Demographic history and genetic structure in pre-Hispanic Central Mexico

Viridiana Villa-Islas, Alan Izarraras-Gomez, Maximilian Larena, Elizabeth Mejía Perez Campos, Marcela Sandoval-Velasco, Juan Esteban Rodríguez-Rodríguez, Miriam Bravo-Lopez, Barbara Moguel, Rosa Fregel, Ernesto Garfias-Morales, et al.

Demographic history and genetic structure in pre-Hispanic Central Mexico

Viridiana Villa-Islas[1], Alan Izarraras-Gomez[1], Maximilian Larena[2], Elizabeth Mejía Perez Campos[3], Marcela Sandoval-Velasco[4-6], Juan Esteban Rodríguez-Rodríguez[1], Miriam Bravo-Lopez[1], Barbara Moguel[1,7], Rosa Fregel[8], Ernesto Garfias-Morales[1], Jazeps Medina Tretmanis[9], David Alberto Velázquez-Ramírez[10], Alberto Herrera-Muñóz[3], Karla Sandoval[11], Maria A. Nieves-Colón[12,13], Gabriela Zepeda García Moreno[14], Fernando A Villanea[15], Eugenia Fernández Villanueva Medina[16], Ramiro Aguayo-Haro[16], Cristina Valdiosera[17,18], Alexander G. Ioannidis[19], Andrés Moreno-Estrada[12], Flora Jay[20], Emilia Huerta-Sanchez[9], J. Víctor Moreno-Mayar[21], Federico Sánchez-Quinto[22], María C. Ávila-Arcos[1]*.

Affiliations:

[1]International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México (UNAM), Querétaro, México

[2]Department of Organismal Biology, Uppsala University, Uppsala, Sweden

[3]National Institute of Anthropology and History, Querétaro, Mexico

[4]Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark

[5]Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA

[6]Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, CDMX, Mexico

[7]Centro de Geociencias, UNAM Juriquilla

[8]Department of Biochemistry, Microbiology, Cell Biology and Genetics, Universidad de La Laguna, San Cristóbal de La Laguna, Spain

[9]Center for Computational Molecular Biology, Brown University, Providence, RI, USA

[10]Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School, 30625 Hannover, Germany.

[11]Equity and Gender Office of the Centre for Research and Advanced Studies (CODIGO-C), Cinvestav, Mexico City, Mexico

[12]Unit of Advanced Genomics, National Laboratory of Genomics for Biodiversity (LANGEBIO), CINVESTAV, Irapuato, Guanajuato, Mexico

[13]Department of Anthropology, University of Minnesota Twin Cities, USA

[14]National Institute of Anthropology and History, Guanajuato, Mexico

[15]Department of Anthropology, University of Colorado Boulder, USA

[16]National Institute of Anthropology and History, Michoacán, Mexico

[17]Universidad de Burgos, Departamento de Historia, Geografía y Comunicaciones, Burgos, Spain

[18]La Trobe University, Department of History and Archaeology, Melbourne, Australia.

[19]Institute for Computational and Mathematical Engineering, Stanford University, USA

[20]Université Paris-Saclay, CNRS, INRIA, Laboratoire Interdisciplinaire des Sciences du Numérique, 91400, Orsay, France

[21]Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark

[22]Computational Genomics, Instituto Nacional de Medicina Genómica, Ciudad de México, Mexico

*Correspondence to: mavila@liigh.unam.mx

Abstract

Aridoamerica and Mesoamerica are two distinct cultural areas in northern and central Mexico, respectively, that hosted numerous pre-Hispanic civilizations between 2,500 BCE and 1,521 CE. The division between these regions shifted southward due to severe droughts ca. 1,100 years ago, allegedly driving a population replacement in central Mexico by Aridoamerican peoples. Here, we present shotgun genome-wide data from 12 individuals and 27 mitochondrial genomes from eight pre-Hispanic archaeological sites across Mexico, including two at the shifting border of Aridoamerica and Mesoamerica. We find population continuity spanning the climate change episode and a broad preservation of the genetic structure across present-day Mexico for the last 2,300 years. Lastly, we identify a contribution to pre-Hispanic populations of northern and central Mexico from two ancient unsampled 'ghost' populations.

Introduction

Before European colonization, the territory comprised by present-day Mexico was home to numerous civilizations that occupied two main cultural areas: Aridoamerica in northern Mexico, inhabited mainly by hunter-gatherers, and Mesoamerica in central and southern Mexico (Fig. 1), where some of the largest pre-Hispanic agriculture-based civilizations in the Americas flourished between 1,200 BCE and 1,500 CE(*1*, *2*). The distinction between Aridoamerica and Mesoamerica is based on the cultural characteristics and subsistence strategies of the peoples that inhabited them, as well as the ecological features of each region (*3*, *4*). Archaeological evidence indicates that the border between these two areas shifted southward between 900 and 1,300 CE following multi-decadal droughts (*5*). This period is also known as the Medieval Warm Period in other regions of the world (*6*). The droughts allegedly led to population replacements in the northern frontier of Mesoamerica by semi-nomadic hunter-gatherers ("Chichimecas") from Aridoamerica (*7*) and precipitated the fall of some pre-Hispanic societies, and the abandonment of Mesoamerican cities in central and southeast Mexico (*5*, *8–10*).

Evidence for the population replacement at the northern frontier of Mesoamerica comes solely from the archaeological record. Whether this change was the product of migration or acculturation

has been debated by archaeologists for years (*8*, *11–13*). Studying the genetic variation of ancient populations that span this period of climate change across these two regions is thus necessary to illuminate the regional population dynamics in response to this drastic environmental change. However, ancient genomic data for pre-Hispanic populations from Mexico is very limited, with only two studies reporting eleven low-depth genomes (<0.3x) for a handful of individuals restricted to northern Mexico (*14*, *15*), and no available genomes from central and southern Mexico.

Here, we report the most extensive set of complete mitogenomes and shotgun genome-wide data to date from pre-Hispanic individuals from Mexico. We analyze our data jointly with publicly available datasets of other ancient Native Americans and present-day Indigenous populations from Mexico. This large compendium of ancient genomic data allows us to study the pre-Hispanic genetic structure in the territory occupied by Mexico, contributing to longstanding questions regarding population dynamics at the northern frontier of Mesoamerica, disentangling the complex structure of central Mexico populations, and revealing a previously unknown ancient contribution of unsampled genetic lineages to Mexican populations.

Results

Sampling and pre-Hispanic genomic data generation

We screened archeological samples from 37 pre-Hispanic individuals excavated at seven sites from within Mesoamerica in central Mexico. Three are in the Sierra Gorda in Querétaro state (Toluquilla and Ranas sites, and a cave in Cadereyta de Montes), three in Michoacán state (Zaragoza, Tanhuato, and La Mina sites) and one in Guanajuato state (Cañada de la Virgen site) (Fig. 1). Furthermore, we produced additional shotgun data for two mummies from the Sierra Tarahumara from Aridoamerica, northern Mexico who were sequenced at lower depths in a previous study (*16*) (Fig. 1). The sampling of the archeological human remains was made following approval by the Archaeology Council of the Instituto Nacional de Antropología e Historia with permit numbers 401.3S.16-2017/990 and 401.3S.16-2019/222. We sampled the minimum possible amount of tissue to avoid unnecessary destruction. For teeth, we tried to separate the root without damaging the crown and returned all leftover material to the archeologists responsible for each collection. Although consultation with Indigenous populations for destructive analysis of human archeological material is not a requirement by law in Mexico and the studied sites are not located within Indigenous communities, we acknowledge the need to involve Indigenous perspectives in delineating the regulations for these types of studies (see (*17*) for a deeper discussion on the subject). To engage the neighboring communities, we delivered public lectures to communicate the genetic results to the village closest to the sites in Sierra Gorda.

After processing and sequencing samples, we obtained informative amounts of shotgun sequence data (0.01x – 4.7 x genomes) for 12 individuals spanning a time transect of 750 years (600 – 1351 CE). These include ten low-depth genomes from within Mesoamerica in central Mexico, and two from Aridoamerica, in the Sierra Tarahumara (Table S1). In addition, we reconstructed the mitochondrial genomes for 27 individuals across these sites (5.7 – 1284.8x), spanning a time transect of 1,600 years (320 BCE to 1,351 CE). (Table S1 and S2). We included in the dataset three previously published ancient genomes (Table S3) as well as nine ancient mitogenomes from northern Mexico (Table S4), yielding a total of 15 ancient genomes and 36 ancient mitogenomes from the Mexican territory.

We named each sample according to burial and individual identifiers provided by archaeologists and added two additional labels. The first label includes 1-3 letters that refer to the archaeological site, while the second refers to whether the individual dates to before (label 'b') or after (label 'a') the drought period. The three pieces of information are divided by an underscore character '_'. Throughout the text we refer to all individuals using these IDs.

Genetic structure and diversity in uniparental markers

All chromosomal sex assignments matched the morphological sex except one (XX chromosomally and male based on morphology) (Table S1) (*18*). It was possible to assign Y-DNA haplogroups to five out of the thirteen individuals assigned as XY, all of which had a Native American Q lineage. This is in agreement with previous studies on ancient and present-day Indigenous Mexican individuals (*19–24*). At the sub-haplogroup level, we found Q1a2a1-L54 and Q1a2a1a1a1-M3, with no apparent differences between Aridoamerica and Mesoamerica (Table S1).

All 27 reconstructed mitochondrial genomes carried one of the mtDNA haplogroups found in Indigenous populations of the Americas: A (n=10), B (n=9), C (n=4), and D (n=4) (*25*) (Table S1, and S4) (*18*). After merging with nine additional pre-Hispanic mitogenomes from Mexico from other studies (*15*, *16*, *26*), we found all four haplogroups in Mesoamerica (n=25), and only C and B in Aridoamerica (n=11) (Table S4). Consistently, we observed higher nucleotide diversity π values in Mesoamerican populations compared to Aridoamerica (*18*).

The pre-Hispanic mtDNA haplogroup distribution resembles the one in present-day individuals (*24*, *26*, *27*), showing an overall continuity of the matrilineal genetic structure for at least 2,300 years. Of the available complete ancient mtDNA genomes from Aridoamerica, 36% harbor haplogroup C, decreasing in frequency to 8% in Mesoamerica, while the opposite occurs for haplogroup A, with 40% of presence in Mesoamerica and 0% in Aridoamerica (Table S4). This gradient has been previously observed in ancient and contemporary populations from Mexico (*19*, *20*, *24*, *26–32*). Moreover, we detected unreported variants in the mtDNA PhyloTree (*33*) for seven individuals assigned to the sub-haplogroups A2d, B2c, and B2l (Table S5) (*18*).

In Sierra Gorda, we identified sub-haplogroup A2d in six individuals, four pre-drought and two post-drought. After quality filtering (*18*), five of them were included in a median-joining network (Table S4). Four (two pre- and two post- drought) clustered together in a single clade, despite spanning a time transect of 1,480 years, while the fifth individual remained in a different clade within A2d (Fig. 2A). Remarkably, P_CCM_b and 2417J_TOL_a are found in the same node within A2d. This reflects a continuity of this maternal lineage, despite the severe droughts in Sierra Gorda. In contrast, the haplotype network for haplogroup B including five Sierra Gorda individuals, shows that the three pre-drought individuals share sub-haplogroup B2I, while the two post-drought are related and found in a node representing sub-haplogroup B2c. Since mtDNA is subject to genetic drift that leads to lineage loss, it cannot accurately inform about genetic structure or admixture events; the discrepant patterns between haplogroups A and B could be explained by a number of scenarios, which are difficult to test due to the small sample size. This highlights the need to explore autosomal data to test the population replacement hypothesis reliably.

We further leveraged the mtDNA data to estimate the past female effective population size (Ne) using Extended Bayesian Skyline Plots (EBSP) in pre-Hispanic (n=29) and present-day Mexico (n=232) (Table S4 and S6), per haplogroup and all haplogroups together (Fig. S6A-H). All EBSPs runs (Table S7) showed wide (95%) Confidence Intervals (CI) except in the EBSP of only present-day Mexico, where we observed a population expansion ca. 5 kybp and a population decline ca. 500 ybp, though the median values do not support the latter (Fig S6G-H). Despite the wide CI, there is a clear distinction between the estimated female Ne per haplogroup, with haplogroup A reaching the highest present-day values (104,048), followed by D (100,812) and decreasing almost an order of magnitude for haplogroups B and C (37,155 and 7,226, respectively). To improve the estimations and get insights about the past female Ne in a wider geographical range, we merged our dataset with available ancient mtDNA sequences from South America (n=137) (Table S6). The EBSPs for haplogroups A, B, and D, respectively also show wide CI (Fig. S6I-J, N), while haplogroup C shows a clearer pattern of population expansion starting ca. 15 kybp and a population decline starting ca. 5 kybp (Fig. S6K). Consistent with our previous observation, the estimated female Ne is notably higher for haplogroup A and D, than for B and C in the merged dataset (Fig. S6I-L).

Autosomal genetic structure in pre-Hispanic Mexico

We performed ADMIXTURE(*34*) and Principal Component Analysis (PCA) (*35*, *36*) to visualize the genetic relationship and genetic structure at the autosomal level between the pre-Hispanic individuals and other ancient individuals representative from North, Central and South America (Table S3), along with present-day Native American from Mexico and continental populations (Table S8) (*16*, *19*, *23*, *37*, *38*). All pre-Hispanic individuals from Mexico, as well as all ancient individuals from California, Belize and Patagonia clustered together in PCA with present-day Indigenous populations from Mexico (Fig. S7) and share similar genetic composition in the ADMIXTURE analysis (Fig. S8).

