0.02.

Statistical power was expressed as the proportion of significant results for 1,000 replicates. We established a limit of 0.80 as the minimum acceptable power level according to Cohen (1988).

The FreeNA software (Chapuis and Estoup 2007) was used to estimate the null allele frequency for every locus and population based on the expectation–maximization (EM) algorithm. Parameters of genetic diversity, such as observed and expected heterozygosity (H_O, H_E), average number of alleles per locus (A) and effective number of alleles per locus (N_{EA}), for each population were estimated using the software Arlequin 3.5.1.2 (Excoffier and Lischer 2010). Also, the rarefied allelic richness (AR) and heterozygosity corrected by sample size (H_{Ec}, Nei 1978) were evaluated with 20,000 permutations with the software SPAGeDi (Hardy and Vekemans 2002). To identify geographic patterns in genetic diversity levels in *P. laevigata*, genetic diversity parameters were regressed against the latitude, longitude and elevation of the populations with the R software (R Core Team, 2016).

The inferences about the inbreeding level in the populations were based on the Bayesian IIM (Individual Inbreeding Model) approach, implemented in the software INest 2.0 (Chybicki and Burczyk 2009). This software allows the calculation of an unbiased inbreeding coefficient (F_{IS}) for multilocus data in presence of null alleles. The analysis evaluates the effects of null alleles (n), inbreeding (f), and genotyping errors (b) on the homozygosity values implementing the full model (nfb) and its comparison with the alternative models (nf, nb, fb) using the deviance information criterion (DIC) among models.

To evaluate the partitioning of genetic diversity within and between populations, we performed an analysis of molecular variance (AMOVA) with the software GenAlEx 6.5 (Peakall and Smouse 2012). The populations were assigned to the six physiographic regions (TP, MA, TMBV, BD, TCV, OCV) to test if the geological limits among them have constituted barriers to gene dispersal in *P. laevigata*. We used both F_{ST} and R_{ST} estimators based, respectively, on the infinite alleles and stepwise mutation models. Also, we carried out a Mantel test (Mantel 1967), to evaluate the existence of isolation by distance (IBD) among the populations of *P. laevigata* using the genetic distances (pairwise F_{ST} values) and the log geographic distances in GenAlEx 6.5.

The genetic structure of the populations was inferred with the program STRUCTURE 2.3.4 (Pritchard et al. 2000; Hubisz et al. 2009) which, through a model-based Bayesian algorithm, identifies groups of individuals by minimizing Hardy–Weinberg and linkage disequilibria. In this analysis, every individual is assigned by probability to a genetic cluster in order to identify the number of genetic clusters (K) with the maximum value of posterior probability [InP(D)] (Pritchard et al. 2000). Based on preliminary runs with

higher values of K, we tested K values from 1 to 5, with 20 iterations for each K. Every run was done using 50,000 iterations as burn-in and 100,000 repetitions of the Markov chain after burn-in. The options of correlated allele frequencies and possible admixture among individuals in the populations were used. The determination of the most probable K value was executed using the maximum value of ΔK according to Evanno et al. (2005) through the online software Structure Harvester (Earl and vonHoldt 2012) As recommended by Janes et al. (2017), we conducted further analysis to test for substructure within the main genetic groups identified, using similar settings and procedures. Results of multiple runs for each K value were summarized using CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007). Finally, we plotted the CLUMPP results with DISTRUCT 1.1 (Rosenberg 2004).

CpDNA sequence analysis

The electropherograms obtained were analyzed and edited using the software BIOEDIT (Hall 1999). The edited sequences were aligned manually using MEGA 4 (Tamura et al. 2007). The analyses of genetic variation were carried out considering nucleotide substitutions only, while the haplotype level analysis considered both nucleotide substitutions and insertion/deletions (indels). We calculated nucleotide diversity (π) and haplotype diversity (h_S) using the software DNAsp 5.0 (Librado and Rozas 2009) and the rarefied haplotype richness with the software SPAGeDi (Hardy and Vekemans 2002) We built a statistical parsimony network with the TCS software (Clement et al. 2000), through the parsimony algorithm of Templeton et al. (1992) using a connection limit of 95% and considering gaps as fifth state. The network was edited using TCS Beautifier (Múrias Dos Santos et al. 2016).

To evaluate the partitioning of haplotype diversity within and among populations and regions, we carried out an AMOVA in Arlequin 3.5 (Excoffier and Lischer 2010) with 10,000 permutations to determine the significance of the test. Populations were grouped according to the six physiographic regions, as previously described for the microsatellite analyses. The presence of phylogeographic structure in the populations was assessed through the comparison of population differentiation with unordered (G_{ST}) and ordered alleles (N_{ST}). If N_{ST} is significantly higher than G_{ST} it means that genealogically close haplotypes tend to occur together in the same populations (i.e., there is phylogeographic structure) (Pons and Petit 1996). This test was performed in SPAGeDi (Hardy and Vekemans 2002) with 20,000 permutations.

The demographic history of the populations was investigated by calculating Tajima's *D* and Fu's *Fs* (Tajima 1989; Fu 1997) with the software DNAsp 5.0 (Librado and Rozas 2009) with 10,000 permutations. Additionally, we performed a mismatch distribution analysis, which assesses the relative frequency of the number of nucleotide differences among all pairs of haplotypes and compares it to the unimodal distribution expected under a recent demographic expansion (Rogers and Harpending 1998; Excoffier et al. 2009). This analysis was carried out in Arlequin 3.5.1.3 (Excoffier and Lischer 2010). The statistical test of no difference between the observed distribution and the distribution expected under the population expansion model was performed through the comparison of the sum of squares differences (SSD).

As an additional historical demography test, we ran a Bayesian Skyline Plot (BSP; Drummond et al. 2005) by means of the software BEAST 1.8.1 implemented online (http://www.phylo.org), to evaluate the effective population size variation over time. We performed two independent runs of 100 million generations using the substitution model GTR + I based on empirical frequencies, a log-normal relaxed clock model and a Bayesian Skyline tree with constant size using 10 initial groups. The parameters and trees were sampled every 1000 iterations with a burn-in of 10%. The time axis was adjusted using rates of 1.0×10^{-9} and 3.0×10^{-9} substitutions per site per year (s/s/y), encompassing the range reported for chloroplast DNA in many angiosperms (Wolfe et al. 1987; Gutiérrez-Rodríguez et al. 2011; Ruiz-Sánchez and Ornelas 2014) The results of every run were analyzed by the software TRACER 1.5 (Drummond and Rambaut 2007) in order to ensure effective sample sizes (ESS) > 200.

Ecological niche modeling (ENM)

The ecological niche of *P. laevigata* was modeled under present climate conditions using the maximum entropy algorithm implemented in MAXENT 3.3.3e (Phillips et al. 2006). One hundred and twenty-nine records of the presence of *P. laevigata* were used, coming from the MEXU herbarium database, the CONABIO database (available at http://www.conabio.gob.mx/informacion/gis/; accessed September 10, 2020) and the GBIF database (available at https://doi.org/10.15468/00000000; accessed September 10, 2020) and after excluding non-georeferenced data, duplicated records, and non-credible data. For each presence record, associated data for 19 bioclimatic variables were taken from WorldClim 2.1 (Fick and Hijmans 2017) for current climate conditions (1970–2000) with 30 arcsec resolution (available at http://worldclim.org/version2). A variance inflation factor (VIF, Brauner and Shacham 1998) analysis was applied to select the least redundant bioclimatic layers (variables with VIF > 0.8 were included). The variables finally used in the niche model were mean diurnal range, isothermality, temperature annual range, mean temperature of coldest quarter, precipitation seasonality, precipitation of wettest quarter,

precipitation of driest quarter and precipitation of warmest quarter.

The model was implemented using the cross-validation resampling method with 50 replicates. As a threshold-independent method of model validation, we used the receiver operating characteristic (ROC) curve analysis. If the value of the area under the curve (AUC) of the ROC is close to 1 it indicates a good model, while values of 0.5 indicate models that are no better than a random model (Phillips et al. 2006). Also, as an additional threshold-dependent method of model validation, we used the true skill statistic (TSS; Allouche et al. 2006). The ecological niche model (ENM) of the present climate conditions was then projected to past climate scenarios, one for the Last Interglacial (LIG, ~ 120 to 140 Ka) (Otto-Bliesner et al. 2006) and two Last Glacial Maximum (LGM, ~ 22 Ka) models provided by the Paleoclimate Modeling Intercomparison Project Phase II (Braconnot et al. 2007), the Community Climate System Model (CCSM4; Collins et al. 2006) and the Model for Interdisciplinary Research on Climate (MIROC-ESM; Hasumi 2007). The LIG and LGM climate models have a resolution of 2.5 arcmin. Additionally, we used the MIROC-ESM and CCSM4

climate scenarios for the mid-Holocene (~ 6 Ka) with a 30 arcsec resolution.

Results

Genetic structure

According to MICROCHECKER, all microsatellite loci showed null alleles, varying in frequency in the populations from 0.09 (Mo07) to 0.29 (Mo13). The results of the statistical power analysis showed that for the different effective population sizes, the six microsatellite markers have sufficient power to detect genetic differentiation of $F_{ST} = 0.01$ at least, which corresponds to low levels of genetic differentiation and high amounts of connectivity (Online Resource 1).