We projected the pre-Hispanic individuals onto a principal components space of only pre-Hispanic and present-day Indigenous populations from Mexico (*16*, *19*, *23*, *37*) using a novel approach named "Missing DNA" PCA (mdPCA) that corrects for the high missing fraction of low-depth ancient genomes (see Materials and Methods). All ancient individuals cluster closely in the mdPCA with present-day populations from the same geographical region, except for individual MOM6_ST_a which is intermediate between northern and central present-day populations. Both pre-, and post- drought Sierra Gorda individuals cluster together with present-day Nahua from Jalisco, Purépecha, Totonac, Nahua from Puebla, and Nahua, all from central Mexico (Fig. 3A). This pattern was replicated when using traditional PCA (Fig. S9) and a Temporal Factor Analysis approach (*39*), which corrects the ancestral relationships using the dates of the pre-Hispanic individuals (Fig. S10) (*18*). In agreement, the ADMIXTURE analysis assuming six ancestral components reveals that pre-Hispanic individuals show similar genetic ancestry proportions to those observed in present-day populations from the same region (Fig. 3B) and that individuals from Sierra Gorda show a homogeneous genetic composition independently of the period in which they lived (Fig. 3B).

We then performed several combinations of D- and Outgroup-f3 statistics to further explore the genetic structure of ancient and present-day populations (*18*). Outgroup-f3 statistics were of the form f3(Test, Source1; YRI), where the Test is the pre-Hispanic individual under analysis, and

Source1 is a present-day Indigenous population. D-statistics were of the form D(Pop1, Pop2; Test, YRI), where Test refers to the pre-Hispanic individual under analysis, Pop1 and Pop2 are all possible combinations of present-day Indigenous populations. YRI represents Yoruba as the outgroup population in both tests. We found F9_ST_a to share higher genetic drift with present-day northern Mexico populations than with any other present-day population (Fig. S11) (Table S9) and to be significantly more related to the present-day Rarámuri than to any other population when tested in the form D(Pop1, Rarámuri; F9_ST_a, YRI) (Fig. S12) (Table S10). However, a population continuity test that considers the read counts in the ancient individual and allele frequencies in the present-day population (*40*) rejected the null hypothesis of population continuity between F9_ST_a and present-day Rarámuri a p-value of 10^-499.3 (*18*).

The genetic ancestries of pre-Hispanic individuals from central Mexico are more complex. They do not seem to have a higher shared genetic drift with any present-day Indigenous population according to the f3 outgroup values, where almost all standard errors overlap (Fig. S13-S15) (Table S9). Furthermore, in D-statistics analysis, we find mostly D=0 in all possible combinations of Pop1 and Pop2 (Figs. S16 – S18) (Table S10) as expected in populations with extensive gene flow between them and not completely diverged from each other.

Genetic diversity in pre-Hispanic Mexico

We estimated the conditional nucleotide diversity (CND) (*41*) and runs of homozygosity (ROH) (*42*) in pre-Hispanic and present-day Indigenous individuals from Mexico (*19*, *23*) to make inferences about past genetic diversity and effective population sizes and investigate changes in patterns of genetic variation that may have arisen after the drought period (*18*). Both CND values and ROH distribution show that pre-Hispanic and present-day individuals from northern Mexico (Pericúes and Akimel O'odham) have the lowest CND values and the highest sum of inferred ROH >4 cM compared to other populations of their respective period (Table S11, Table S12, Table S13). Pericú hunter gatherers had the lowest CND values (Fig. S19) (Table S12) but the ROH proportion revealed a larger population size than another ancient hunter gatherer population from Patagonia (Fig. S19) (Table S12) (*18*). These results are in agreement with previous studies reporting lower genetic diversity and longer ROH in hunter-gatherer populations from South America than those observed in other regions of the Americas (*43*).

Pre-Hispanic individuals from Sierra Tarahumara, Toluquilla and Cañada de la Virgen show similar values of CND and sum of inferred ROH segments. In Toluquilla we found that the CND value increases with the date difference between pairs of individuals, probably reflective of the accumulation of new mutations during the 489-680 years of difference between pre- (2417Q_TOL_b) and post- (2417J_TOL_a and 333B_TOL_a) drought individuals. Lastly, the segment size distribution of ROHs suggests that the pre-Hispanic individuals studied here belonged to populations with small effective population (Ne) sizes (2Ne= 1,600-6,400) (Fig. S20) (*18*), which is in agreement with Ne previously calculated for present-day northern and southern populations (*44*).

Genetic continuity before and after the 900 – 1,300 CE droughts in the Sierra Gorda

To formally test the hypothesis of population replacement in the Sierra Gorda from the NFM during the 900 – 1,300 CE droughts, we applied different combinations of outgoup-f3 and D-

statistics. We used individuals from Sierra Gorda from the pre- (TOL_b and R_b) or post- (TOL_a) drought period as the test population in an outgroup-f3 of the form f3(Test, Source; YRI), and Source representing another ancient individual from Mexico or ancient representatives of other regions in the Americas (North, Central and South America) (Table S3).

Under a population replacement scenario, we would expect the Toluquillan individuals who lived after the drought to be more closely related to pre-Hispanic individuals with higher proportions of Aridoamerican genetic ancestry (Pericúes and F9_ST_a) than to Toluquillan individuals who lived before the drought (2417Q_TOL_b). Instead, what we observe is that the pre-Hispanic individuals from Sierra Gorda before and after the climate change episode share higher genetic drift between them than with any other pre-Hispanic individual (Fig. 4A), except for the individual 11R_R_b, who shows higher shared genetic drift with an individual from Cañada de la Virgen (Fig. S21) (Table S14). However, standard errors for 11R_R_b are high, due to its low coverage (Fig. S21). In the case of the two individuals from Sierra Tarahumara, the highest value of genetic drift is between them (Fig. S22). In contrast, outgroup-f3 tests with individuals from Cañada de la Virgen (CdV) as the Test population in f3(CdV, Source; YRI) (Fig. S23), have the highest outgroup-f3 values when a related individual (inferred using a relatedness analysis, (*18*)) from the same site is used as the source, but the Source with the second highest value does not belong to Cañada de la Virgen.

We performed D-statistic tests in the form D(SG_b, SG_a; Pop3, YRI), with the individuals from Sierra Gorda of each period, first without collapsing them per period and then merging the individuals in their respective group, and Pop3 being any pre-Hispanic or present-day individual or population from Mexico. SG_b represents the pre-drought individuals 2417Q_TOL_b and 11R_R_b, while SG_a includes the two post-drought individuals 2417J_TOL_a and 333B_TOL_a. Notably, 2417Q_TOL_b shared more alleles with California or Pericú individuals than the post-drought individuals, but this sharing was not observed when only using transversions (Fig. S24) (Table S15). When merging the individuals per period, none of the comparison under this model deviates significantly from D=0, meaning that SG_b and SG_a form a clade to the exclusion of other individuals and are more closely related to each other than with any other pre-Hispanic or present-day population. The only exception was when Pericúes were used as Pop3, where D>0 with a |Z|= 3.625, indicating that the pre-drought individuals are more closely related to Pericúes (Fig. S25) (Table S16). However, this significant deviation disappears when repeating the analysis using only transversion sites (Fig. 4B) (Table S17).

We then used f-statistic-based admixture graphs (qpGraph) (*45*) to further evaluate the population continuity scenario in Sierra Gorda by testing the fit of models that grouped individuals from before and after the drought period in the same clade. For the model, we used the 11,500-kyo USR1 (*46*) as the surrogate for ancient Beringian, 10,700-kyo Anzick (*47*) individual as the surrogate for ancestral Southern Native American (SNA) (*16*) and the 4,200-kyo Ancient Southwestern Ontario individual (ASO) (*15*) as the surrogate for Northern Native American (NNA) (*16*). In the analysis, we included the pre-Hispanic individuals with the highest coverage 2417Q_TOL_b, 2417J_TOL_a, and 333B_TOL_a from Toluquilla and F9_ST_a as the surrogate for Aridoamerica. The three individuals can in fact be modeled in the same clade separated from F9_ST_a, with no contribution to this (Fig. 4C).

Finally, we used qpWave (*45*) to assess if pre- and post-drought Sierra Gorda individuals can share a demographic history without the need of an additional wave of genetic ancestry from Aridoamerica or other population. Indeed, we find that pre- and post-drought Sierra Gorda

individuals can be modeled in the same clade without the need of an admixture event (p-value >0.12). In contrast, an additional genetic source is needed to reflect genetic variation in F9_ST_a, Pericúes, Akimel O' odham, and Maya (Table S18).

Admixture model for Central Mexico populations

Given the complex relationships observed for ancient individuals in central Mexico, we tested additional demographic models using qpGraph. The base model that fitted the pre-Hispanic populations includes Anzick (*47*) individual as the ancestral SNA (*16*), present-day Athabascan (*16*) as the representative for NNA (*16*), and present-day Indigenous populations in Mexico (*37*) representing the northern (Konkaak), central (Nahua from Puebla), and southern (Triqui) regions. Furthermore, this model involves a split of the Southern Native American into S2A and S2B sources. The pre-Hispanic populations from Sierra Gorda and Cañada de la Virgen show different levels of genetic ancestry from Southern Native American and Northern Native American. We found that the population from Sierra Gorda had a higher percentage of the Southern Native American branch from S2B (64%) than the population from Cañada de la Virgen (37%) who had 62% genetic ancestry shared with the Northern Native American branch, indicating a different demographic history for each of these populations despite its relative geographic proximity (Fig. S26). This model supports previous studies that report several admixture events between the two branches that have given rise to Central and South American populations (*15*). For the Michoacán population, we obtained terminal branches with no support (zeros); thus, we did not make conclusions with this model for the Michoacán pre-Hispanic population.

Ghost population contribution to pre-Hispanic Mexico

Earlier studies report the contribution of a "ghost" genetic ancestry from an unsampled group, designated UpopA, among the present-day Mixe from Mexico (*48*). We tested the presence of this UpopA in the pre-Hispanic individuals, using combinations of admixture graph models. Interestingly, Sierra Tarahumara (represented by F9_ST_a) and Cañada de la Virgen (represented by E8_CdV_b) show a genetic ancestry contribution from a 'ghost' population, at 28% and 17%, respectively (Fig. S27A-B), consistent with the "ghost" UpopA genetic ancestry previously reported in Mixe (Fig. S27C), while models including individuals from the other archaeological sites were rejected. When we modeled the pre-Hispanic individuals and Mixe together we found that Sierra Tarahumara and Mixe share the same ghost genetic ancestry UPopA (with a better Z-score when Sierra Tarahumara receives NA1 genetic ancestry) (Fig. S28), while Cañada de la Virgen requires the contribution of a second "ghost" genetic ancestry from that found in Mixe (models without this second "ghost" genetic ancestry are rejected, $|Z| > 3$), which we name "UPopA2" (Fig. S29). Finally, we included Sierra Tarahumara, Cañada de la Virgen and Mixe and the two "ghost" populations in the same model and confirmed that F9_ST_a and Mixe share the contribution from UpopA1, while Cañada de la Virgen requires the contribution an additional unsampled group UpopA2 (Fig. 5) (*18*).

Discussion

We have generated an ancient genome dataset from Mexico to address long-standing questions about population dynamics between Aridoamerica and Mesoamerica. The data allowed us to i) describe levels of population structure and genetic diversity in Mexico prior to European

colonization, ii) test a previous hypothesis of population replacement in central Mexico following a drastic climate change between 1,200 BCE and 1,500 CE, iii) demonstrate a complex admixture model for central Mexico populations, and iv) detect a contribution from unsampled population A (UpopA) to some northern and central Mexico populations.

Our findings exhibit a geographical structure in pre-Hispanic individuals that differentiates northern and central populations from Mexico, which is consistent with the northwest-southeast cline observed in the genetic structure of present-day Indigenous populations (*37*). The pre-Hispanic individuals cluster in proximity to present-day Indigenous populations from the same geographical area in the mdPCA and exhibit similar ancestral components based on ADMIXTURE analysis, except for individual MOM6_ST_a (Fig. 3 and S9-S10). This reflects an overall conservation of the genetic structure of the populations inhabiting the Mexican territory (*37*) for at least 1,400 years (which is the date of the most ancient individual in the dataset). This is consistent with demographic models based on present-day Indigenous populations (*44*),(*49*) proposing a northern/southern population split between 4,000 and 10,000 years ago followed by multiple waves of admixture events between them (*50*). This geographical structure is also reflected in the maternal lineages (Table S5). The spatial distribution of the haplogroups found, namely A, B, C and D, closely resembles that of present-day Mexico (*32*, *51*).

Although we lacked resolution to make accurate past female population size inferences based solely on the ancient mitogenomes in the EBSP analyses, we were able to recapitulate a population expansion ca. 5kybp when using modern mitogenomes, as it has been previously observed with whole-genome data from present-day Indigenous people from Mexico (*52*). This expansion is likely related to the domestication and propagation of maize cultivation in Mexico(*53*). Interestingly, we did not recover a signal of a population bottleneck ca. 500 ybp after European colonization in the mitogenome data. We speculate that the demographic impact of European colonization on the female population was variable across Mexico, and less drastic than in insular and other isolated regions of the Americas (*54–56*). The recovery of additional ancient mitogenomes from a wider temporal and spatial range will provide resolution to paint a more accurate picture of the female population history.

When focusing on the genetic diversity at the autosomal level, our results revealed that ancient Aridoamericans and the populations at the northern frontier of Mesoamerica have similar levels of conditional nucleotide diversity (CND) between them and those observed in present-day Indigenous populations from northern (Akimel O'odham) and southern Mexico (Mixe and some Mayan) (Fig. S19). However, we caution that we lack genome-wide genetic information of present-day individuals from the exact same site as the pre-Hispanic individuals in central Mexico analyzed here to make a direct temporal comparison. We observed, however, that the lowest CND and the highest total number of segments in ROH were found in the ancient hunter gatherer Pericúes and the present-day Akimel O'odham population, both from northern Mexico. In contrast, pre-Hispanic individuals from Sierra Gorda and Cañada de la Virgen and present-day populations from central, southern, and southeastern Mexico have the lowest total number of segments in ROH (Fig. S20). Altogether, these results are reflective of the lower population sizes maintained in isolated Aridoamerica populations and larger population sizes in pre-Hispanic and present-day populations in Mesoamerica. It is important to note that the pre-Hispanic individuals from this study come from small-sized villages (*57*, *58*) and do not necessarily reflect the demography of other Mesoamerican pre-Hispanic populations from larger and multiethnic pre-Hispanic metropoleis (e.g., Teotihuacán, Tenochtitlán, or Palenque) (*59*, *60*), where higher levels of

diversity and lower ROH would be expected. Ancient DNA studies on individuals from these sites would be needed to reveal the extent of genetic diversity in these sites prior to the population collapse inflicted by European colonization.

To directly test the hypothesis of a population replacement at the northern frontier of Mesoamerica by Aridoamerican hunter-gatherers after the 900-1,300 CE droughts, we studied Sierra Gorda individuals from the time before and after the climate change episode. Outgroup-f3 analyses showed that individuals from Sierra Gorda pre- and post-drought shared higher genetic drift between them than with any other pre-Hispanic individual. Using D statistics, we observed that pre- and post-drought individuals from Sierra Gorda always form a clade to the exclusion of the rest of pre-Hispanic or present-day populations from Mexico. Then, qpGraphs showed that pre- and post-drought individuals from Sierra Gorda can be modeled in the same clade, different from the one formed by F9_ST_a as an Aridoamerican surrogate. Regardless F9_ST_a might not belong to the population that allegedly replaced the population in Sierra Gorda following the drought; we used him safely as a surrogate for northern genetic ancestry since previous studies have estimated that the split time between the northern and southern ancestries occurred ca. 7,200ya (*44*). Therefore, even though F9_ST_a likely did not belong to the population that allegedly replaced Sierra Gorda inhabitants following the drought, we can confidently use this genome as a surrogate for northern genetic ancestry since, by the time of the climate change episode, both ancestries were already well differentiated. Furthermore, with qpWave, we confirmed that pre- and post-drought individuals from Sierra Gorda genetic make-up can be explained within the same genetic history without an additional source of genetic variation. Even if this replacement had happened by an unsampled northern population, we would still expect to see signals that significantly differentiated the post- drought individuals with an input from Aridoamerica, in the qualitative (mdPCA and ADMIXTURE) and the quantitative (f3, Dstats, qpGrap and, qpWave) analyses, which is not the case. Instead, the evidence points to population continuity in Sierra Gorda after the climate change episode.

A possible explanation for a population continuity despite the droughts is that the favorable climatic conditions at the northern Sierra Gorda maintained higher humidity than in other arid sites of the northern frontier of Mesoamerica, such as Cañada de la Virgen. Because agriculture was the main subsistence strategy in Cañada de la Virgen, the droughts forced its inhabitants to migrate to other regions during the heavy droughts, which resulted in an abandonment of the site between 1,000-1,100 CE (*58*). In contrast, the main subsistence strategy in Toluquilla and Ranas was the mining and trade of cinnabar, a valuable mineral of sacred value in pre-Hispanic cultures (*57*, *61*, *62*). We hypothesize that the cinnabar trade and the landscape of the Sierra Gorda allowed the peoples of Toluquilla and Ranas to subsist despite low rainfall conditions during the drought. Notably, we found shorter ROH segments in the pre-drought pre-Hispanic individual than in the post-drought individuals. This could indicate a reduction in the population size after the climate change episode and suggest a possible demographic impact in the same population. Additional assessment of this hypothesis at other sites at the northern frontier of Mesoamerica will shed light into the population migrations and dynamics in a wider geographical range.

We were also interested in obtaining insights into the demography of central Mexico populations since previous attempts of demographic modeling have been hampered by the high genetic heterogeneity observed in these populations (*44*, *52*). Our outgroup-f3 and D statistics results show

that, although pre-Hispanic individuals from Mesoamerica, in central Mexico share higher genetic drift with present-day populations from that region, none of these relationships were statistically significant. Using qpGraph we found that the pre-Hispanic populations from central Mexico, all have different genetic ancestry sharing with the NNA and SNA branches. This was expected given previous studies on ancient genomes from Central and South America reporting multiple admixture events between these two branches after their split ca. 15,000 years ago (*15*, *50*). Together, these observations point to a scenario in which populations from central Mexico have not completely diverged from one another, possibly due to extensive gene flow as expected based on the active commercial exchange between different Mesoamerican populations for centuries (*63*). This interaction was mainly through trade routes and alliances between different nations (*63*), as revealed in present-day Indigenous populations who share identity-by-descent (IBD) segments (*37*, *50*) and through the analysis of STR loci (56). The study of IBD segments in present-day Indigenous populations from Mexico has also evidenced gene flow between Mesoamerican and Aridoamerican populations in pre-Hispanic times (*50*). However, this gene flow occurred less frequently than within Mesoamerica. The individual MOM6_ST_a from Aridoamerica may have been the result of such admixture between Mesoamerican and Aridoamerican ancestors.

Furthermore, the qpGraph admixture models explored for the pre-Hispanic populations showed F9_ST_a and the ancient individuals from Cañada de la Virgen have genetic ancestry from an "ghost" or unsampled population. A contribution from an unsampled population, named 'UpopA' was previously identified in present-day Mixe (*48*) as well as in present-day northern and central Indigenous populations from Mexico (*50*). UpopA was estimated to have diverged ~24,700 years ago from Native Americans (*48*). Interestingly, while Sierra Tarahumara shares the contribution of UpopA with Mixe, a second "ghost" population ('UpopA2') is needed to model Cañada de la Virgen in the admixture model together with Sierra Tarahumara and Mixe (Fig. 5). This observation reveals a complex population history in the Americas during the late Pleistoscene that needs to be further characterized. Additional aDNA studies from the Americas could contribute to the identification of the source of both "ghost" ancestries, seemingly contributing to many present-day Indigenous populations from Mexico.

In summary, we find that the pre-Hispanic population structure from over a thousand years ago can still be observed today. Our work, together with previous studies (*15*, *37*, *44*, *50*, *52*), shows that the demographic events that gave rise to Aridoamerican and Mesoamerican populations are more complex than previously thought. Commercial trade routes may have contributed to increased mobility, facilitating gene flow between different populations within and between various cultural areas. Furthermore, we found genetic continuity in the Sierra Gorda region at the northern frontier of Mesoamerica which suggest the local population stayed in their homeland despite the longstanding droughts that forced other populations to abandon their cities. The identification of a second "ghost" genetic ancestry contribution to some pre-Hispanic central Mexico population reveals a complex past that needs to be characterized through the ethical study of ancient genomes from Mexico. Our study opens the door for further research to address the questions of the unknown genetic past and population dynamics of Mexican pre-Hispanic populations, whose genetic legacy is retained today among Indigenous and admixed populations.

Materials and Methods

Laboratory procedures

DNA extraction and library preparation (before amplification) were performed in the Human Paleogenomics Laboratory, a clean lab facility at the International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México (LIIGH-UNAM). Bones and teeth surfaces were cleaned with a 1% sodium hypochlorite solution, followed by a solution of 75% ethanol. Then, the surface was UV irradiated (256 nm) for 1.5 minutes on each side using a UVP CL-1000 crosslinker. A Dremel tool was used to remove the outer surfaces of bones and teeth. Bones were cut to get an inner sample of around 100-200 mg for DNA extraction. Teeth were cut at the cementoenamel junction, the roots were wrapped in aluminum foil and pulverized using a hammer. Around 100-200 mg of the root was used for DNA extraction. DNA extraction was performed using the methods described in (*64*, *65*), as described in Table S2.

DNA extraction from individual P_CCM_b was performed taking 150 mg of mummified skin. The sample was washed in deionized sterile water and dried. Then, the sample was UV irradiated (256 nm) for 1.5 minutes using a UVP CL-1000 crosslinker. Epithelium was cut in small pieces and incubated at 50ºC for 24 hours in lysis buffer (10mM Tris-HCl, 10mM NaCl, 5mM $CaCl_2$, 2.5 mM EDTA, 1% SDS, 10 mg/ml proteinase K, 10 mg/mL DTT). Followed by centrifugation at 16,100 x g for 5 minutes and recovery of the supernatant. Then, DNA purification was performed according to the protocol in (*65*).

Double-stranded libraries for Illumina sequencing were prepared according to (*66*), using single-indexed adapters with 6-bp barcodes or double-indexed with 7-bp barcodes, depending on the sequencing platform, NextSeq550 or NovaSeq, respectively. Libraries were analyzed with qPCR to determine the optimum number of cycles during the indexing PCR step. Barcoded libraries were sequenced for a first screening on the NextSeq550 equipment from Illumina using a 2x75 run at either LANGEBIO's genomics core facility (National Laboratory of Genomics for Biodiversity, Irapuato, Guanajuato) or INMEGEN genomics facilities (National Institute of Genomic Medicine, Mexico City). Depending on the quality of the libraries (% endogenous and % clonality), some were chosen to be subjected to mitochondrial genome capture or whole-genome capture using Daicel Arbor Biosciences (Ann Arbor, MI, USA) commercial kits (Table S2). Captured libraries were sequenced to assess complexity and yield with the tool preseq (*67*) and sequenced to higher depth using the NextSeq 550 (2x75 cycles) at LANGEBIO's core facility or in the NovaSeq at the Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) with an S1, 2x100 cycles run (Table S2). Sequencing runs for Sierra Tarahumara individuals were carried out at the Sci LifeLab Uppsala Sequencing Center and at the Bustamante Lab in the Department of Genetics at Stanford University (Table S2).

Sequence data processing

Raw reads were processed using Adapterremoval (*68*) for trimming Illumina adapter sequences and collapsing of pairs of reads with an overlap of at least 11 bp (--trimns --trimqualities --qualitybase 33 --minlength 30 --collapse). Collapsed reads with >30bp and quality above 33 were retained for downstream processing.

The retained reads were mapped to the human reference genome b37 (hg19), with the mitochondrial sequence replaced by the revised Cambridge reference sequence (rCRS, NC_012920). Mapping was done with bwa 0.7.13, aln algorithm, seed disabled (-l 500), and keeping reads with mapping quality >25. Clonal duplicates and reads mapping to more than one

place in the genome were removed with SAMtools rmdup (*69*) and eliminating reads with labels 'XT:A:R' and 'XA:Z', respectively. Then, reads were realigned using GATK with RealignerTargetCreator and IndelRealigner (*70*). To authenticate ancient DNA sequence data, mapDamage2 (*71*) was used to assess the damage patterns and distribution of reads length, using default parameters (Fig S1-S4). Alignments to rCRS were separated into new bam files using SAMtools (*69*) to analyze the mitochondrial genome for haplogroup assignment, haplotype network analysis and Extended Bayesian Skyline Plot (EBSP). For the analysis of autosomal variants, base qualities were rescaled with mapDamage2 (*71*). Depth of coverage was estimated using ATLAS (*72*).

mtDNA analysis

The program Schmutzi (*73*) was used to estimate mtDNA contamination and to obtain consensus sequences. The reads aligned to the Cambridge reference sequence (rCRS, NC_012920) were analyzed with the script contDeam.pl with parameters --library double and the default --uselength, and schmutzi.pl was then run with 8 threads. Haplogrep 2 (*74*) was used to assign a mitochondrial haplogroup and quality of assignment for each library using the fasta file with the consensus mitochondrial genome generated by Schmutzi (Table S1).

For the estimation of the mitochondrial genetic diversity, we aligned the 27 mitochondrial genomes reconstructed in this study along with the eight public mitochondrial genomes from Mexico (Table S4), using the BioEdit sequence alignment editor v. 7.0.5.3 (*75*). The haplotype list was obtained with DNA Sequence Polymorphism software (DNAsp) v. 6.12.03x64 (*76*). The sites with gaps were not considered. Then, we estimated the nucleotide diversity Pi through the genetic distance method of Tajima and Nei, with the software Arlequin v. 3.5.1.3 using default parameters (*77*).

Consensus fasta sequences of the ancient mtDNA genomes from this and previous studies were aligned together with previously published modern sequences (*26*, *33*, *78–91*) (Table S4 and S6) using MEGA X (*92*). Multiple sequence alignments of mitochondrial genomes of the same haplogroup (A, B, C or D) were further visually inspected and edited (i.e. removing common indels, as well as to confirm variants and making sure sequences remained at the same length), to assure an accurate alignment. Common hypervariable sites that are phylogenetically uninformative were excluded from the analysis (309, 315, 515-522 AC indels, 3107, 16182-16183, 16193, and 16519) (*33*). mtDNA sequences were excluded from the alignments based on two criteria: 1) > 200 "N's" as observed in fasta files or 2) Visual evaluation of an excess of C > T mutations that cannot be explained by random mutation. Four of the mtDNA sequences (333B_TOL_a, 333O_TOL_b, E8_CdV_b and, BC30) were excluded from EBSP analysis but not from haplotype network analysis (Table S4), as their high number of N's, would have likely resulted in overestimation of the mutation rate and biased inferences in ESBP analyses.

Haplotype median-joining networks were constructed in Popart (*93*) after collapsing aligned full mtDNA sequences into haplotypes with DnaSP v. 6.12.03x64 (*76*). Haplotype networks were constructed with both ancient (from this study and previously published) (n=25) and present-day (n=78) mtDNA of individuals in the Mexican territory (Table S4 and S6). All present-day samples were retrieved from the PhyloTree (version 17) database (*33*, *82*). Only sequences that were

reported as being of the same sub haplogroup or neighboring sub-haplogroup as our ancient samples were chosen for sequence comparison.

Past female effective population sizes were reconstructed using a Bayesian skyline approach (*94*) in BEAST 2 (*95*). The ancient mtDNA sequences, mentioned above were merged with present-day mtDNA sequences collected from published public databases (Table S4 and S6). Alignments were partitioned into five concatenated regions in the following order (Dloop, Coding, rRNA, and tRNA) as proposed by (*81*). HKY+G was found to be the best substitution model for this arrangement of the data using PartitionFinder 2 software (*96*). Substitution rates were estimated following a strict molecular clock model starting from point estimations as in (*97*). Additionally, we used tip calibrations for ancient samples using dating estimates (Table S1). We ran Markov Chain Monte Carlo (MCMC) chains of 100 million steps for each haplogroup alignment with a sampling of parameters every 10,000 generations, discarding the first 10 million steps as burn-in. Extended Bayesian skyline analysis was plotted using SkyViz reported in (*56*). Two independent runs of EBSP were run, showing a similar behavior (see Table S7 for statistics of each run). Figures S6A-N correspond to the first of the two EBSP runs.