The results of the analysis with FreeNA showed a proportion of null alleles from 0 to 0.38 depending on the locus and the population; however, null alleles did not affect the estimation of genetic differentiation ($F_{ST} = 0.152$ and 0.155 with and without the ENA correction, respectively). In general, genetic diversity levels were moderately high. We found a total of 96 alleles, summed across the six microsatellite loci

Table 2 Parameters of genetic diversity at six microsatellite loci for 21 populations of *Prosopis laevigata* in Mexico

Population	A	N_{EA}	AR	H_O	H_E	H_{EC}	F
1	3.5	2.32	1.54	0.372	0.514	0.544	0.330*
2	2.33	2.01	1.21	0.148	0.195	0.207	0.297
3	4.83	3.25	1.6	0.386	0.565	0.596	0.367*
4	4.33	3.62	1.63	0.444	0.579	0.63	0.319*
5	2.83	2.46	1.54	0.235	0.482	0.543	0.602*
6	3.33	2.19	1.48	0.318	0.455	0.476	0.344*
7	4.83	2.78	1.61	0.449	0.58	0.612	0.277*
8	4.33	3.19	1.61	0.43	0.064	0.612	0.310*
9	5.33	2.93	1.59	0.427	0.564	0.589	0.285*
10	4.33	2.87	1.6	0.481	0.569	0.598	0.205
11	3.83	2.44	1.47	0.334	0.445	0.468	0.297*
12	4	2.47	1.46	0.23	0.119	0.462	0.515*
13	2.66	1.59	1.32	0.267	0.303	0.319	0.172
14	3.83	3.61	1.55	0.317	0.508	0.547	0.441*
15	4	2.55	1.52	0.436	0.485	0.515	0.167
16	5.16	3.66	1.67	0.42	0.639	0.669	0.385*
17	5.5	4.51	1.66	0.45	0.629	0.659	0.328*
18	4.83	3.61	1.64	0.559	0.611	0.639	0.132
19	6	4.6	1.71	0.551	0.68	0.71	0.233*
20	5.5	3.43	1.66	0.518	0.625	0.66	0.226
21	5.83	3.7	1.65	0.495	0.62	0.65	0.247*
Total	4.33	3.03	1.55	0.393	0.487	0.557	0.323

*Significant deviation from Hardy–Weinberg equilibrium ($P < 0.05$)

A mean number of alleles per locus; N_{EA} mean effective number of alleles per locus; *AR* rarefied allelic richness; H_O observed heterozygosity; H_E expected heterozygosity; H_{EC} expected heterozygosity corrected by sample size (Nei, 1978); *F* fixation index

analyzed. The total number of alleles per locus fluctuated from nine (Mo05) to 27 (Mo16), with an average number of alleles per locus of A = 4.34. At all loci, we found a few alleles in high frequency and many rare alleles in low frequency (alleles with a frequency > 0.10). The rarefied allelic richness showed an average value of AR = 1.68 (Table 2).

The average expected heterozygosity (H_E) was 0.527, ranging from H_E = 0.195 (population 21) to H_E = 0.680 (population 4). The average observed heterozygosity (H_O) was 0.393, varying between H_O = 0.148 (population 21) and H_O = 0.558 (population 17). The percentage of polymorphic loci was 100% in most populations, except for populations 1, 2, 19 (P = 83.33) and 21 (P = 33.33) (Table 2). The regression analysis between genetic diversity estimates and latitude indicated positive and significant relations for N_{EA} (R^2 = 0.45, P = 0.008), A (R^2 = 0.47, P = 0.006), AR (R^2 = 0.20, P = 0.01), H_O (R^2 = 0.38, P = 0.003), and H_{Ec} (R^2 = 0.29, P = 0.01), showing substantial evidence of increased genetic diversity in populations northwards. In addition, we found positive correlations of AR (R^2 = 0.19, P = 0.04) and H_O (R^2 = 0.19, P = 0.04) with longitude, showing an increased value of these parameters in eastern populations (Fig. 1).

The software INest revealed a significant contribution of null alleles to inbreeding values according to the DIC value for the nfb (8214.56) and fb (8286.39) models. The revised estimate of inbreeding (F = 0.177), was considerably lower than the estimate without correcting for null alleles (F = 0.323).

The AMOVA using the F_{ST} estimator revealed a moderate overall genetic structure (F_{ST} = 0.16; P < 0.001). We found slight differences between the six physiographic regions indicating that only 3% (F_{CT} = 0.03; P < 0.001) of the genetic diversity is found among regions and that 13% (F_{SC} = 0.13; P < 0.001) is found among populations within regions, while 84% is within populations. The value of the R_{ST} estimator was higher than the F_{ST} value (0.20; P < 0.001), with 0% (R_{CT} = -0.01; P = 0.983) of the genetic diversity being due to differences among regions, and 20% (R_{SC} = 0.2; P < 0.001) due to differences among populations within regions (Table 3). The Mantel test between log geographic distances and the pairwise F_{ST} values showed a non-significant relationship (r = 0.064; P = 0.257), indicating no isolation by distance among populations.

The ΔK statistic showed K = 2 as the most probable number of genetic groups in the data set (Fig. 2a, b; Online Resource 2a, b). Populations in the Balsas Depression (i. e. 1, 2) were almost completely assigned to the light green cluster (proportion of ancestry = 0.97), whereas populations in the western Trans-Mexican Volcanic Belt (9, 10, 11, 12 and 13) had high proportions (> 0.7) of the dark green cluster. The other populations had mixed proportions of both clusters (Fig. 2a, b).

When testing for substructure within the two main genetic clusters, we found that K = 3 was the most probable number of genetic groups within the dark green group (Fig. 2c, d; Online Resource 2c, d). In this case, populations from the TP (18, 19, 20, 21), the east of the MA (17) and the TCV (4, 5) showed a high frequency of the light blue genetic group and the rest of the populations showed more or less equal proportions of the other genetic groups, except for populations 6 and 11 in which the dark blue genetic group predominated and population 13, which showed a high frequency of the purple genetic group (Fig. 2c, d). The analysis also suggested K = 3 as the optimal number of groups within the light green main genetic cluster (Fig. 2e, f; Online Resource 2e, f). Population 2 from the BD had an almost complete assignment (94%) to the brown genetic group, while the rest of the populations had in general less than 10% assignment to this group. Meanwhile, the yellow genetic group was predominant in the other population from the BD (1) and in population 6, while the red group was frequent in MA (14, 16) and TP populations (21) (Fig. 2f).

Chloroplast DNA and haplotype network

We sequenced 427 bp of the intergenic spacer region psbA3′-trnH in 109 individuals belonging to 19 populations of *P. laevigata* covering the whole distribution. A total of 18 haplotypes were found, which formed a network with a star-like shape, with the most common haplotype (77 individuals) located at the center of the network (Table 4, Fig. 3). This haplotype was found in all the populations sampled. Most of the other haplotypes were separated by one mutational step from the central haplotype, although, there were some haplotypes separated by two to six mutational steps from the central haplotype (Fig. 3).

Most of the derived haplotypes were exclusive to a single population (private haplotypes), except for the haplotypes H2 (found in populations 1, 3, 10, 12, 15 and 17), H3 (found in populations 18, 20 and 21) and H4 (found in populations 18 and 21). Eight populations were monomorphic for the most common haplotype. On the other hand, the populations with the highest number of haplotypes were 1, 12, 20 and 21, located at the southwest, center and northeast of the distribution of the species (Table 4, Fig. 3).