Sex assignment

Determination of biological sex was made using the tool reported in (*98*). This approach computes the proportion of reads mapped to the Y chromosome with respect to the reads mapping to the X chromosome (Ry) and Y chromosomes. According to the method, Ry > 0.075 corresponds to XY, while Ry < 0.016 corresponds to XX (Table S1).

Contamination estimation based on the X chromosome

To account for the contamination present in the nuclear DNA, we estimated the contamination with the software hapCon (*99*), which is based on detecting polymorphic sites on the X chromosome of male individuals. The estimation of contamination was performed only in three individuals meeting the inclusion criteria (XY assignment and > 0.02x on the X chromosome). Estimates with hapCon are reported in Table S1 and were lower (<1%) than the ones estimated with the mitochondrial genome (1-2%) (Table S2).

Y-chromosome haplogroup inference

Y chromosome genotype calling and haplogroup assignments were made as reported in (*100*). Genotype calling was performed sampling one random base at each site of the Y chromosome covered at least once, with ANGSD (-dohaplocall 1 -doCounts 1 -r chrY: -minMinor 0 -maxMis 4). Then, haplogroup assignments were made using the phylogenetic tree of Y chromosome single nucleotide polymorphisms constructed from the 1000 Genomes Project Phase 3, as in (*100*). The most derived haplogroup was assigned.

Autosomal reference panels

For analysis at a continental level, we constructed a reference panel with modern populations (Supplementary Table S5) that includes genetic information of: Yoruba (YRI), European genetic ancestry (CEU), and Chinese (CHB) from the 1000 Genomes Project Phase 3 (*38*, *101*), and available genome-wide information for Indigenous populations from Mexico previously published by the Human Genome Diversity Project (HGDP) (*19*, *23*). These genome-wide data were intersected using plink (*102*) with previously published reference panels of genotype information

of Indigenous populations from Mexico (*103*), after masking non-Native American sites as in (*104*) and keeping individuals with > 85% genotype information (mind 0.15) and SNVs with a minor allele frequency of 0.01 (maf 0.01), and with a 90% genotyping rate (geno 0.1) using the software plink (*102*). The final reference panel includes genotype information of 576,409 SNVs from 596 individuals from 4 continental populations (YRI, CEU, CHB, and Indigenous from Mexico) (Table S8).

For the analysis within Mexico, we used the reference panel including only the Indigenous populations from Mexico (without the other three continental populations: YRI, CEU, and CHB) and keeping the individuals with >90% of Native American genetic ancestry. In total, this reference panel included genetic information of 561,327 SNVs from 268 individuals (Table S5).

We also integrated our generated ancient genome-wide data with previously reported low-depth genomes (0.09 – 0.3 x) from three ancient Pericúes from Aridoamerica reported in (*78*, *79*), resulting in a total of 15 pre-Hispanic individuals from Mexico. Our analysis also included ancient genomic data from previous studies for 21 individuals from across the Americas (Table S3)(*46*, *78*, *79*, *105–109*).

Pseudo-haploid calls

Pseudo-haploid calls of the ancient genomic data were made for the positions included in the reference panel. This was done by randomly sampling one read (when more than one read covered the site) at each site present in the reference panel and keeping the observed base at that site if it had a minimum base quality of 30. If only one read covered the site, the observed base was kept if it had a base quality above 30. Then, the calls were turned into homozygous genotypes. Similarly, for the genotype data in the reference panel, one allele was randomly sampled at heterozygous sites and turned into a homozygous genotype. For downstream analyses, we kept ancient individuals with 0.01x genome-wide coverage (Table S1), all having > 4,650 SNVs intersected with the reference panel.

Relatedness inference

Relatedness between ancient individuals from the same site was assessed using READ (*110*). A normalization step with a panel of non-related individuals is required previous to estimating the relationship in the ancient individuals, this step was performed with all Nahuas from the reference panel as they are the biggest group in the reference panel. Then, the normalization value was used to run READ with the ancient individuals. Two individuals from Toluquilla (333C_TOL_a and 333B_TOL_a) were identified as first-degree relatives. Thus, the one with the highest coverage (333B_TOL_a) was used for downstream analyses. Furthermore, individuals E4_CdV_b and E8_CdV_b were identified as second-degree relatives, as well as E7_CdV_b and E19_CdV_b.

ADMIXTURE analysis

For each dataset analyzed with ADMIXTURE, one hundred replicates were run with randomly generated seed values and calculating the cross-validation error with parameter -cv. The run with the best likelihood for each k was plotted using pong software (*111*). A first ADMIXTURE (*34*) analysis was carried out on the genotype reference panel including Indigenous individuals from Mexico and the three continental populations (YRI, CEU, and CHB) and k=2 through k=9 to identify present-day Indigenous individuals with <90% Native American genetic ancestry. We found k=4 to be the k with the lowest mean cv error (Fig. S8A), separating YRI, CEU, CHB, and

Native American genetic ancestry. The run of k=4 with the best likelihood was chosen and the Indigenous individuals with <90% Native American genetic ancestry were removed from the panel. A second ADMIXTURE analysis was carried out with only the present-day Indigenous populations with >90% Native American genetic ancestry, from k=2 to k=8, being k=2 the one with the lowest cv error (Fig. S8B) and separating populations with northern and southern ancestries. Then, a third ADMIXTURE run was carried out including the present-day Indigenous populations and pre-Hispanic individuals (Fig. 3).

Missing DNA PCA

Missing DNA PCA (mdPCA) is a principal component analysis that corrects for genotypes missing due to genetic ancestry masking or degradation in ancient DNA samples. The correction is performed by comparing genetic distances between all samples. The genetic distance is computed as the average number of pairwise differences (pi as defined in (*112*). The manual of the method and the description of the parameters can be found in the mdPCA GitHub repository https://github.com/AI-sandbox/mdPCA. The dataset included 15 ancient samples (12 from this study and 3 Pericúes previously published) and the reference panel including present-day Indigenous individuals from Mexico. Genotyping data in plink format was converted to an unphased VCF format with Plink version 1.90 beta (*102*). The method used this VCF file as an input with the following parameters: each individual was plotted as an average of both parental haplotypes (AVERAGE_PARENTS=True), weights for each individual were used inversely proportional to the number of samples from the corresponding population (IS_WEIGHTED=True), the simple weighted of covariance PCA without any optimization was used (METHOD=1), no genotype was masked based on local genetic ancestry calls (IS_MASKED=False), and only the three first PC's were calculated (NUM_DIMS=3). Individuals with <0.15x whole genome coverage (19,851 SNVs or less) were further down-weighted to 0.1x to avoid having the principal components be selected based on noise variance stemming from these samples. These included BC30, MOM6_ST_a, 11R_R_b, E4_CdV_b, E7_CdV_b, E19_CdV_b, E2_Mich_b, and E4_Mich_b.

Principal component analysis

We performed principal component analyses (PCA) using smartpca from the software eigensoft v6.0.1 (*35*) and projecting the individuals (lsqproject: YES) on the PCs estimated for the modern populations. The PCA (Fig. S7) with only the ancient individuals as well as the present-day Indigenous populations from Mexico, only included present-day individuals masked and with >90% Native American genetic ancestry.

Temporal Factor Analysis

Temporal Factor Analysis (TFA) (*39*) were run with the reference panel that includes only the Indigenous populations from Mexico. The ancient samples included in the TFA consisted exclusively of individuals from Mexican territory from this study with >1x genome-wide coverage: one individual from Sierra Tarahumara (F9_ST_b), and three from Toluquilla (2417Q_TOL_b , 2417J_TOL_a, and 333B_TOL_a). Genotyping data was manipulated with Plink version 1.90 beta (*102*). The dataset was converted from Plink format to Eigenstrat's geno format with the command convertf from Eigensoft version 6.0.1.1 (*35*). Missing genotypes were imputed with the snmf function from the LEA package, using the parameters K=2, entropy= TRUE, and repetitions=5. The runs with the lowest cross-entropy value were considered for

posterior steps. Genotypes were corrected according to the autosomal coverage depth of each sample. We assigned the average depth from the HGDP dataset mentioned in the publication: 35X (*23*) (Table S8). We assigned the respective coverage read depth for all 11 individuals from the Simons Genome Diversity Project (*113*)(Table S8).Modern samples genotyped with microarrays (*103*) did not have a coverage depth value as they represent a different technology without sequencing reads, therefore, we assigned the maximum value from the dataset to all microarray samples, i.e. 56.19X. TFA was applied to these imputed and corrected genotypes with the following parameters: lambda = 5e-1, and K=2. Imputation, coverage correction, TFA, and the manipulation of the genotyping matrix in geno format were performed with R version 4.0.2. Variance explained by lambda did not change considerably across logarithmic values, thus we chose the default lambda value 5e-1. We plotted factor 1 and 2 as negative when required to match geography and facilitate the comparison between TFA plots (Fig. S10).

Outgroup f3 and D statistics

To estimate each pre-Hispanic individual's genetic relatedness with a particular pre-Hispanic or present-day Indigenous population, we performed outgroup f3 statistics using ADMIXTOOLS v5.0 (*45*). Outgroup f3 was calculated in the form (Test, Source1; YRI), where the "Test" is the pre-Hispanic individual under analysis, and Source1 being another pre-Hispanic individual or present-day Indigenous population. We used all YRI from the 1000 Genomes Project (*38*, *101*) as an outgroup population. Assuming there was no admixture in the tree, the f3 value is proportional to the shared genetic drift between Test and Source1.

We performed D statistics using ADMIXTOOLS v5.0 (*45*) to identify whether a pre-Hispanic individual had higher genetic affinities with: i) other pre-Hispanic individuals from the same archaeological site than to pre-Hispanic individuals from other sites; or ii) a specific present-day indigenous population than to other present-day population. D values were computed in the form (Pop1, Pop2; Test, YRI). The "Test" refers to the pre-Hispanic individual under analysis, while Pop1 and Pop2 were all possible combinations between pre-Hispanic individuals and present-day Indigenous populations. Under the null hypothesis, Test individual would be equally related to Pop1 and Pop2, and we expect no significant deviations from D=0. Significant D<0 indicates closer relation between Test individual and Pop2. Moreover, a significant D>0 indicates closer relation between Test individual and Pop1. Only absolute values of Z-score > 3 were considered as significant (which corresponds to a p-value of ~ 0.0027).

Population continuity test between F9_ST_a and present-day Rarámuri

According to the PCA, TFA, ADMIXTURE, outgroup-f3 and D statistics, individual F9_ST_a from north Mexico is significantly closer to present-day Rarámuri. Thus, we analyzed whether individual F9 belonged to the present-day Rarámuri ancestral population using the method published in (*40*). Derived alleles in individual F9 and present-day Rarámuri were identified using ancestral alleles' information from (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/pilot_data/technical/reference/ancestral_alignments). Then, allele counts were performed for each site in the present-day Rarámuri (reference population), excluding alleles with frequency 0 or 1. We performed allele counts for the pre-Hispanic individual F9. Parameters alpha and beta necessary for the analysis were estimated using min_samples = 14, which refers to the minimum number of individuals required in the reference panel that have information for each SNV. Under the null hypothesis, it is assumed that reference

population and ancient individual belong to a continuous population, rejection of the null hypothesis would mean no population continuity.

qpGraph

We used the qpGraph tool of the ADMIXTOOLS v5.0 software package (*45*) to construct admixture graphs and test which various models of demographic history best fit the data. We applied different models according to the analysis: 1)For the model of population continuity, we used the 11,500-kyo USR1 (*46*) as the surrogate for ancient Beringian, 10,700-kyo Anzick (*107*) individual as the surrogate for ancestral Southern Native American (SNA) (*78*) and the 4,200-kyo Ancient Southwestern Ontario individual (ASO) (*79*) as the surrogate for Northern Native American (NNA) (14). 2) For the admixture model for Central Mexico populations, the base model included the pre-Hispanic populations, Anzick (*107*) individual as the ancestral SNA (*78*), present-day Athabascan as the representative for NNA (*78*), and present-day Indigenous populations in Mexico representing the northern (Konkaak), central (Nahua from Puebla), and southern (Triqui) regions. 3) For UpopA we used the different models described in main text and (*18*). pre-Hispanic individuals were added to each model in various combinations to test the fit. The following parameters were applied: outpop (NULL), diag (0.0001), hires (YES), blgsize (0.05), and lsqmode (YES). A worst fitting f statistical test of |Z| score >3 is set as a threshold for model rejection.

qpWave

We used qpWave version 410 (*45*) to test whether pre- and post-drought Sierra Gorda (SG) individuals could be modeled using either SG individuals or other ancient or modern Mexican indigenous population relative to a set of reference groups (Mbuti, Onge, USR1; Papuan, Anzick, Han, Mixe, MA1, Zapotec, Karitiana, Piapoco) as in (*114*). In order to maximize the power for this test we increased the number of SNVs to 2,224,359k transversion sites ascertained in 1KGP and having a 5% MAF. Present-day human sequencing data from HGDP genomes (23) was used to represent Mexican indigenous genetic variation. With these data, we tested whether we could model the genetic variation of pre (2417Q_TOL_b) and post-drought Sierra Gorda (2417J_TOL_a and 333B_TOL_a) individually, using as a source either another Sierra Gorda individual or another Mexican indigenous population. For the latter case we used as references the pre-Hispanic individual F9_ST_a, and Pericúes, along with present-day Akimel O' odham and Mayan. If pre- and post-drought Sierra Gorda individuals arise from different demographic histories to each other, then when analyzed together it would be required to add one or more different sources of genetic variation to explain the data, which would be denoted by a p-value < 0.05. We also performed qpWave analysis with the panel of Indigenous populations from Mexico reported in (104) to increase the number of populations, though at the cost of reducing the number of overlapping sites. Results are consistent with those obtained using the HGDP Akimel O'odham and Mayan in that post-drought genetic variation can be explained with pre-drought genetic variation, suggesting genetic continuity. However, using the Moreno-Estrada reference panel resulted in a low number of overlapping SNVs for the qpWave analysis and, consequently, some p-values could not be computed by the program (e.g. when using the pre and post-drought samples as both target (Pop1) and sources (Pop2) (Table S18) .