The haplotype diversity (h_S) ranged between 0 and 1 in the populations and was on average moderate (0.4). The nucleotide diversity (π) was in general low, ranging between 0 and 0.00392 with an average of 0.0013 (Table 4). The results of the AMOVA indicated an F_{ST} = 0.09 (P < 0.001), meaning that 9% of the genetic diversity is found among populations, while 91% of the genetic diversity is found within populations (Table 5). Differences among regions were low and non-significant (F_{CT} = 0.043; P = 0.06) signaling a high genetic similarity of *P. laevigata* populations
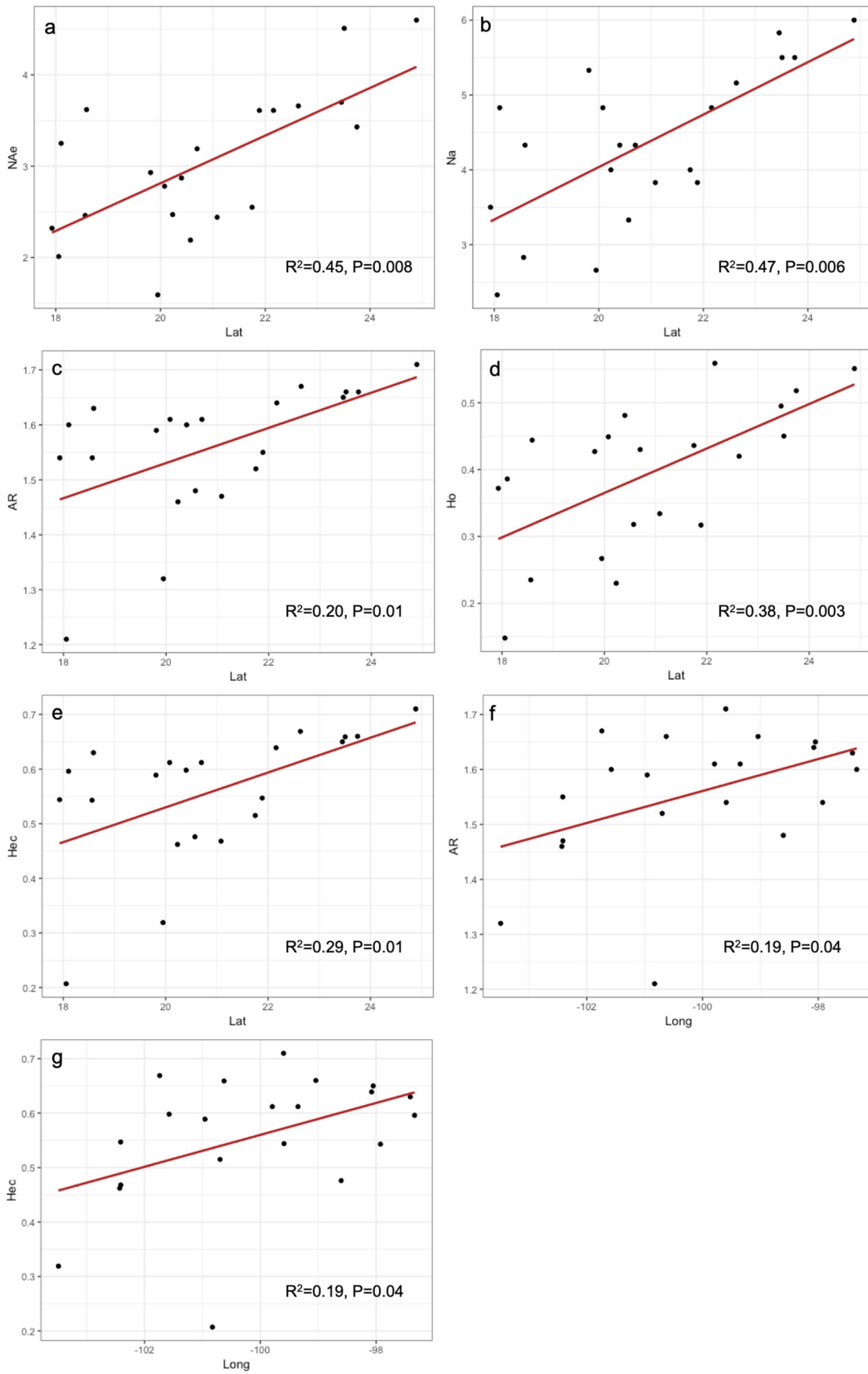
◀**Fig. 1** Regression analysis between genetic diversity parameters of *Prosopis laevigata* populations and geographic variables: **a** mean effective number of alleles (N_{EA}) per locus versus latitude; **b** mean number of alleles per locus (A) versus latitude; **c** rarefied allelic richness (AR) versus latitude; **d** observed heterozygosity (H_O) versus latitude; **e** expected heterozygosity corrected by sample size (H_{Ec}) versus latitude; **f** rarefied allelic richness (AR) versus longitude; **g** expected heterozygosity corrected by sample size (H_{Ec}) versus longitude

along its distribution. The G_{ST} value (0.090) also indicated a low level of genetic structure across the distribution of *P. laevigata*. The differentiation value considering distances among haplotypes (ordered alleles) was $N_{ST} = 0.101$ and was not significantly different from G_{ST} (P = 0.497), denoting a lack of phylogeographic structure in this species.

The values of Tajima's D and Fu's Fs were not significant for individual populations but both overall values were negative and significant (D = -2.018, P < 0.001; Fs = -10.868, P < 0.001), suggesting a recent demographic expansion in *P. laevigata*. The mismatch analysis showed a unimodal distribution, with a main peak located on a single difference between pairs of haplotypes (Online Resource 3). The sum of square deviations (SSD) did not indicate significant differences in comparison to a constant expansion model (SSD = 0.058; P = 0.312), therefore supporting a recent demographical expansion. Likewise, the Bayesian Skyline Plot analysis showed a slight increase in effective population size of the species from 70 Ka to the beginning of the LGM, followed by a more marked expansion from around 20 Ka to the beginning of the Holocene (10 Ka) (Fig. 4).

Ecological Niche model

The potential distribution map based on the ENM of *P. laevigata* for the present-day climate conditions corresponded well to the known distribution of the species (AUC = 0.954 and TSS = 0.825) (Fig. 5a). When the model was projected onto the LIG, we found a substantial reduction of the suitable area for *P. laevigata*, which appeared mainly restricted to the western portion of the Balsas Depression (BD), with scattered small predicted areas in the central BD, the westernmost part of the TMBV and the OCV (Fig. 5b). For the LGM, somewhat different results were obtained with the MIROC-ESM and CCSM4 models (Fig. 5c, d). In the warmer and dryer MIROC-ESM model, a larger area of suitable habitat was predicted in comparison to the LIG, including most of the BD, a large portion of the Sierra Madre del Sur (SMS), the OCV and even the Central Depression of Chiapas; and to the north, some parts of the TMBV, the SMO and the MA (Fig. 5c). In contrast, the CCSM4 model suggested fragmented areas of suitable habitat in the westernmost and the central part of the BD, the OCV, the northwestern Pacific coastal plain and the MA (Fig. 5d). For the

MH, differences between the MIROC-ESM and CCSM4 models were less marked (Fig. 5e, f), both showing suitable habitat for *P. laevigata* in the western and eastern parts of the BD and in the OCV, the TMBV, the MA (in the south eastern and central portions) and the western ridges of the SMO. The whole set of models suggests a refugial area for *P. laevigata* in the west of the BD during the LIG, an expansion to the east and the south in the LGM, and a northwards expansion in the MH and the present.

Discussion

In this study, we used nuclear and chloroplast DNA markers to assess genetic diversity and structure of *Prosopis laevigata* throughout its distribution in Mexico. Despite the ecological and economic importance of *Prosopis* species in the arid zones of North America, this is the first study focusing on the evolutionary history of a species of the genus in this geographic region. Overall, we found moderate levels of genetic diversity in *P. laevigata* ($H_E = 0.527$) which are similar or lower than the values previously reported for a few other *Prosopis* species from South America calculated also from nuclear microsatellites (*Prosopis alba* Griseb, *P. chilensis* and *P. flexuosa*), with H_E values of 0.67, 0.55 and 0.71, respectively (Bessega et al. 2016, 2019; Moncada et al. 2019). Genetic differentiation values are more difficult to compare given the different geographic scales and number of populations analyzed and, unfortunately, analyses of cpDNA diversity based on sequences have not been published yet for any other mesquite species, making comparisons impossible. Nevertheless, some of the shared biological characteristics of *Prosopis* species (pollination by bees, self-compatible, animal dispersed seeds) may explain to some extent the similarities in their patterns of genetic diversity (Duminil et al. 2007).

Interestingly, topographic elements associated with the limits of the physiographic regions, and particularly areas such as the highlands of the TMBV and the SMO that represent interruptions to the continuity of the Mexican arid zones and, therefore, were hypothesized to constitute barriers to gene flow among *P. laevigata* populations, were not associated with any significant genetic breaks in the species. Previously, it has been demonstrated that the TMBV impaired historical gene flow between northern and southern populations of trees such as *Liquidambar styraciflua* L. (Ruiz-Sanchez and Ornelas 2014) and even promoted species diversification in plants and animals (McCormack et al. 2011; Gándara and Sosa 2014). However, the lower elevation parts of the TMBV are probably permeable for some species of drier habitats, providing connectivity between populations to the south and the north.

Table 3 Analysis of molecular variance (AMOVA) using the infinite alleles (IAM) and the stepwise mutation models (SMM) for six microsatellite loci in 21 *Prosopis laevigata* populations from Mexico. The proportions of genetic variation among regions and within and between populations are shown

Mutation model	Source	Degrees of freedom	Sum of Squares	Variance Components	Percentage of variance (%)
IAM (F_{ST})					
	Among Regions	5	59.451	0.067	3
	Among Populations	15	116.748	0.277	13
	Within Populations	427	767.481	1.797	84
	Total	447	943.681	2.144	100
(SMM) R_{ST}					
	Among Regions	5	372,489.101	0.000	0
	Among Populations	15	1,388,609.680	3628.317	20
	Within Populations	427	5,991,846.883	14,032.428	80
	Total	447	7,752,945.665	17,660.746	100

$F_{ST} = 0.16$ $(P < 0.001)$; $F_{SC} = 0.13$ $(P < 0.001)$; $F_{CT} = 0.03$ $(P < 0.001)$; $R_{ST} = 0.20$ $(P < 0.001)$; $R_{SC} = 0.21$ $(P < 0.001)$; $R_{CT} = -0.01$ $(P = 0.983)$

In contrast, results of both the genetic and ENM analysis strongly suggest that the recent evolutionary history of *P. laevigata* was characterized by a significant population expansion event. Current genetic diversity and structure of the species seem to have been determined mostly by a range contraction during the LIG, when suitable habitat for the species possibly reached its minimum extent, and later expansion both northwards and southwards. Such recent demographic expansion in *P. laevigata* is supported by the cpDNA haplotype network, which showed a clear star-like shape with a high-frequency central haplotype which is present in all the populations and several closely related haplotypes at low frequency restricted to single populations. These patterns are typically observed in populations that have recently experienced a fast demographic growth (Hewitt 2004; Xun et al. 2016; Stefenon et al. 2019). All the historical demography analyses, including Tajima's D, Fu's Fs, mismatch distribution and skyline plot also supported this inference. Particularly, the skyline plot suggested that the effective population size of *P. laevigata* was steadily increasing before the LGM, showed a more marked increase between the LGM and the early Holocene, and then stabilized at about 5 Ka. Besides, nuclear microsatellites also showed a high-frequency allele and several rare alleles at most of the loci analyzed, a pattern that also may be explained by a population expansion (Keinan et al. 2007; Hallatschek and Nelson 2008). In terms of genetic structure, moderate levels of differentiation were found for both cpDNA and nuclear markers, as well as a lack of phylogeographic structure in the cpDNA marker. This is also typical of expanding populations because during the range expansion process the ancestral alleles or haplotypes are spread into the newly colonized areas, resulting in low or moderate differentiation values among populations. In contrast, in populations with stable historical population sizes, genetic differentiation depends on the equilibrium between

mutation, gene flow and genetic drift (Excoffier et al. 2009; Hofer et al. 2009).