Conditional nucleotide diversity

CND, a relative measure of genetic diversity, was estimated by genetic pairwise comparison of the genomic positions that are variable in African individuals (*41*). For these comparisons we grouped

the individuals per archeological site. This analysis included 16 pre-Hispanic individuals (Pericúes, Sierra Tarahuamara, Sierra Gorda, Cañada de la Virgen and Patagonia) with at least 10,000 SNVs overlapping between them (Table S11 and S12) and present-day populations from Mexico (*19*, *23*).

This consists of counting the differences between the ascertained alleles present in one pair of individuals from the same population. For this analysis, we used a panel of 1,938,919 transversions SNVs with a minor allele frequency of 0.01 in the YRI population. In all cases, pseudohaploid genotype calls were performed before the comparisons. CND values were calculated for all pairs of individuals for each present-day Indigenous population and for all pairs of pre-Hispanic individuals within the same site. For present-day populations we masked the sites with no Native American genetic ancestry to avoid counting levels of diversity from a different genetic ancestry. Mean and standard error values were estimated 100 times using a window size of 50 Kb even for pairs of ancient individuals. Results are reported in a violin plot, where each dot inside the violin corresponds to the CND value for a unique pair of individuals (Fig. S19).

Runs of homozygosity

We used the tool hapROH (*42*) to detect runs of homozygosity in pre-Hispanic individuals. The distribution of ROHs was used as an indirect way of measuring genetic diversity and effective population size. This method detects the ROH's lengths longer than four centimorgans(cM) and up to 300 cM. The sum of the ROH detected gives insights into the relatedness of the parents of each individual and informs about population sizes. Length of ROH and relatedness between parents are positively correlated. While the higher the number of ROH segments, the smaller the population size and the smaller the number of heterozygote sites, which means less genetic diversity. hapROH uses a built-in set of 1240K SNVs from to perform pseudohaploid genotype calls and a reference panel of 5,008 present-day human haplotypes from (*54*). We performed a pseudohaploid variant calling for the 1240K SNVs for each pre-Hispanic genome. hapROH's manual recommends using samples with a minimum of 400K SNVs intersected with the 1240K SNVs. However, we also included the individual E8_CdV_b, which intersected 369,504 SNVs, and the Pericú individual B03, which intersected 370,026 SNVs, to have at least one individual of these sites in the analyses and considering the number of SNVs is not too far from 400K (Table S13). Detection of runs of homozygosity was made using the set of 5,008 haplotypes mentioned above.

References:

1. A. L. Austin, L. L. Luján, *El pasado indígena* (Fondo de Cultura Económica, 2019).
2. B. B. Cornejo, *La frontera septentrional de Mesoamérica* (Instituto Nacional de Antropología e Historia, México, ed. 2, 2000), vol. 1 of *Historia Antigua de México*.
3. P. Kirchhoff, Mesoamerica: its geographic limits, ethnic composition and cultural characteristics. *Heritage of conquest*, 17–30 (1952).
4. M. del C. Solanes, E. Vela, Atlas del México prehispánico. *Arqueología Mexicana*. **5**, 5–18 (2000).
5. D. W. Stahle, J. V. Diaz, D. J. Burnette, J. C. Paredes, R. R. Heim, F. K. Fye, *et al.*, Major Mesoamerican droughts of the past millennium: MESOAMERICAN DROUGHTS.

Geophysical Research Letters. **38** (2011) (available at http://doi.wiley.com/10.1029/2010GL046472).
6. M. K. Hughes, H. F. Diaz, Was there a 'medieval warm period', and if so, where and when? *Climatic Change*. **26**, 109–142 (1994).
7. P. Armillas, Condiciones ambientales y movimientos de pueblos en la frontera septentrional de Mesoamérica. *Homenaje a Fernando Márquez-Miranda*, 62–82 (1964).
8. K. H. Wogau, H. W. Arz, H. N. Böhnel, N. R. Nowaczyk, J. Park, High resolution paleoclimate and paleoenvironmental reconstruction in the Northern Mesoamerican Frontier for Prehistory to Historical times. *Quaternary Science Reviews*. **226**, 106001 (2019).
9. D. J. Kennett, S. F. M. Breitenbach, V. V. Aquino, Y. Asmerom, J. Awe, J. U. L. Baldini, *et al.*, Development and Disintegration of Maya Political Systems in Response to Climate Change. *Science*. **338**, 788–791 (2012).
10. J. A. Hoggarth, S. F. M. Breitenbach, B. J. Culleton, C. E. Ebert, M. A. Masson, D. J. Kennett, The political collapse of Chichén Itzá in climatic and cultural context. *Global and Planetary Change*, 18 (2016).
11. S. L. O'Hara, S. E. Metcalfe, F. A. Street-Perrott, On the arid margin: The relationship between climate, humans and the environment. A review of evidence from the highlands of central Mexico. *Chemosphere*. **29**, 965–981 (1994).
12. B. Braniff, Oscilación de la frontera norte mesoamericana: un nuevo ensayo. *Arqueología, segunda época*, 99–114 (1989).
13. A. Gómez Aí-za, S. Sánchez Vázquez, La Frontera Cultural Meso-Aridoamericana: Construcción De Imaginarios Nacionalistas En La Historia Mexicana. *Xih*. **9** (2014), doi:10.37646/xihmai.v9i18.244.
14. M. Raghavan, M. Steinrucken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, *et al.*, Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*. **349**, aab3884–aab3884 (2015).
15. C. L. Scheib, H. Li, T. Desai, V. Link, C. Kendall, G. Dewar, *et al.*, Ancient human parallel lineages within North America contributed to a coastal expansion. *Science*. **360**, 1024–1027 (2018).
16. M. Raghavan, M. Steinrucken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, *et al.*, Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*. **349**, aab3884–aab3884 (2015).
17. M. C. Ávila-Arcos, C. de la Fuente Castro, M. A. Nieves-Colón, M. Raghavan, Recommendations for Sustainable Ancient DNA Research in the Global South: Voices From a New Generation of Paleogenomicists. *Front. Genet.* **13**, 880170 (2022).
18. Materials and methods are available as supplementary materials on Science Online.
19. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, *et al.*, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. **538**, 201–206 (2016).
20. I. De La Cruz, A. González-Oliver, B. M. Kemp, J. A. Román, D. G. Smith, A. Torre-Blanco, Sex Identification of Children Sacrificed to the Ancient Aztec Rain Gods in Tlatelolco. *Current Anthropology*. **49**, 519–526 (2008).
21. D. Perez-Benedico, J. La Salvia, Z. Zeng, G. A. Herrera, R. Garcia-Bertrand, R. J. Herrera, Mayans: a Y chromosome perspective. *Eur J Hum Genet*. **24**, 1352–1358 (2016).
22. K. Sandoval, A. Moreno-Estrada, I. Mendizabal, P. A. Underhill, M. Lopez-Valenzuela, R. Peñaloza-Espinosa, *et al.*, Y-chromosome diversity in Native Mexicans reveals continental

transition of genetic structure in the Americas. *Am. J. Phys. Anthropol.* **148**, 395–405 (2012).
23. A. Bergström, S. A. McCarthy, R. Hui, M. A. Almarri, Q. Ayub, P. Danecek, *et al.*, Insights into human genetic variation and population history from 929 diverse genomes. *Science*. **367**, eaay5012 (2020).
24. B. Z. González-Sobrino, A. P. Pintado-Cortina, L. Sebastián, F. Morales-Mandujano, A. V. Contreras, J. Chávez-Benavides, *et al.*, Genetic Diversity and Differentiation in Urban and Indigenous Populations of Mexico: Patterns of Mit, 21.
25. H.-J. Bandelt, C. Herrnstadt, Y.-G. Yao, Q.-P. Kong, T. Kivisild, C. Rengo, *et al.*, Identification of Native American Founder mtDNAs Through the Analysis of Complete mtDNA Sequences: Some Caveats. *Ann Human Genet*. **67**, 512–524 (2003).
26. A. Y. Morales-Arce, M. H. Snow, J. H. Kelley, M. Anne Katzenberg, Ancient mitochondrial DNA and ancestry of Paquimé inhabitants, Casas Grandes (A.D. 1200–1450). *American Journal of Physical Anthropology*. **163**, 616–626 (2017).
27. J. Mata-Míguez, L. Overholtzer, E. Rodríguez-Alegría, B. M. Kemp, D. A. Bolnick, The genetic impact of aztec imperialism: Ancient mitochondrial DNA evidence from Xaltocan, Mexico. *American Journal of Physical Anthropology*. **149**, 504–516 (2012).
28. A. Y. Morales-Arce, G. McCafferty, J. Hand, N. Schmill, K. McGrath, C. Speller, Ancient mitochondrial DNA and population dynamics in postclassic Central Mexico: Tlatelolco (ad 1325–1520) and Cholula (ad 900–1350). *Archaeological and Anthropological Sciences* (2019), doi:10.1007/s12520-018-00771-7.
29. A. González-Oliver, L. Márquez-Morfín, J. C. Jiménez, A. Torre-Blanco, Founding Amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo: mtDNA Lineages in Ancient Maya. *American Journal of Physical Anthropology*. **116**, 230–235 (2001).
30. A. Y. Morales-Arce, C. A. Hofman, A. T. Duggan, A. K. Benfer, M. A. Katzenberg, G. McCafferty, *et al.*, Successful reconstruction of whole mitochondrial genomes from ancient Central America and Mexico. *Scientific Reports*. **7**, 18100 (2017).
31. B. M. Kemp, A. Gonzalez-Oliver, R. S. Malhi, C. Monroe, K. B. Schroeder, J. McDonough, *et al.*, Evaluating the Farming/Language Dispersal Hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proceedings of the National Academy of Sciences*. **107**, 6759–6764 (2010).
32. M. Bodner, U. A. Perego, J. E. Gomez, R. M. Cerda-Flores, N. Rambaldi Migliore, S. R. Woodward, *et al.*, The Mitochondrial DNA Landscape of Modern Mexico. *Genes*. **12**, 1453 (2021).
33. M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*. **30**, E386–E394 (2009).
34. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. **19**, 1655–1664 (2009).
35. N. Patterson, A. L. Price, D. Reich, Population Structure and Eigenanalysis. *PLoS Genetics*. **2**, e190 (2006).
36. D. Reich, A. L. Price, N. Patterson, Principal component analysis of genetic data. *Nature Genetics*. **40**, 491–492 (2008).
37. A. Moreno-Estrada, C. R. Gignoux, J. C. Fernandez-Lopez, F. Zakharia, M. Sikora, A. V. Contreras, *et al.*, The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science*. **344**, 1280–1285 (2014).
38. The 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature*. **526**, 68–74 (2015).

39. O. François, F. Jay, Factor analysis of ancient population genomic samples. *Nat Commun*. **11**, 4661 (2020).
40. J. G. Schraiber, Assessing the Relationship of Ancient and Modern Populations. *Genetics*. **208**, 383–398 (2018).
41. P. Skoglund, H. Malmstrom, A. Omrak, M. Raghavan, C. Valdiosera, T. Gunther, *et al.*, Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science*. **344**, 747–750 (2014).
42. H. Ringbauer, J. Novembre, M. Steinrücken, Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat Commun*. **12**, 5425 (2021).
43. D. I. Cruz Dávalos, Y. O. Arizmendi Cárdenas, M. J. Bravo-Lopez, S. Neuenschwander, S. Reis, M. Q. R. Bastos, *et al.*, "Indigenous peoples in eastern Brazil: insights from 19 th century genomes and metagenomes" (preprint, Genomics, 2022), , doi:10.1101/2022.01.27.477466.
44. M. C. Ávila-Arcos, K. F. McManus, K. Sandoval, J. E. Rodríguez-Rodríguez, V. Villa-Islas, A. R. Martin, *et al.*, Population History and Gene Divergence in Native Mexicans Inferred from 76 Human Exomes. *Molecular Biology and Evolution*. **37**, 994–1006 (2020).
45. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, *et al.*, Ancient Admixture in Human History. *Genetics*. **192**, 1065–1093 (2012).
46. J. V. Moreno-Mayar, B. A. Potter, L. Vinner, M. Steinrücken, S. Rasmussen, J. Terhorst, *et al.*, Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature*. **553**, 203–207 (2018).
47. M. Rasmussen, S. L. Anzick, M. R. Waters, P. Skoglund, M. DeGiorgio, T. W. Stafford, *et al.*, The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*. **506**, 225–229 (2014).
48. J. V. Moreno-Mayar, L. Vinner, P. de Barros Damgaard, C. de la Fuente, J. Chan, J. P. Spence, *et al.*, Early human dispersals within the Americas. *Science*. **362**, eaav2621 (2018).
49. H. García-Ortiz, et al., The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas. *Nat Commun*. **12**, 5942 (2021).
50. H. García-Ortiz, The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas, 12.
51. A. Gorostiza, V. Acunha-Alonzo, L. Regalado-Liu, S. Tirado, J. Granados, D. Sámano, *et al.*, Reconstructing the History of Mesoamerican Populations through the Study of the Mitochondrial DNA Control Region. *PLoS ONE*. **7**, e44666 (2012).
52. S. Romero-Hidalgo, Demographic history and biologically relevant genetic variation of Native Mexicans inferred from whole-genome sequencing. *NATURE COMMUNICATIONS*, 8.
53. J. Ramos-Madrigal, B. D. Smith, J. V. Moreno-Mayar, S. Gopalakrishnan, J. Ross-Ibarra, M. T. P. Gilbert, *et al.*, Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. *Current Biology*. **26**, 3195–3201 (2016).
54. M. A. Nieves-Colón, W. J. Pestle, A. W. Reynolds, B. Llamas, C. de la Fuente, K. Fowler, *et al.*, Ancient DNA Reconstructs the Genetic Legacies of Precontact Puerto Rico Communities. *Molecular Biology and Evolution*. **37**, 611–626 (2020).
55. B. Llamas, L. Fehren-Schmitz, G. Valverde, J. Soubrier, S. Mallick, N. Rohland, *et al.*, Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*. **2**, e1501385 (2016).
56. B. D. O'Fallon, L. Fehren-Schmitz, Native Americans experienced a strong population bottleneck coincident with European contact. *Proceedings of the National Academy of Sciences*. **108**, 20444–20448 (2011).