This inferred range expansion in *P. laevigata* is congruent with paleoclimatic reconstructions for the Neotropics, as well as with previous reports of demographic history for Mexican arid zones species (Ruiz-Sanchez et al. 2012; Angulo et al. 2017; Cornejo-Romero et al. 2017; Loera et al. 2017; Scheinvar et al. 2017; Ornelas et al. 2018) during the glacial/interglacial cycles of the Pleistocene. Available data indicate that during the LIG the climatic conditions were in general warmer and wetter than in the present (Scheinvar et al. 2017; Shadik et al. 2017). Afterward, a trend toward increased aridification has been inferred from various sources of evidence and multiple sites (Caballero et al. 2010; Ortega et al. 2010; Chávez-Lara et al. 2012; Caballero-Rodríguez et al. 2018). For example, Caballero et al. (2010), suggested that during the LGM arid conditions predominated in the central highlands of Mexico while the coasts were more humid. This was probably due to a moisture decline from ~40 Ka to 10 Ka BP resulting from the orographic shadow exerted by the eastern mountains of Mexico which blocked tradewinds from the Gulf of Mexico as a consequence of lower temperatures (Correa-Metrio et al. 2012). This trend apparently continued during the Holocene, since areas of Central Mexico that were wet during the LGM, such as the south of Guanajuato state, became dry by the end of the middle Holocene and into the late Holocene (Domínguez-Vázquez et al. 2019). A similar pattern of climate change probably occurred in northern Mexico according to the evidence provided by fossil packrat middens, which indicate the presence of *Pinus* L. and *Juniperus* L., woodlands at the end of the Pleistocene and the beginning of the Holocene that were rapidly replaced by columnar cacti and mesquite at around 5 Ka (Rhode 2002; Van Devender et al. 2007) suggesting and overall increase in aridity in these regions toward the present. However, these patterns

Fig. 2 Genetic structure in the populations of *Prosopis laevigata*, inferred with the STRUCTURE algorithm (Pritchard et al. 2000; Hubisz et al. 2009) from microsatellite data. **a** bar plot showing the assignment of all sampled individuals into two main genetic groups, **b** map showing the geographic distribution of these two main genetic groups, **c** bar plot showing the genetic assignment corresponding to the substructure within the dark green main genetic group, **d** map showing the geographic distribution of these three genetic subgroups, **e** bar plot showing the genetic assignment corresponding to the substructure within the light green genetic group, **f** map showing the geographic distribution of these three genetic subgroups. The numbering of populations follows Table 1. The gray shading indicates elevation with darker tones corresponding to higher elevations. *SMO* Sierra Madre Oriental; *TP* Tamaulipas plains; *MA* Mexican Altiplano; *TMBV* Trans-Mexican Volcanic Belt; *BD* Balsas Depression; *TCV* Tehuacán-Cuicatlán Valley; *OCV* Oaxaca Central Valleys
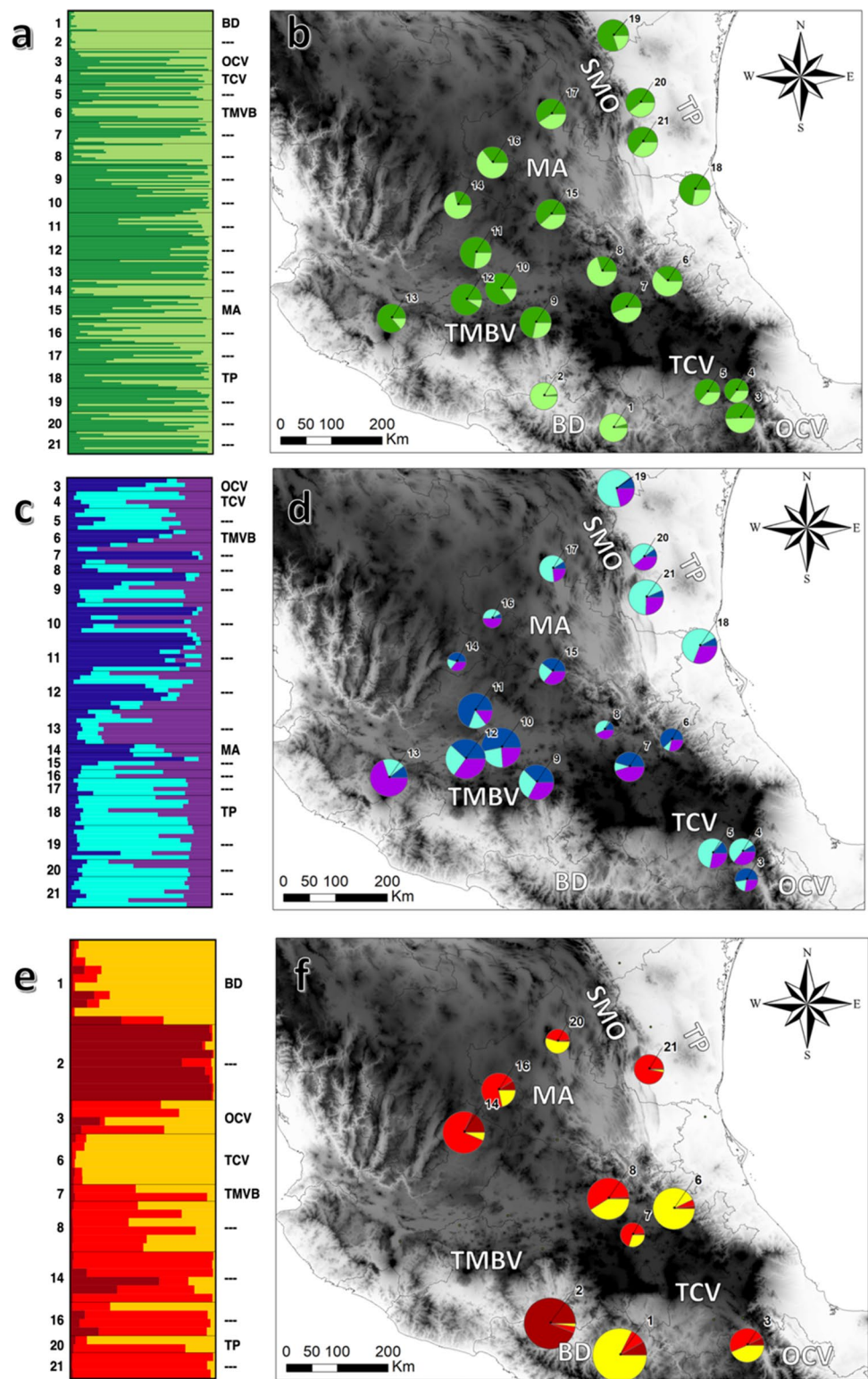
are also thought to have been very variable in space and time during the end of the Pleistocene and the beginning of the Holocene.