57. A. J. Herrera Muñoz, E. Mejía Campos, Un minero en la Sierra Gorda: caso de contaminación ocupacional multielemental de metales pesados a finales del periodo Clásico. *Arqueología (México, DF)*, 127–135 (2017).
58. E. Nalda, La arqueología de Guanajuato: trabajos recientes. *Arqueología mexicana*. **16**, 36–43 (2008).
59. M. E. Smith, Mesoamerica's First World City: Teotihuacan in Comparative Context. *Teotihuacan: The World Beyond the City* (2020).
60. B. A. Álvarez-Sandoval, L. R. Manzanilla, M. González-Ruiz, A. Malgosa, R. Montiel, Genetic Evidence Supports the Multiethnic Character of Teopancazco, a Neighborhood Center of Teotihuacan, Mexico (AD 200-600). *PLoS ONE*. **10**, e0132371 (2015).
61. E. Campos, J. Campos, A. Muñoz, Mercurialism Determination in Fetuses Bone Remains from Toluquilla, Queretaro, Mexico. *JSRR*. **8**, 1–10 (2015).
62. E. M. Pérez Campos, A. J. Herrera Muñoz, MINAS Y MINEROS: PRESENCIA DE METALES EN SEDIMENTOS Y RESTOS HUMANOS AL SUR DE LA SIERRA GORDA DE QUERÉTARO EN MÉXICO. *Chungará (Arica)*. **45**, 161–176 (2013).
63. D. A. Corona, Expansión Territorial Comercial en Mesoamérica y Mesoamérica Septentrional por medio del Intercambio. **19**, 10 (2015).
64. N. Rohland, M. Hofreiter, Ancient DNA extraction from bones and teeth. *Nat Protoc*. **2**, 1756–1762 (2007).
65. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, *et al.*, Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*. **110**, 15758–15763 (2013).
66. M. Meyer, M. Kircher, Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*. **2010** (2010) (available at http://www.cshprotocols.org/cgi/doi/10.1101/pdb.prot5448).
67. T. Daley, A. D. Smith, Predicting the molecular complexity of sequencing libraries. *Nature Methods*. **10**, 325–327 (2013).
68. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Research Notes*. **9**, 88 (2016).
69. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
70. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, *et al.*, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. **20**, 1297–1303 (2010).
71. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. **29**, 1682–1684 (2013).
72. V. Link, A. Kousathanas, K. Veeramah, C. Sell, A. Scheu, D. Wegmann, "ATLAS: Analysis Tools for Low-depth and Ancient Samples" (preprint, Bioinformatics, 2017), , doi:10.1101/105346.
73. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology*. **16**, 224 (2015).
74. H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt, *et al.*, HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res*. **44**, W58–W63 (2016).

75. T. A. Hall, in *Nucleic acids symposium series* ([London]: Information Retrieval Ltd., c1979-c2000., 1999), vol. 41, pp. 95–98.
76. J. Rozas, A. Ferrer-Mata, J. C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado, S. E. Ramos-Onsins, *et al.*, DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*. **34**, 3299–3302 (2017).
77. L. Excoffier, H. E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*. **10**, 564–567 (2010).
78. M. Raghavan, et. al, Genomic evidence for the Pleistocene and recent population history of Native Americans. **349**, aab3884–aab3884 (2015).
79. C. L. Scheib, et al., Ancient human parallel lineages within North America contributed to a coastal expansion. *Science*. **360**, 1024–1027 (2018).
80. S. Mallick, et. al, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. **538**, 201–206 (2016).
81. B. Llamas, et al., Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances*. **2**, e1501385 (2016).
82. M. Van Oven, PhyloTree Build 17: Growing the human mitochondrial DNA tree. *Forensic Science International: Genetics Supplement Series*. **5**, e392–e394 (2015).
83. S. Kumar, C. Bellis, M. Zlojutro, P. E. Melton, J. Blangero, J. E. Curran, Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. *BMC evolutionary biology*. **11**, 1–17 (2011).
84. U. A. Perego, A. Achilli, N. Angerhofer, M. Accetturo, M. Pala, A. Olivieri, *et al.*, Distinctive Paleo-Indian Migration Routes from Beringia Marked by Two Rare mtDNA Haplogroups. *Current Biology*. **19**, 1–8 (2009).
85. A. Achilli, U. A. Perego, H. Lancioni, A. Olivieri, F. Gandini, B. H. Kashani, *et al.*, Reconciling migration models to the Americas with the variation of North American native mitogenomes. *Proceedings of the National Academy of Sciences*. **110**, 14308–14313 (2013).
86. A. Achilli, A. Olivieri, M. Pala, B. Hooshiar Kashani, V. Carossa, U. A. Perego, *et al.*, Mitochondrial DNA Backgrounds Might Modulate Diabetes Complications Rather than T2DM as a Whole. *PLoS ONE*. **6**, e21029 (2011).
87. U. A. Perego, N. Angerhofer, M. Pala, A. Olivieri, H. Lancioni, B. H. Kashani, *et al.*, The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. *Genome Research*. **20**, 1174–1179 (2010).
88. D. M. Behar, M. van Oven, S. Rosset, M. Metspalu, E.-L. Loogväli, N. M. Silva, *et al.*, A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. *The American Journal of Human Genetics*. **90**, 675–684 (2012).
89. A. Hartmann, M. Thieme, L. K. Nanduri, T. Stempfl, C. Moehle, T. Kivisild, *et al.*, Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum. Mutat.* **30**, 115–122 (2009).
90. R. S. Just, T. M. Diegoli, J. L. Saunier, J. A. Irwin, T. J. Parsons, Complete mitochondrial genome sequences for 265 African American and U.S. "Hispanic" individuals. *Forensic Science International: Genetics*. **2**, e45–e48 (2008).
91. M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, Mitochondrial genome variation and the origin of modern humans. *Nature*. **408**, 708–713 (2000).
92. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*. **35**, 1547 (2018).

93. J. W. Leigh, D. Bryant, POPART : full-feature software for haplotype network construction. *Methods Ecol Evol*. **6**, 1110–1116 (2015).
94. J. Heled, A. J. Drummond, Bayesian inference of population size history from multiple loci. *BMC Evol Biol*. **8**, 289 (2008).
95. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, *et al.*, BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol*. **10**, e1003537 (2014).
96. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. *Mol Biol Evol*, msw260 (2016).
97. P. Soares, L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Röhl, *et al.*, Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *The American Journal of Human Genetics*. **84**, 740–759 (2009).
98. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science*. **40**, 4477–4482 (2013).
99. Y. Huang, H. Ringbauer, hapCon: estimating contamination of ancient genomes by copying from reference haplotypes. *Bioinformatics*. **38**, 3768–3777 (2022).
100. The 1000 Genomes Project Consortium, D. G. Poznik, et al., Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics*. **48**, 593–599 (2016).
101. The 1000 Genomes Project Consortium, An integrated map of genetic variation from 1,092 human genomes. *Nature*. **491**, 56–65 (2012).
102. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, *et al.*, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. **81**, 559–575 (2007).
103. A. Moreno-Estrada, et al., The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science*. **344**, 1280–1285 (2014).
104. B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: A Discriminative Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of Human Genetics*. **93**, 278–288 (2013).
105. M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, *et al.*, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. **463**, 757–762 (2010).
106. P. Flegontov, N. E. Altınışık, P. Changmai, N. Rohland, S. Mallick, N. Adamski, *et al.*, Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature*. **570**, 236–240 (2019).
107. M. Rasmussen, et al., The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*. **506**, 225–229 (2014).
108. C. Posth, N. Nakatsuka, I. Lazaridis, P. Skoglund, S. Mallick, T. C. Lamnidis, *et al.*, Reconstructing the Deep Population History of Central and South America. *Cell*. **175**, 1185-1197.e22 (2018).
109. C. de la Fuente, M. C. Ávila-Arcos, J. Galimany, M. L. Carpenter, J. R. Homburger, A. Blanco, *et al.*, Genomic insights into the origin and diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proceedings of the National Academy of Sciences*. **115**, E4006–E4012 (2018).
110. J. M. Monroy Kuhn, M. Jakobsson, T. Günther, Estimating genetic kin relationships in

prehistoric populations. *PLoS ONE*. **13**, e0195491 (2018).
111. A. A. Behr, K. Z. Liu, G. Liu-Fang, P. Nakka, S. Ramachandran, pong: fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics*. **32**, 2817–2823 (2016).
112. Nei, Masatoshi, *Molecular Evolutionary Genetics* (Columbia University Press, New York).
113. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, *et al.*, The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. **538**, 201–206 (2016).
114. K. Nägele, C. Posth, M. Iraeta Orbegozo, Y. Chinique de Armas, S. T. Hernández Godoy, U. M. González Herrera, *et al.*, Genomic insights into the early peopling of the Caribbean. *Science*. **369**, 456–460 (2020).
115. The 1000 Genomes Project Consortium, R. A. Gibbs, E. Boerwinkle, H. Doddapaneni, Y. Han, V. Korchina, *et al.*, A global reference for human genetic variation. *Nature*. **526**, 68–74 (2015).

Acknowledgments

Acknowledgments: We greatly acknowledge the IT support of Luis Alberto Aguilar Bautista, Alejandro de León Cuevas, Carlos Sair Flores Bautista and Jair Garcia Sotelo of the Laboratorio Nacional de Visualización Científica Avanzada (LAVIS UNAM). We also thank Alejandra Castillo Carbajal and Carina Uribe Díaz from LIIGH-UNAM for wetlab technical support during the project. We thank Alexandra Sockell for technical support at the Bustamante Lab. We thank Leonardo Correa-Mendoza for technical support with qpWave. We thank Diego Ortega-Del Vecchyo and Angélica González Oliver for enriching comments throughout the project on earlier versions of the manuscripts. We thank Yazmín Jiménez Islas for her help in the Summary figure. We also thank Rafael Montiel Duarte for his early contribution in the Cañada de la Virgen project.

Funding: Viridiana Villa-Islas is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and has received CONACyT fellowship (Número de becaria: 282188). M.C.A.-A. received support from UNAM-PAPIIT (Project numbers: IA201219 and IA203821) and the Wellcome Trust Seed Award in Science (grant number: 208934/Z/17/Z), and the International Centre for Genetic Engineering and Biotechnology (project number: CRP/MEX10-03). M.C.A.-A., E.H.-S., and F.J. receive support from the Human Frontiers Science Program (grant number: RGY0075/201). R.F. was supported by a grant (PALEUCOL; PGC2018-094101-A-I00) from FEDER/Ministerio de Ciencia e Innovación—Agencia Estatal de Investigación and from CajaCanarias and La Caixa Banking foundations (GENPAC; 2018PATRI16). Data generation for the Sierra Tarahumara Individuals was funded with a Trainee Research Grant to M.C.A.-A. by Stanford Center for Computational, Evolutionary and Human Genomics. E.H.-S. r received support from NIH R35GM128946 and the Alfred P. Sloan Foundation. Data generation for the Cañada de la Virgen individuals was supported by The Mexican Biobank Project (Grants FONCICYT/50/2016 and MR/N028937/1 from CONACYT and Newton Fund, respectively) awarded to A.M.-E. Also, this work has been partially supported by the government of Guanajuato (Grant No. 093/2016 CINVESTAV)

awarded to K.S. And, data generation for the Aridoamerican individuals was supported in part by a Leakey Foundation granted to F.A.V.

Author contributions: M.C.A.-A. and V.V.-I. conceived and designed the study with input from E.M.P.C., M.S.-V., M.B.-L., B.M., R.F., K.S., and M.A.N.-C. performed the wetlab work. V.V.-I. processed all sequence data. V.V.-I. carried out most analyses with help from, A.I.-G., M.L., J.M.T., J.E.R.-R., D.A.V.-R., and F.S.-Q. E.M.P.C., A.H.-M., K.S., G.Z. E.G.-M., M.V.-M, R.A.-H., A.G.-O., C.V. contributed samples, provided the archaeological context and discussed results. F.A.V., A.M.-E., F.J., A.I., E.H.-S.V.M.-M., and F.S.-Q. suggested analyses, discussed results and provided feedback throughout the project. V.V.-I. and M.C.A.-A. wrote and edited the manuscript with input from all authors.

Competing interests: The authors declare no competing interests.

Data and materials availability: Data are available from the European Nucleotide Archive under the project accession PRJEB51440.