In general, our results seem to support the interglacial refugia hypothesis (IRH). It is probable that during the LIG the distribution of *P. laevigata* was most restricted in extent and located mainly in the western part of the BD, with other smaller potential refugial areas in the central BD, the westernmost part of the TMBV and the OCV. These areas acted as sources for population expansion and migration both

Table 4 Genetic diversity estimates for the psbA3′-trnH region of the cpDNA in populations of *Prosopis laevigata* in Mexico

Population	N	h	AR	h_S	π	D	Fs
1	7	4	2.2	0.7143	0.00538	− 1.0354	0.1255
2	4	1	1	0	0	0	0
3	4	3	2.5	0.8333	0.00118	− 0.6123	− 0.2876
4	7	1	1	0	0	0	0
5	7	1	1	0	0	0	0
6	5	1	1.6	0.4	0	0	1.0404
7	3	2	2	0.67	0.00314	0	1
8	5	2	1.6	0.4	0.00094	− 0.816	0.0902
9	5	1	1	0	0	0	0
10	6	2	1.5	0.33	0.00078	− 0.933	− 0.0027
12	4	4	3	1	0.00392	0.167	− 2.1811
13	4	2	1.75	0.5	0.00118	− 0.6123	1.0986
15	6	2	1.9	0.6	0.00141	1.445	0.7951
16	6	1	1	0	0	0	0
17	5	3	2.2	0.7	0.00282	− 1.0484	− 0.1858
18	6	3	2.25	0.7333	0.0022	0.3106	− 0.3041
19	6	1	1	0	0	0	0
20	9	4	2.46	0.8056	0.00249	− 1.1499	− 1.5044
21	9	3	2.31	0.75	0.00183	0.1959	− 1.0387
Total		2.05	1.68	0.4	0.0013	− 2.018***	− 10.868***

***$P < 0.001$

N number of individuals analyzed; *h* number of haplotypes; *AR* rarefied haplotypic richness; h_S haplotype diversity (Pons and Petit, 1996); π nucleotide diversity; *D* Tajima´s D; *Fs* Fu's Fs
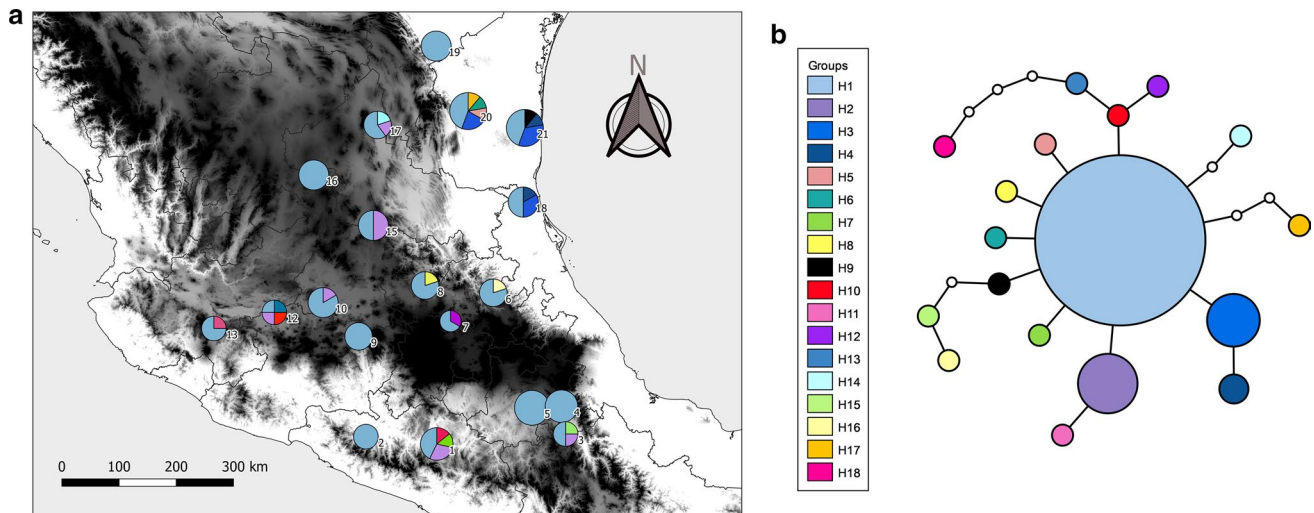
Fig. 3 **a** Geographic distribution of the haplotypes observed in the psbA3′-trnH region of the cpDNA in *Prosopis laevigata*. The numbering of populations follows Table 1. Note that no data are presented for populations 11 and 14 due to amplification problems. The gray shading indicates elevation with darker tones corresponding to higher elevations. **b** Haplotype network obtained with statistical parsimony in TCS (Clement et al. 2000). Each circle indicates a haplotype, with circle size proportional to the frequency of the haplotype. White dots indicate mutational steps

northwards and southwards with the aridification trends that occurred in Mexico toward the LGM and the present, and particularly with the combination of warming and aridification of the Pleistocene/Holocene transition. However, as is frequently observed (e. g. Ornelas et al. 2018; Guevara

2020; Peñaloza-Ramírez et al. 2020), the MIROC-ESM and CCSM4 scenarios differed in the extent of the predicted suitable area for *P. laevigata*, even though the two suggested the same trend. In contrast, the two models for the MH were much more concordant.

Table 5 Analysis of molecular variance for the psbA3′-trnH region of the cpDNA in *Prosopis laevigata* populations in Mexico. The proportions of genetic variation among regions and within and between populations are shown

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of the variance
Among Regions	5	3.979	0.01803	4.36%
Among populations within Regions	13	6.307	0.01914	4.63%
Within Populations	90	33.90	0.3767	91.02%
Total	108	44.19	0.4171	100%

$F_{ST} = 0.089$ ($P < 0.001$); $F_{SC} = 0.048$ ($P = 0.08$); $F_{CT} = 0.043$ ($P = 0.06$)

Interestingly, genetic diversity levels calculated from the nuclear microsatellites were positively correlated with latitude. Usually, areas within the geographic range of a species that have been recently colonized, harbor less genetic diversity in comparison to areas that sustained stable populations over time (i. e. refugia) because of successive founder effects during the process of colonization (Hewitt 2000; Gugger et al. 2008). However, our available evidence favors the scenario in which the central-west portions of the current range of *Prosopis laevigata* acted as refugial areas during the LIG, while southern, eastern and northern areas were more recently colonized. Still, the latitudinal trend in genetic diversity might be explained by current differences in effective population size rather than a historical signature of the colonization process. Indeed, during the collection trips in the north, we could observe very large populations with many thousands of individuals, while the current populations in the BD, the TCV and the OCV are small and scattered, even making it difficult to sample ten individuals in some of them. This pattern also suggests that the 'leading edge' populations of *P. laevigata* may be still expanding while the 'rear edge' populations may be decreasing. Additionally, recent ecophysiological studies of *P. laevigata* in populations from the whole distribution have shown better performance in germination under water deficit, salinity and temperature, as well as photosynthetic performance under drought conditions in northern populations in relation to southern ones (Contreras-Negrete et al. unpublished data), probably leading to more effective regeneration of northern populations.

The range shift and expansion in the distribution of *P. laevigata* during these historical climate transitions are also congruent with the ecology of the species. The seed dispersal of *Prosopis* species is carried out mainly by mammalians as rabbits, hares, humans, and by water courses (Campos and Ojeda 1997; de Noir et al. 2002; Pasiecznik et al. 2002). Endozoochory has been identified as an efficient mechanism in long-distance dispersal, and the seed dispersal of some species of *Prosopis* by the megafauna in the last 12.000 years has been proposed (Janzen and Martin 1982), which could have allowed the rapid expansion of *P. laevigata* populations under favorable ecological conditions during the large-scale warming and aridification process of the Holocene. Mesquites are known to be effective colonizers of open areas, since these plants have high light requirements and show high germination rates and resistance to drought and another stressful conditions typical of arid zones (Zimmermann 1991; Golubov et al. 2001; Pasiecznik et al. 2002).