References Supplementary Material:

116. J. Bracho, Narrativa e identidad. El mestizaje y su representación historiográfica. *LREL*, 55 (2009).
117. P. Wade, in *The Companion to Latin American Studies* (Routledge, 2014), pp. 197–211.
118. E. Mejía Pérez Campos, X. Chávez Balderas, R. Chávez Sánchez, Pepita. La momia de la Sierra Gorda de Querétaro. *Arqueología mexicana*, 70–75 (2009).
119. E. Mejia Pérez Campos, thesis, Escuela Nacional de Antropología e Historia, Ciudad de México, Méx (2010).
120. G. Zepeda García Moreno, Cañada de la Virgen, refugio de los muertos y los ancestros. *Guanajuato, Ediciones La Rana* (2010).
121. I. D. Lara Barajas, G. Islas Estrada, G. Zepeda García Moreno, "Informe Final del Análisis Antropofísico del Entierro 19," *Proyecto Arqueológico Cañada de la Virgen* (Centro INAH-Querétaro, 2011).
122. A. Goguitchaichvili, V. Ortega, J. Archer, J. Morales, A. T. Guerrero, Absolute geomagnetic intensity record from pre-Columbian pottery dates elite Tlailotlacan Woman in ancient Teotihuacan. *Journal of Archaeological Science: Reports*. **14**, 146–151 (2017).
123. A. Brito-Mayor, A dos y a cuatro patas: el Occidente de Mesoamérica y la relación entre el perro, el humano y viceversa. Una aproximación a través de estudios de caso. *A*. **28**, 169 (2019).
124. T. Maricic, M. Whitten, S. Pääbo, Multiplexed DNA Sequence Capture of Mitochondrial Genomes Using PCR Products. *PLoS ONE*. **5**, e14004 (2010).
125. M. L. Carpenter, J. D. Buenrostro, C. Valdiosera, H. Schroeder, M. E. Allentoft, M. Sikora, *et al.*, Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *The American Journal of Human Genetics*. **93**, 852–864 (2013).
126. A. García, R. Nores, J. M. B. Motti, M. Pauro, P. Luisi, C. M. Bravi, *et al.*, Ancient and modern mitogenomes from Central Argentina: new insights into population continuity, temporal depth and migration in South America. *Human Molecular Genetics*. **30**, 1200–1217 (2021).
127. A. González-Martín, A. Gorostiza, L. Regalado-Liu, S. Arroyo-Peña, S. Tirado, I. Nuño-Arana, *et al.*, Demographic History of Indigenous Populations in Mesoamerica Based on mtDNA Sequence Data. *PLOS ONE*. **10**, e0131791 (2015).
128. B. Z. González-Sobrino, et al., Genetic Diversity and Differentiation in Urban and Indigenous Populations of Mexico: Patterns of Mitochondrial DNA and Y-Chromosome Lineages. *Biodemography and Social Biology*. **62**, 53–72 (2016).
129. S. Romero-Hidalgo, et al., Demographic history and biologically relevant genetic variation of Native Mexicans inferred from whole-genome sequencing. *Nature Communications*. **8**, 1005 (2017).
130. J. Ramos-Madrigal, B. D. Smith, J. V. Moreno-Mayar, S. Gopalakrishnan, J. Ross-Ibarra, M. T. P. Gilbert, *et al.*, Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into the Early Stages of Maize Domestication. *Current Biology*. **26**, 3195–3201 (2016).
131. R. Acuna-Soto, Megadrought and Megadeath in 16th Century Mexico. *Emerging Infectious Diseases*. **8**, 360–362 (2002).
132. G. Kubler, *The Population of Central Mexico in the Sixteenth Century.* (JSTOR, 1948).
133. P. Skoglund, B. H. Northoff, M. V. Shunkov, A. P. Derevianko, S. Pääbo, J. Krause, *et al.*, Separating endogenous ancient DNA from modern day contamination in a Siberian

Neandertal. *Proceedings of the National Academy of Sciences*. **111**, 2229–2234 (2014).
134. R. Fregel, F. L. Méndez, Y. Bokbot, D. Martín-Socas, M. D. Camalich-Massieu, J. Santana, *et al.*, Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proc Natl Acad Sci USA*. **115**, 6774–6779 (2018).
135. S. J. Macfarlan, C. N. Henrickson, Inferring Relationships Between Indigenous Baja California Sur and Seri/Comcáac Populations Through Cultural Traits. *Journal of California and Great Basin Anthropology*. **30**, 18 (2021).
136. J. V. Moreno-Mayar, et al., Early human dispersals within the Americas. *Science*. **362**, eaav2621 (2018).
137. E. M. P. Campos, INTERPRETACIÓN PRELIMINAR RESPECTO A LA TEMPORALIDAD DE TOLUQUILLA, QUERÉTARO, 24.
138. H. Rangel-Villalobos, V. M. Martínez-Sevilla, G. Martínez-Cortés, J. A. Aguilar-Velázquez, M. Sosa-Macías, R. Rubi-Castellanos, *et al.*, Importance of the geographic barriers to promote gene drift and avoid pre- and post-Columbian gene flow in Mexican native groups: Evidence from forensic STR Loci: GEOGRAPHIC BARRIERS, DIFFERENTIATION AND ADMIXTURE IN AMERINDIANS. *Am. J. Phys. Anthropol.* **160**, 298–316 (2016).

Figures and Tables

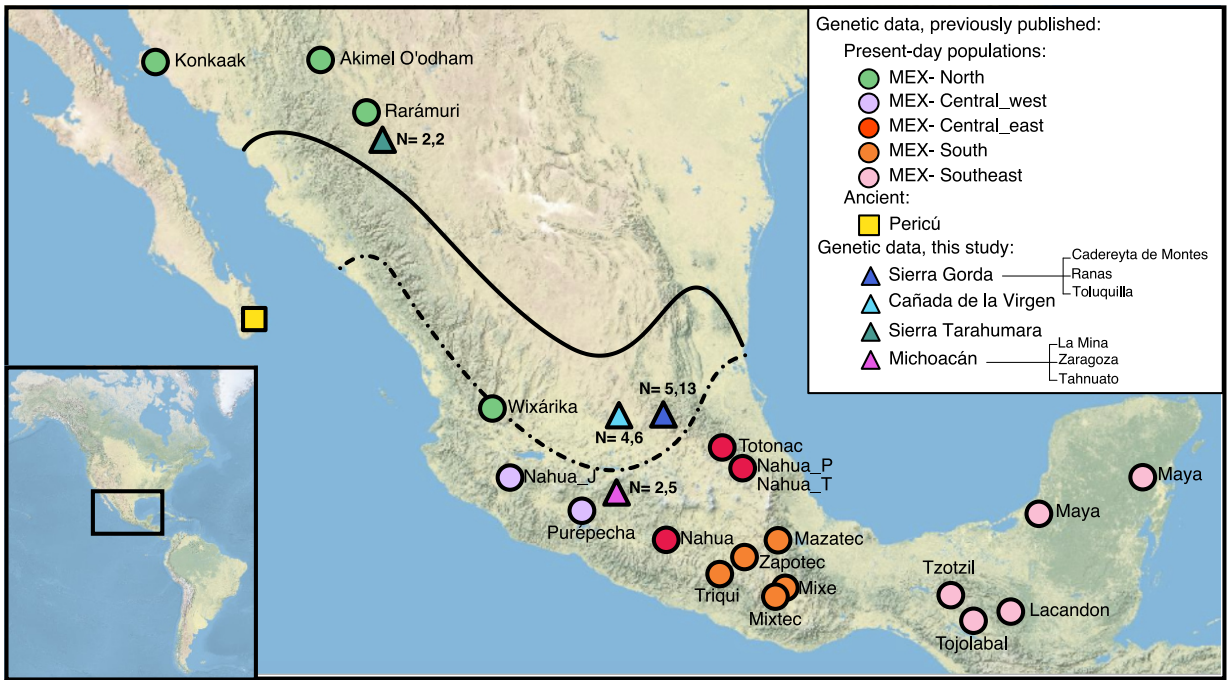

Fig.1. Map of archeological sites and modern populations. Map of Mexico indicating the location of the pre-Hispanic sites analyzed here (triangles) and in previous studies (squares) (*15*, *16*) and the approximate locations of present-day Indigenous populations sampled in previous studies (circles) (*16*, *23*, *37*, *113*). The continuous line indicates the northern frontier of Mesoamerica, with Aridoamerica north to this and Mesoamerica south to this one during the year ca. 800-900 CE. Dashed line indicates this border after the year ca. 1500 CE, based on (*2*). Numbers next to archaeological sites from this study indicate the number of whole-genome data and mitochondrial genomes recovered per site (separated by a comma).

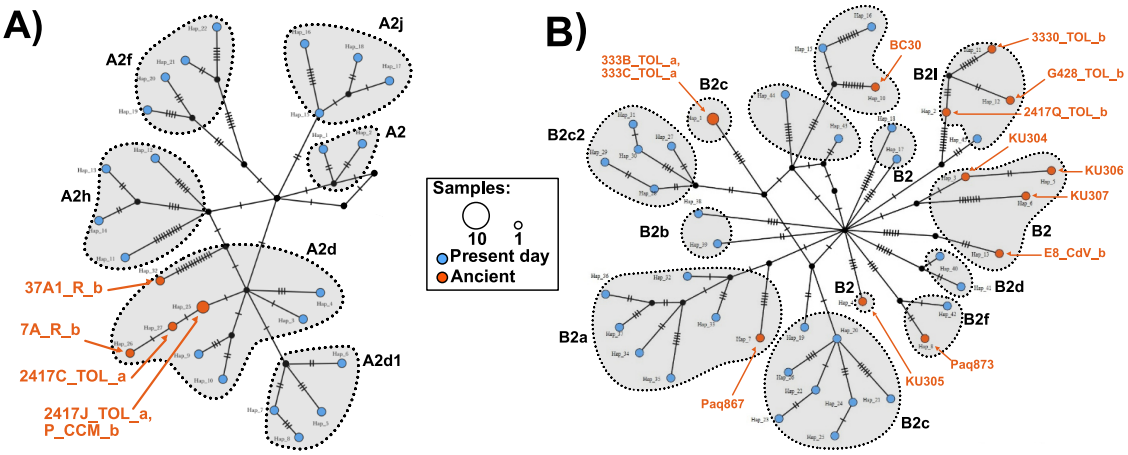

Fig. 2. mtDNA Haplotype networks. (A) Haplotype network of mitochondrial sub haplogroups A2 including present-day and pre-Hispanic mitochondrial haplotypes from Mexico. Arrows point to individuals from Sierra Gorda within haplogroup A2d from before and after the drought period. (B) Haplotype network of mitochondrial sub haplogroups of the lineage B2 including present-day and pre-Hispanic mitochondrial haplotypes from Mexico. Arrows point to pre-Hispanic individuals from Cueva Candelaria, Paquimé, Sierra Gorda, and Cañada de la Virgen.

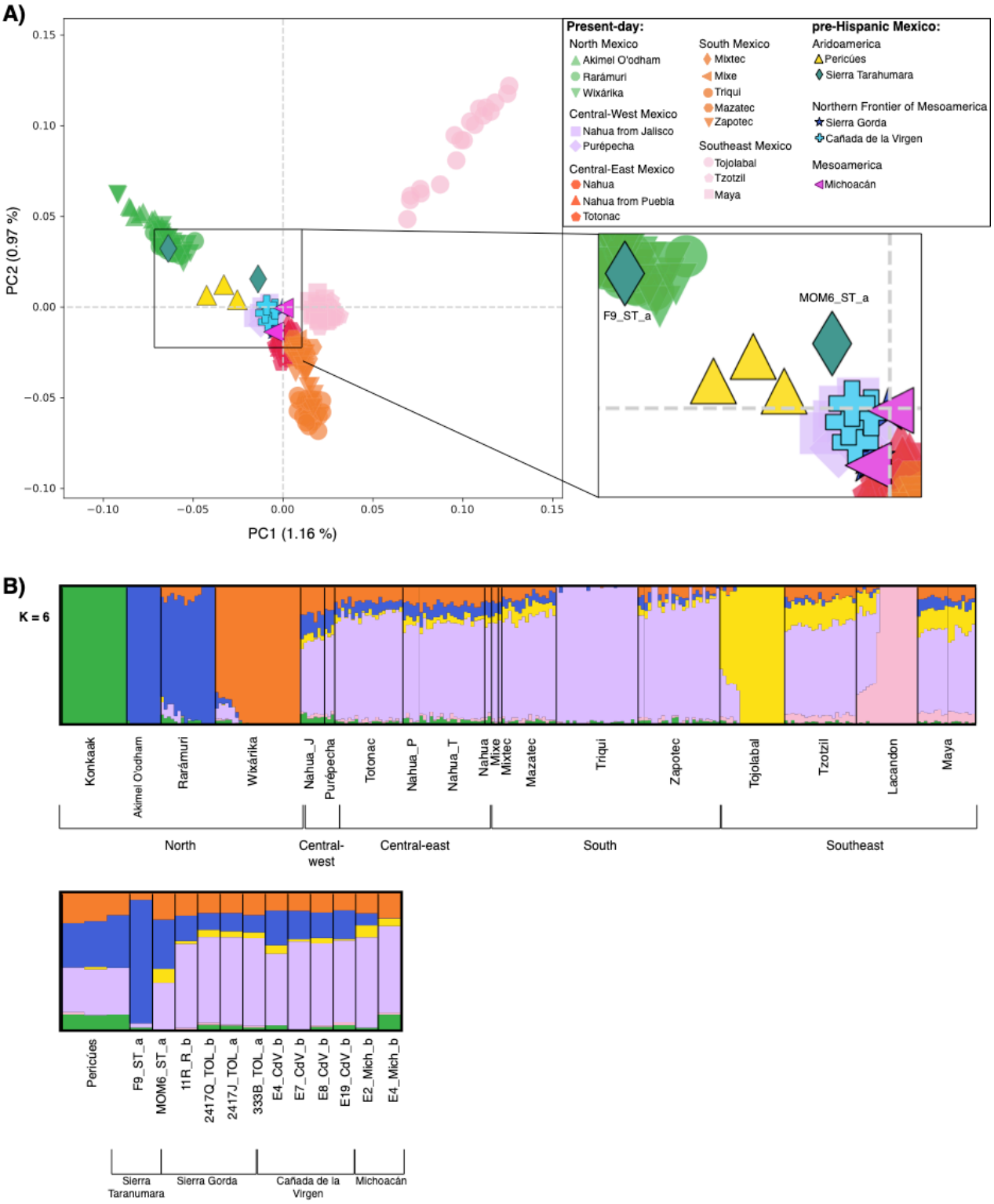

Fig.3. Genetic structure of Indigenous and pre-Hispanic individuals. (A) Missing DNA Principal Component Analysis (mdPCA) of present-day Indigenous Populations with the ancient and pre-Hispanic genomes from Mexico. (B) Unsupervised clustering analysis of same data using ADMIXTURE with k=6, showing

component green, blue, and orange in northern populations, component yellow and pink in southeast populations and component violet in south and central Mexico populations.