Fig. 4 Bayesian skyline plot of *Prosopis laevigata* effective population size through time in Mexico, based on psbA3′-trnH cpDNA sequences from 109 individuals. The thick solid blue line is the median estimate, and the area among the light blue lines shows the 95% highest probability density (HPD) (Drummond et al., 2005)
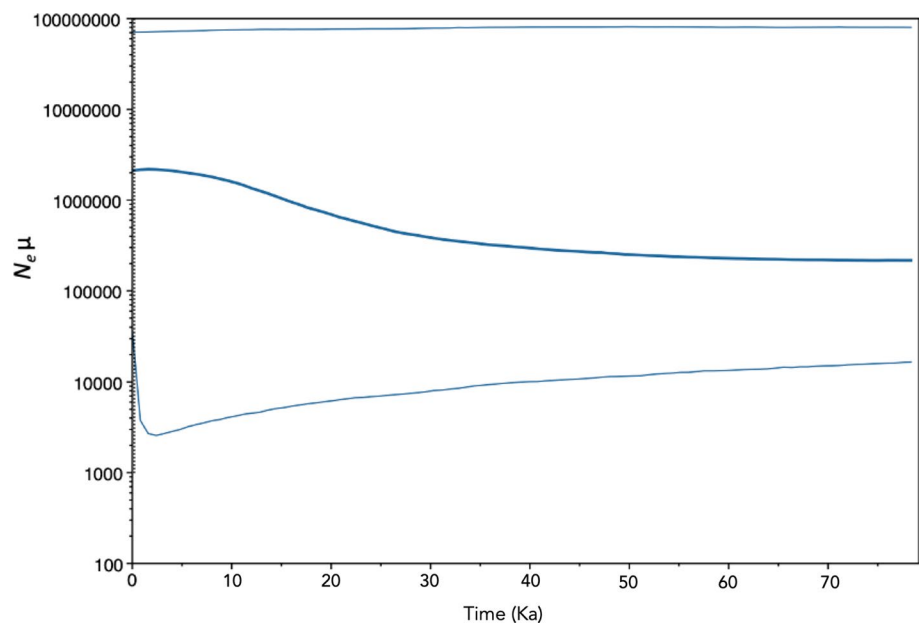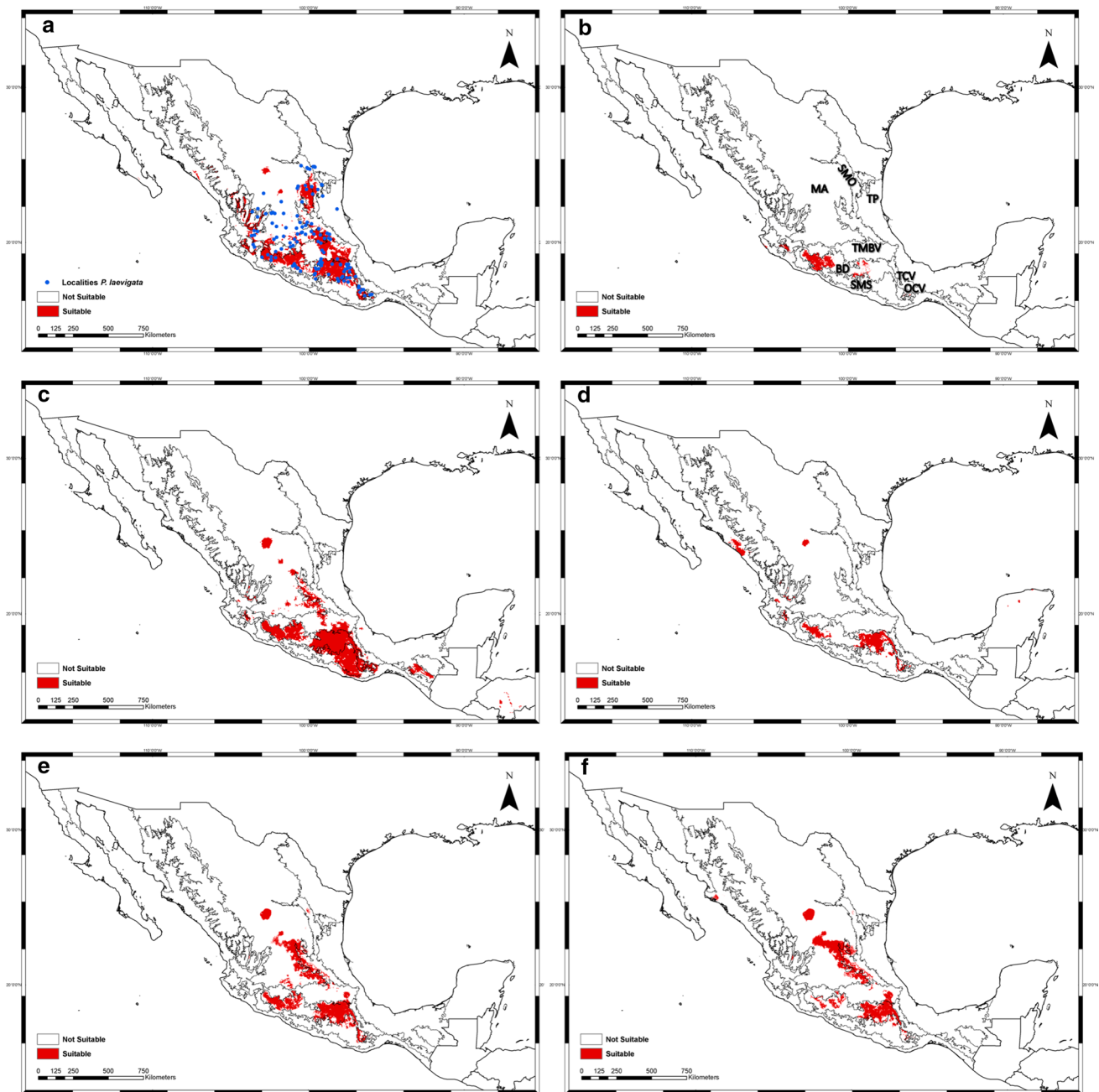
Fig. 5 Ecological niche models for *Prosopis laevigata* in Mexico, **a** contemporary climate conditions, **b** Last interglacial (LIG) climatic conditions, **c** Last Glacial Maximum (LGM) under the MIROC-ESM model, **d** LGM under the CCSM4 model, **e** Mid-Holocene under the MIROC-ESM model, **f** Mid-Holocene under the CCSM4 model.

The red area represents the potential distribution of the species in each period. *SMO* Sierra Madre Oriental; *TP* Tamaulipas Plains; *MA* Mexican Altiplano; *TMBV* Trans-Mexican Volcanic Belt; *BD* Balsas Depression; *SMS* Sierra Madre del Sur; *TCV* Tehuacán-Cuicatlán Valley; *OCV* Oaxaca Central Valleys

These ecological characteristics are responsible for the rapid spread of some *Prosopis* species in contemporary time and some of them have become invasive in arid zones of the Middle East, Africa and Australia (Pasiecznik et al. 2002).

Conclusion

In conclusion, we observe that the recent demographic history, as well as the current patterns of genetic diversity

and structure found in *P. laevigata,* have been decisively shaped by the climatic fluctuations experienced by the species since the LIG, following a pattern consistent with the IRH for dryland species. Likewise, the spread of *P. laevigata* to the north and south of its current distribution since the LIG is consistent with the ecological characteristics of species of the genus, which have been recognized as effective colonizers, even under adverse environmental conditions. Given the widespread distribution of *P. laevigata* as well as its ecological importance, the present study represents relevant information on the ecological and environmental historical patterns that have shaped the biogeographic history of the arid zones of Mexico.

Data archiving statement

Sequence data have been deposited in the Genbank (accession numbers MK618446- MK618463). Microsatellite data and the sequence alignment file have been deposited in the Dryad database (https://doi.org/10.5061/dryad.w3r2280pr).

Information on Electronic Supplementary Material

Online Resource 1. Results of the POWSIM analysis to evaluate the statistical power of our data for six microsatellite loci to detect different values of genetic differentiation in *Prosopis laevigata* populations..
Online Resource 2. Values of ΔK and mean Ln P (D) as a function of K from STRUCTURE analysis for **a**, **b** the main genetic structure, **c**, **d** the genetic substructure with the dark green main genetic group and **e**, **f** the genetic structure within the light green main genetic group.
Online Resource 3. Mismatch distribution analysis. The plot shows the frequency distribution of the observed number of pairwise differences among haplotypes (blue line) compared to the distribution expected under a recent population expansion (red line).

Compliance with ethical standards

References

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43:1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Angulo DF, Amarilla LD, Anton AM, Sosa V (2017) Colonization in North American arid lands: The journey of Agarito (*Berberis trifoliolata*) revealed by multilocus molecular data and packrat midden fossil remains. PLoS ONE 12:1–26. https://doi.org/10.1371/journal.pone.0168933

Bessega C, Cony M, Saidman BO et al (2019) Genetic diversity and differentiation among provenances of *Prosopis flexuosa* DC (Leguminosae) in a progeny trial: Implications for arid land restoration. Forest Ecol Managem 443:59–68. https://doi.org/10.1016/j.foreco.2019.04.016

Bessega C, Pometti CL, Ewens M et al (2016) Fine-scale spatial genetic structure analysis in two Argentine populations of *Prosopis alba* (Mimosoideae) with different levels of ecological disturbance. Eur J Forest Res 135:495–505. https://doi.org/10.1007/s10342-016-0948-9

Braconnot P, Otto-Bliesner B, Harrison S et al (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and last glacial maximum - Part 1: Experiments and large-scale features. Clim Past 3:261–277. https://doi.org/10.5194/cp-3-261-2007

Burkart A (1976) A monograph of the genus *Prosopis* (Leguminosae subfam. Mimosoideae). J Arnold Arbor 57:219–249

Brauner N, Shacham M (1998) Role of Range and Precision of the Independent Variable in Regression of Data. AIChE J 44:603–611. https://doi.org/10.1002/aic.690440311

Caballero-Rodríguez D, Correa-Metrio A, Lozano-García S et al (2018) Late-Quaternary spatiotemporal dynamics of vegetation in Central Mexico. Rev Palaeobot Palynol 250:44–52. https://doi.org/10.1016/j.revpalbo.2017.12.004

Caballero M, Lozano García S, Vázquez Selem L, Ortega B (2010) Evidencias de cambio climático y ambiental en registros glaciales y en cuencas lacustres del centro de México durante el último máximo glacial. Bol Soc Geol Mex 62:359–377. https://doi.org/10.18268/bsgm2010v62n3a4

Calderón G, Rzedowski J (2001) Flora fanerogámica del Valle de México. Instituto de Ecología, A.C. y Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Pátzcuaro

Campos CM, Ojeda RA (1997) Dispersal and germination of *Prosopis flexuosa* (Fabaceae) seeds by desert mammals in Argentina. J Arid Environm 35:707–714. https://doi.org/10.1006/jare.1996.0196

Cervantes M (2005) Plantas de importancia económica en zonas áridas y semiáridas de México. X Encuentro de Geógrafos de América Latina, São Paulo, Brasil, 20 al 25 de marzo de 2005. Invest Geogr 57:3388–3407

Challenger A (1998) La zona ecológica árida y semiárida In: Utilización y conservación de los ecosistemas terrestres de México. Pasado, presente y futuro. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad/Instituto de Biología, UNAM/Agrupación Sierra Madre, México, DF, pp 617–713

Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Molec Biol Evol 24:621–631. https://doi.org/10.1093/molbev/msl191

Chávez-Lara CM, Roy PD, Caballero MM et al (2012) Lacustrine ostracodes from the Chihuahuan desert of Mexico and inferred late quaternary paleoecological conditions. Revista Mex Ci Geol 29:422–431

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. J Heredity 100:106–113. https://doi.org/10.1093/jhered/esn088

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Molec Ecol 9:1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences. Routledge, New York. https://doi.org/10.4324/9780203771587

Collins WD, Bitz CM, Blackmon ML et al (2006) The Community Climate System Model Version 3 (CCSM3). J Clim 19:2122–2143. https://doi.org/10.1175/JCLI3761.1

Cornejo-Romero A, Vargas-Mendoza CF, Aguilar-Martínez GF et al (2017) Alternative glacial-interglacial refugia demographic hypotheses tested on *Cephalocereus columna-trajani* (Cactaceae) in the intertropical Mexican drylands. PLoS ONE 12:1–17. https://doi.org/10.1371/journal.pone.0175905

Correa-Metrio A, Lozano-García S, Xelhuantzi-López S, Sosa-Najera S, Metcalfe SE (2012) Vegetation in western Central Mexico during the last 50 000 years: Modern analogs and climate in the Zacapu Basin. J Quatern Sci 27:509–518. https://doi.org/10.1002/jqs.2540

de Noir, F. Bravo, S. Abdala R (2002) Mecanismos de dispersión de algunas especies de leñosas nativas del Chaco Occidental y Serrano. Quebracho Revista Ci Forest 140–150

Domínguez-Vázquez G, Osuna-Vallejo V, Castro-López V, Israde-Alcántara I, Bischoff J (2019) Changes in vegetation structure during the Pleistocene-Holocene transition in Guanajuato, central Mexico. Veg Hist Archaeobot 28:81–91. https://doi.org/10.1007/s00334-018-0685-8

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:1–8. https://doi.org/10.1186/1471-2148-7-214

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Molec Biol Evol 22:1185–1192. https://doi.org/10.1093/molbev/msi103

Duminil J, Fineschi S, Hampe A, Jordano P, Salvani D, Vendramin GG, Petit RJ (2007) Can population genetic structure be predicted from life-history traits? Amer Naturalist 169:662–672. https://doi.org/10.1086/513490

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet Resources 4:359–361. https://doi.org/10.1007/s12686-011-9548-7

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Molec Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. Annual Rev Ecol Evol Syst 40:481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molec Ecol Resources 10:564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37:4302–4315. https://doi.org/10.1002/joc.5086

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–925

Galindo-Almanza S (1992) Potencial de Hibridacion natural en el Mezquite (*Prosopis laevigata* y *P.glandulosa* Var. *torreyana*, Leguminosae) De La Altiplanicie De San Luis Potosi. Acta Bot Mex 20:101–117. https://doi.org/10.1017/CBO9781107415324.004

Gándara E, Sosa V (2014) Spatio-temporal evolution of *Leucophyllum pringlei* and allies (Scrophulariaceae): A group endemic to North American xeric regions. Molec Phylogen Evol 76:93–101. https://doi.org/10.1016/j.ympev.2014.02.027

Garrick RC, Nason JD, Meadows CA, Dyer RJ (2009) Not just vicariance: Phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja peninsula. Molec Ecol 18:1916–1931. https://doi.org/10.1111/j.1365-294X.2009.04148.x

Genise J, Palacios RA, Hoc PS, Carrizo R, Moffat L, Mom MP, Torregrosa S (1990) Observaciones sobre la biología floral de *Prosopis* (Leguminosae, Mimosoideae). II. Fases florales y visitantes en el Distrito Chaqueño Serrano. Darwiniana 30:71–85.

Golubov J, Mandujano MC, Eguiarte LE (2001) La paradoja de mezquites (*Prosopis* spp.): La invasión de especies o potenciadores de la biodiversidad? Bot Sci 69:23–30. https://doi.org/10.17129/botsci.1644

González-Rodríguez A, Bain JF, Golden JL, Oyama K (2004) Chloroplast DNA variation in the *Quercus affinis-Q. laurina* complex in Mexico: Geographical structure and associations with nuclear and morphological variation. Molec Ecol 13:3467–3476. https://doi.org/10.1111/j.1365-294X.2004.02344.x

Guevara L (2020) Altitudinal, latitudinal, and longitudinal responses of cloud forest species to Quaternary glaciations in the northern Neotropics. Biol J Linn Soc 130:615–625. https://doi.org/10.1093/biolinnean/blaa070

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98

Gugger PF, Gonzalez-Rodriguez A, Rodriguez-Correa H, Sugita S, Cavender-Bares J (2011) Southward Pleistocene migration of Douglas-fir into Mexico: Phylogeography, ecological niche modeling, and conservation of "rear edge" populations. New Phytol 189:1185–1199. https://doi.org/10.1111/j.1469-8137.2010.03559.x

Gugger PF, McLachlan JS, Manos PS, Clark JS (2008) Inferring long-distance dispersal and topographic barriers during postglacial colonization from the genetic structure of red maple (*Acer rubrum* L.) in New England. J Biogeogr 35:1665–1673. https://doi.org/10.1111/j.1365-2699.2008.01915.x

Gutiérrez-Rodríguez C, Ornelas JF, Rodríguez-Gómez F (2011) Chloroplast DNA phylogeography of a distylous shrub (*Palicourea padifolia*, Rubiaceae) reveals past fragmentation and demographic expansion in Mexican cloud forests. Molec Phylogen Evol 61:603–615. https://doi.org/10.1016/j.ympev.2011.08.023

Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. Theor Popul Biol 73:158–170. https://doi.org/10.1016/j.tpb.2007.08.008

Hardy OJ, Vekemans X (2002) spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molec Ecol Notes 2:618–620. https://doi.org/10.1046/j.1471-8286.2002.00305.x

Hewitt G (2000) The genetic legacy of the quaternary ice ages. Nature 405:907–913. https://doi.org/10.1038/35016000

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philos Trans Roy Soc B Biol Sci 359:183–195. https://doi.org/10.1098/rstb.2003.1388

Hofer T, Ray N, Wegmann D, Excoffier L (2009) Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than

by selection. Ann Hum Genet 73:95–108. https://doi.org/10.1111/j.1469-1809.2008.00489.x

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Molec Ecol Resources 9:1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Janes JK, Miller JM, Dupuis JR, Malenfant RM, Gorrell JC, Cullingham CI, Andrew LI (2017) The K = 2 conundrum. Molec Ecol 26:3594–3602. https://doi.org/10.1111/mec.14187

Janzen DH, Martin PS (1982) Neotropical anachronisms: The fruits the gomphotheres ate. Science 215:19–27. https://doi.org/10.1126/science.215.4528.19

Keinan A, Mullikin JC, Patterson N, Reich D (2007) Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. Nat Genet 39:1251–1255. https://doi.org/10.1038/ng2116

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Loera I, Ickert-Bond SM, Sosa V (2017) Pleistocene refugia in the Chihuahuan Desert: the phylogeographic and demographic history of the gymnosperm *Ephedra compacta*. J Biogeogr 44:2706–2716. https://doi.org/10.1111/jbi.13064

Mantel N (1967) The Detection of Disease Clustering and a Generalized Regression Approach. Cancer Res 27:64–71

Masuelli RS, Balboa O (1989) Self-incompatibility in *Prosopis flexuosa*. Pl Cell Incompat Newslett 21:44–48

McCormack JE, Heled J, Delaney KS, Peterson AT, Knowles LL (2011) Calibrating divergence times on species trees versus gene trees: Implications for speciation history of *Aphelocoma* jays. Evolution 65:184–202. https://doi.org/10.1111/j.1558-5646.2010.01097.x

Metcalfe SE (2006) Late Quaternary environments of the northern deserts and central transvolcanic belt of Mexico. Ann Missouri Bot Gard 93:258–273. https://doi.org/10.3417/0026-6493(2006)93[258:LQEOTN]2.0.CO;2

Metcalfe SE, O'Hara SL, Caballero M, Davies SJ (2000) Records of Late Pleistocene-Holocene climatic change in Mexico - A review. Quatern Sci Rev 19:699–721. https://doi.org/10.1016/S0277-3791(99)00022-0

Moncada X, Plaza D, Stoll A, Payacan C, Seelenfreund D, Martínez E, Bertin A, Squeo FA (2019) Genetic diversity and structure of the vulnerable species *Prosopis chilensis* (Molina) Stuntz in the Coquimbo region, Chile. Gayana Bot 76:91–104. https://doi.org/10.4067/S0717-66432019000100091

Mottura MC, Finkeldey R, Verga AR, Gailing O (2005) Development and characterization of microsatellite markers for *Prosopis chilensis* and *Prosopis flexuosa* and cross-species amplification. Molec Ecol Notes 5:487–489. https://doi.org/10.1111/j.1471-8286.2005.00965.x

Múrias Dos Santos A, Cabezas MP, Tavares AI, Xavier R, Branco M (2016) TcsBU: A tool to extend TCS network layout and visualization. Bioinformatics 32:627–628. https://doi.org/10.1093/bioinformatics/btv636

Napier JD, de Lafontaine G, Heath KD, Hu FS (2019) Rethinking long-term vegetation dynamics: multiple glacial refugia and local expansion of a species complex. Ecography 42:1056–1067. https://doi.org/10.1111/ecog.04243

Nason JD, Hamrick JL, Fleming TH (2002) Historical Vicariance and Postglacial Colonization Effects on the Evolution of Genetic Structure in Lophocereus, a Sonoran Desert Columnar Cactus. Evolution 56:2214. https://doi.org/10.1554/0014-3820(2002)056[2214:hvapce]2.0.co;2

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Ornelas JF, García JM, Ortiz-Rodriguez AE, Licona-Vera Y, Gándara E, Molina-Freaner F, Vásquez-Aguilar AA (2019) Tracking host trees: The phylogeography of endemic *Psittacanthus sonorae* (loranthaceae) mistletoe in the Sonoran desert. J Heredity 110:229–246. https://doi.org/10.1093/jhered/esy065

Ornelas JF, Licona-Vera Y, Vásquez-Aguilar AA (2018) Genetic differentiation and fragmentation in response to climate change of the narrow endemic *Psittacanthus auriculatus*. Trop Conserv Sci 11:1–15. https://doi.org/10.1177/1940082918755513

Ornelas JF, Sosa V, Soltis DE, Daza JM, González C, Soltis PS, Gutiérrez-Rodríguez C, Espinosa de los Monteros A, Castoe TA, Bell C, Ruiz-Sanchez E, (2013) Comparative Phylogeographic Analyses Illustrate the Complex Evolutionary History of Threatened Cloud Forests of Northern Mesoamerica. PLoS ONE 8:1–9. https://doi.org/10.1371/journal.pone.0056283

Ortega B, Vázquez G, Caballero M, Israde I, Lozano-García S, Schaaf P, Torres E (2010) Late Pleistocene: Holocene record of environmental changes in Lake Zirahuen, Central Mexico. J Paleolimnol 44:745–760. https://doi.org/10.1007/s10933-010-9449-x

Otero-Arnaiz A, Casas A, Hamrick JL, Cruse-Sanders J (2005) Genetic variation and evolution of *Polaskia chichipe* (Cactaceae) under domestication in the Tehuacán Valley, central Mexico. Molec Ecol 14:1603–1611. https://doi.org/10.1111/j.1365-294X.2005.02494.x

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A (2006) Simulating arctic climate warmth and icefield retreat in the last interglaciation. Science 311:1751–1753. https://doi.org/10.1126/science.1120808

Palacios R (2006) Los Mezquites mexicanos: Biodiversidad y Distribución Geográfica. Bol Soc Argent Bot 41:99–121

Pasiecznik N, Felker P, Harsh HP, LN, Cruz G, Tewari JC, Cadoret K, Maldonado LJ, (2002) The '*Prosopis juliflora*'-'*Prosopis pallida*' complex: a monograph. For Trees Livelihoods 12:229–229

Peakall R, Smouse PE (2012) GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460

Pedersen MW, Overballe-Petersen S, Ermini L, SarkissianC D, Haile J, Hellstrom M, Spens J, Thomsen PF, Bohmann K, Cappellini E, Schnell IB, Wales NA, Carøe C, Campos PF, Schmidt AMZ, Gilbert TP, Hansen AJ, Orlando L, Willerslev E (2015) Ancient and modern environmental DNA. Philos Trans Roy Soc B Biol Sci 370:1–12. https://doi.org/10.1098/rstb.2013.0383

Peñaloza-Ramírez JM, Rodríguez-Correa H, González-Rodríguez A, Rocha-Ramírez V, Oyama K (2020) High genetic diversity and stable Pleistocene distributional ranges in the widespread Mexican red oak *Quercus castanea* Née (1801) (Fagaceae). Ecol Evol 10:4204–4219. https://doi.org/10.1002/ece3.6189

Pérez-Crespo MJ, Ornelas JF, González-Rodríguez A, Ruiz-Sanchez E, Vasquez-Aguilar AA¸ Ramírez-Barahona S, (2017) Phylogeography and population differentiation in the *Psittacanthus calyculatus* (Loranthaceae) mistletoe: a complex scenario of climate–volcanism interaction along the Trans-Mexican Volcanic Belt. J Biogeogr 44:2501–2514. https://doi.org/10.1111/jbi.13070

Phillips SB, Aneja VP, Kang D, Arya SP (2006) Modelling and analysis of the atmospheric nitrogen deposition in North Carolina. Int J Glob Environ Issues 6:231–252. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Polhill RM, Raven PH, Stirton CH (1981) Evolution and systematics of the Leguminosae. Advances Legume Syst 1:1–26

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. Genetics 144:1237–1245

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. https://doi.org/10.1007/s10681-008-9788-0

Ramírez-Barahona S, Eguiarte LE (2013) The role of glacial cycles in promoting genetic diversity in the Neotropics: The case of cloud forests during the Last Glacial Maximum. Ecol Evol 3:725–738. https://doi.org/10.1002/ece3.483

Ramírez-Barahona S, Eguiarte LE (2014) Changes in the distribution of cloud forests during the last glacial predict the patterns of genetic diversity and demographic history of the tree fern *Alsophila firma* (Cyatheaceae). J Biogeogr 41:2396–2407. https://doi.org/10.1111/jbi.12396

Rhode D (2002) Early Holocene juniper woodland and chaparral taxa in the Central Baja California Peninsula, Mexico. Quatern Res 57:102–108. https://doi.org/10.1006/qres.2001.2287

Roberts DR, Hamann A (2015) Glacial refugia and modern genetic diversity of 22 western North American tree species. Proc Roy Soc B Biol Sci. https://doi.org/10.1098/rspb.2014.2903

Rodríguez-Correa H, Oyama K, Quesada M, Fuchs EJ, Quezada M, Ferrufino M, Valencia-Ávalos S, Cascante-Marín A, González-Rodríguez A (2017) Complex phylogeographic patterns indicate Central American origin of two widespread Mesoamerican *Quercus* (Fagaceae) species. Tree Genet Genomes 13:1–14. https://doi.org/10.1007/s11295-017-1147-7

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Molec Biol Evol 9:552–569. https://doi.org/10.1093/oxfordjournals.molbev.a040727

Rosenberg NA (2004) Distruct MANUAL. Molec Ecol Notes 4:137–138

Ruiz-Sanchez E, Ornelas JF (2014) Phylogeography of *Liquidambar styraciflua* (Altingiaceae) in Mesoamerica: Survivors of a Neogene widespread temperate forest (or cloud forest) in North America? Ecol Evol 4:311–328. https://doi.org/10.1002/ece3.938

Ruiz-Sanchez E, Rodriguez-Gomez F, Sosa V (2012) Refugia and geographic barriers of populations of the desert poppy, *Hunnemannia fumariifolia* (Papaveraceae). Organisms Diversity Evol 12:133–143. https://doi.org/10.1007/s13127-012-0089-z

Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Molec Ecol Notes 6:600–602. https://doi.org/10.1111/j.1471-8286.2006.01378.x

Rzedowski J (1988) Análisis de la distribución geográfica del complejo *Prosopis* (Leguminosae, Mimosoideae) en Norteamérica. Acta Bot Mex 3:7. https://doi.org/10.21829/abm3.1988.566

Rzedowski J (1993) Diversity and origins of the phanerogamic flora of Mexico. In: Ramamoorthy TP, Bye R, Lot A, Fa J (eds) Biological Diversity of Mexico: Origins and Distribution. Oxford Univ. Press, New York, pp 129–144

Scheinvar E, Gámez N, Castellanos-Morales G, Aguirre-Planter E, Eguiarte LE (2017) Neogene and Pleistocene history of *Agave lechuguilla* in the Chihuahuan Desert. J Biogeogr 44:322–334. https://doi.org/10.1111/jbi.12851

Shadik CR, Cárdenes-Sandí GM, Correa-Metrio A, Edwards RL, Min A, Bush MB (2017) Glacial and interglacials in the Neotropics: a 130,000-year diatom record from central Panama. J Paleolimnol 58:497–510. https://doi.org/10.1007/s10933-017-0006-8

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. Amer J Bot 94:275–288. https://doi.org/10.3732/ajb.94.3.275

Simpson BB (1977) Mesquite, its biology in two desert scrub ecosystems. US/IBP Synthesis Series. Dowden, Hutchinson & Ross, New York

Stefenon VM, Klabunde G, Lemos RPM, Rogalski M, Nodari RO (2019) Phylogeography of plastid DNA sequences suggests post-glacial southward demographic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. Sci Rep 9:1–13. https://doi.org/10.1038/s41598-019-39308-w

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molec Biol Evol 24:1596–1599. https://doi.org/10.1093/molbev/msm092

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III Cladogram estimation. Genetics 132:619–633

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Molec Ecol Notes 4:535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Vásquez-Cruz M, Sosa V (2016) New insights on the origin of the woody flora of the Chihuahuan Desert: The case of Lindleya. Amer J Bot 103:1694–1707. https://doi.org/10.3732/ajb.1600080

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058. https://doi.org/10.1073/pnas.84.24.9054

Xun H, Li H, Li S, Wei S, Zhang L, Song F, Jiang P, Yang H, Han F, Cai W (2016) Population genetic structure and post-LGM expansion of the plant bug *Nesidiocoris tenuis* (Hemiptera: Miridae) in China. Sci Rep 6:1–11. https://doi.org/10.1038/srep26755

Zimmermann HG (1991) Biological control of mesquite, *Prosopis spp.* (Fabaceae), in South Africa. Agric Ecosyst Environm 37:175–186. https://doi.org/10.1016/0167-8809(91)90145-N