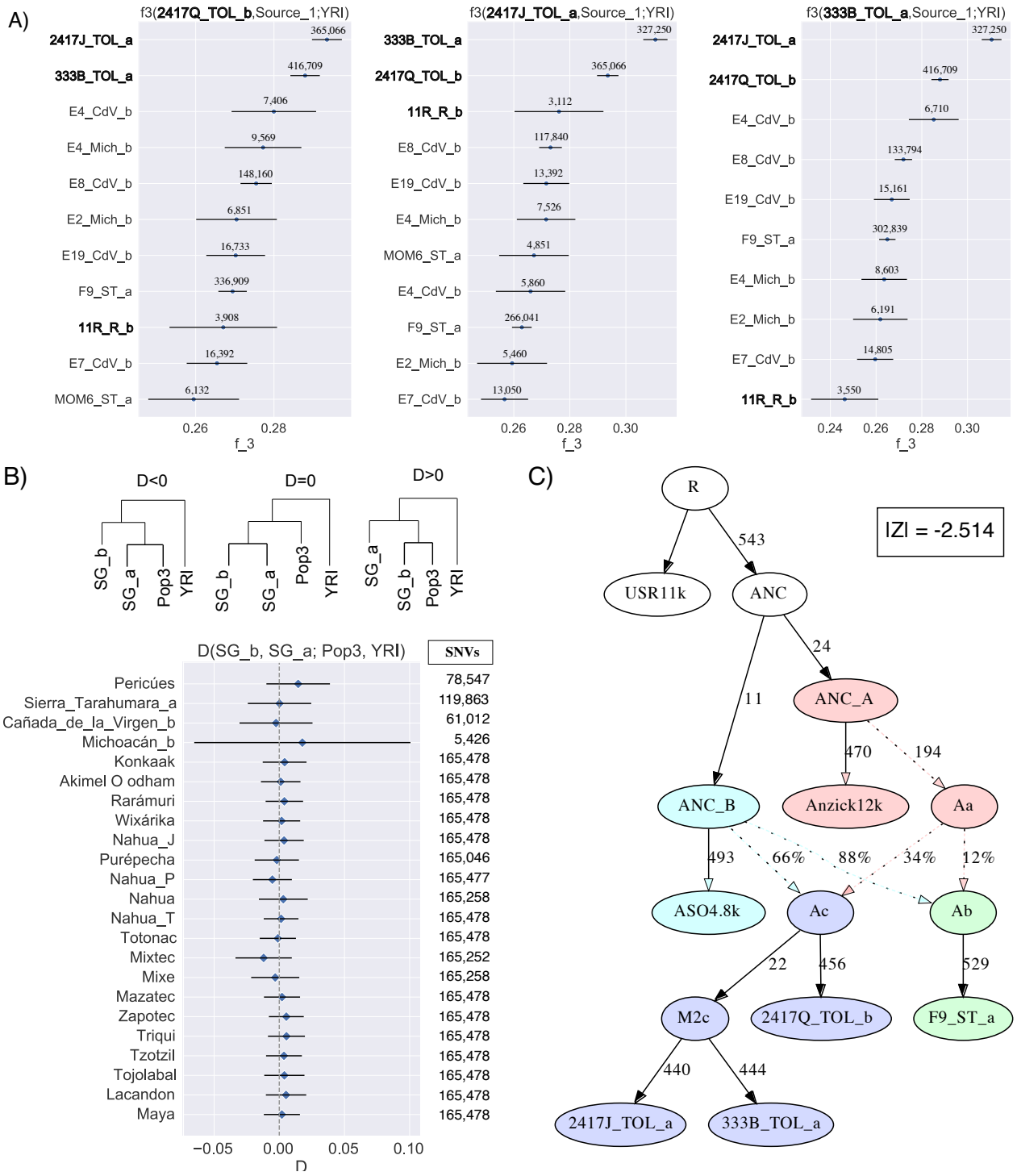

Fig.4. Population continuity in Sierra Gorda. (A) Shared genetic drift and genetic affinities of pre-Hispanic individuals from Toluquilla, 2417Q_TOL_b, 2417J_TOL_a, and 333B_TOL_a compared with pre-Hispanic individuals from Mexico. Higher values of f3 indicate higher shared genetic drift. Point estimates and one standard error are shown. (B) D statistics in the form D(SG_b, SG_a; Pop3,YRI) including transversions only. SG_b includes 2417Q_TOL_b and

11R_R_b, and SG_a 2417J_TOL_b and 333B_TOL_a. Pop3 is any of the other pre-Hispanic or present-day populations from Mexico and is shown in the Y axis. YRI as an outgroup corresponds to Yoruba from Africa from the 1000 Genomes Project (*38*, *101*). Expected tree topologies according to different D values are drawn on the top of the plot. Blue dots indicate no significant deviations from D=0 (|Z|>3). The number of SNV's used in each test is shown in the right. (C) qpGraph model of individuals from Sierra Gorda from before and after the drought are shown in the same clade, without contribution of F9_ST_a to the post-drought individuals.

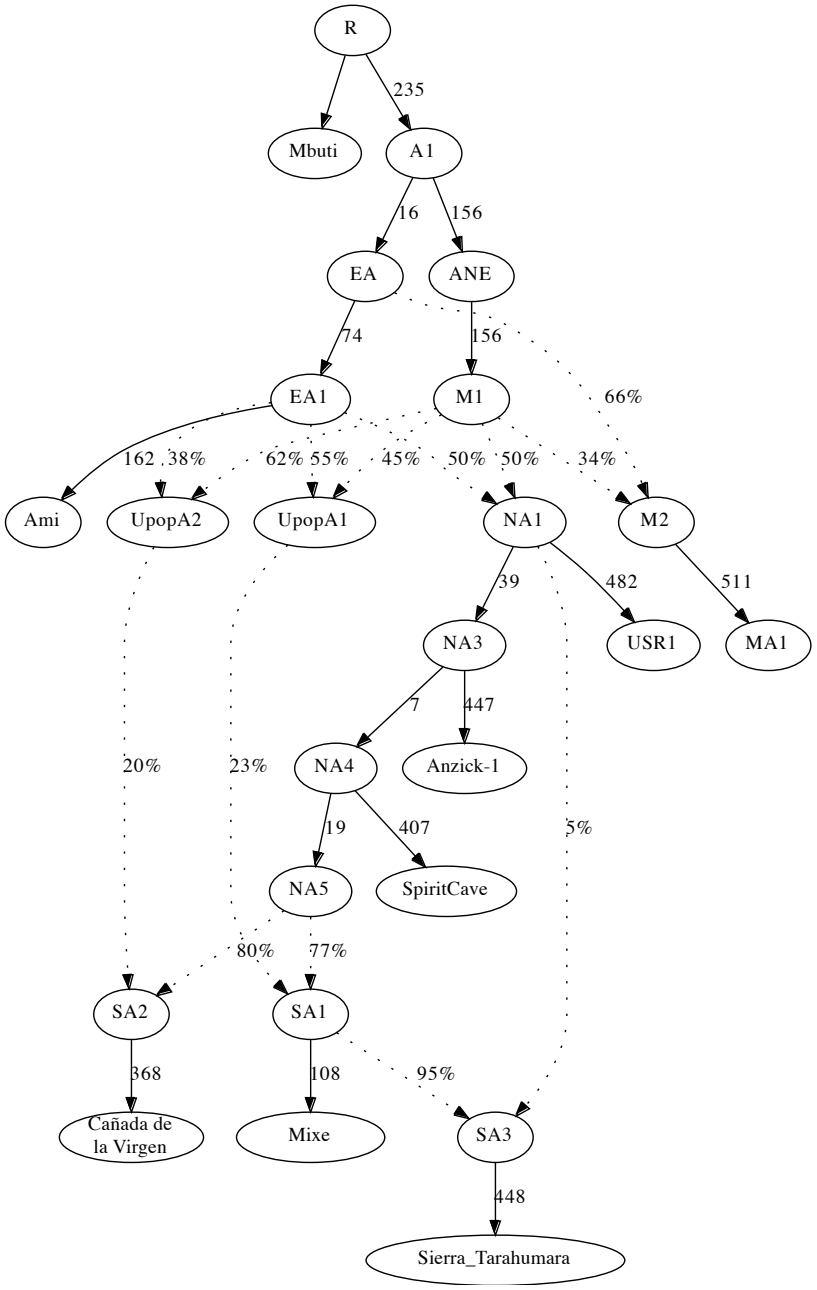

Fig.5. Ghost ancestries in pre-Hispanic individuals from Mexico. qpGraph model for Sierra Tarahumara and Cañada de la Virgen including present-day Mixe. The tree has no inner 0 branches and a |Z-score| of 2.415. Model shows different ghost ancestries, UPopA1 and UPopA2, in Sierra Tarahumara and Cañada de la Virgen individuals, respectively. UPopA1 stands for the previously unsampled population genetic ancestry reported in Mixe, while UPopA2 is a previously unidentified genetic ancestry. All f statistics are within 1.75 standard error.

Supplementary Materials for

Demographic history and genetic structure in Pre-Hispanic Central Mexico

Viridiana Villa-Islas[1], Alan Izarraras-Gomez[1], Maximilian Larena[2], Elizabeth Mejía Perez Campos[3], Marcela Sandoval-Velasco[4-6], Juan Esteban Rodríguez-Rodríguez[1], Miriam Bravo-Lopez[1], Barbara Moguel[1,7], Rosa Fregel[8], Ernesto Garfias-Morales[1], Jazeps Medina Tretmanis[9], David Alberto Velázquez-Ramírez[10], Alberto Herrera-Muñóz[3], Karla Sandoval[11], Maria A. Nieves-Colón[12,13], Gabriela Zepeda García Moreno[14], Fernando A Villanea[15], Eugenia Fernández Villanueva Medina[16], Ramiro Aguayo-Haro[16], Cristina Valdiosera[17,18], Alexander G. Ioannidis[19], Andrés Moreno-Estrada[12], Flora Jay[20], Emilia Huerta-Sanchez[9], J. Víctor Moreno-Mayar[21], Federico Sánchez-Quinto[22], María C. Ávila-Arcos[1]*.

*Corresponding author. Email: mavila@liigh.unam.mx

This PDF file includes:

Supplementary Text
Figs. S1 to S33
References (116-138)

Other Supplementary Materials for this manuscript include the following:

Supplementary_Table_S1_Sample_summary_infomation.xlsx
Supplementary_Table_S2_Sample_infomation_per_sequencing.xlsx
Supplementary_Table_S3_Reference_public_ancient_nuclear_DNA.xlsx
Supplementary_Table_S4_mtDNA_Mexico.xlsx
Supplementary_Table_S5_mtDNA_new_variants.xlsx
Supplementary_Table_S6_mtDNA_for_Networks_and_EBSP.xlsx
Supplementary_Table_S7ab_EBSP_statistics.xlsx
Supplementary_Table_S8_Reference_panel_present-day.xlsx
Supplementary_Table_S9_F3outgroup_prehispanic-present-day.xlsx
Supplementary_Table_S10_Dstats_prehispanic-present-day.xlsx
Supplementary_Table_S11_individuals_sites_for_CND_analysis.xlsx
Supplementary_Table_S12_CND_pairwise_values.xlsx
Supplementary_Table_S13_hapROH.xlsx
Supplementary_Table_S14_F3outgroup_prehispanic-prehispanic.xlsx
Supplementary_Table_S15_Dstats_SG_and_prehispanic_each_individual.xlsx
Supplementary_Table_S16_Dstats_SG_and_prehispanic_per_site_Mexico.xlsx

Supplementary_Table_S17_Dstats_SG_and_prehispanic_per_site_Mexico_Transversions.xlsx

Supplementary_Table_S18_qpWave.xlsx

Supplementary Text

Full Ethics statement

The sampling of the archeological human remains was made upon approval by the Consejo de Arqueología (Archaeology Council) of the Instituto Nacional de Antropología e Historia (National Institute of Anthropology and History) of each collection. The permit numbers are: 401.3S.16-2017/990, 401.3S.16-2019/222 and 401.1S.3-2022/1152.

We sampled the minimum possible amount of tissue to avoid unnecessary destruction of these precious materials. For teeth, we tried to separate the root from the crown without damaging the crown. We returned all leftover material to the archeologists responsible for each collection. In addition, we took photographs of each sample during several steps of the sample processing to have a visual archive. Because of historical and socio-political reasons, together with the prominent admixture of different continental ancestries in the territory and the imposed institutionalized narrative of "mestizaje" (*116*, *117*) to minimize the contribution of present-day Indigenous peoples to our national identity (something utterly condemnable), it becomes quite complex to assign a present-day population in Mexico as the single direct descendant of the remains of human ancestors recovered from archaeological contexts. This complexity raises several questions regarding consultation of descendant populations for this kind of research (e.g. should admixed "mestizo" populations who share the geographical area be consulted? who should be considered a representative for them? under what capacity? In part because of these uncertainties, consultation with Indigenous populations for destructive analysis of archeological remains is not a standard procedure in the country. This is on its own an emerging focus of discussion that warrants further investigation and debate beyond the scope of this research (but see Ávila-Arcos et. al 2022). However, a series of efforts to communicate genetic findings resulting from this work to local communities from the sites of Toluquilla and Ranas have taken place over the course of the project.

Sample overview

Samples include skeletal remains from Northern Mesoamerica: Sierra Gorda - Querétaro (n=13), Cañada de la Virgen - Guanajuato (n=16) and Central Mesoamerica: Michoacán (n=8); along with updated genomic data from Aridoamerica: Sierra Tarahumara – Chihuahua (n=2), previously reported at less resolution (*1*) (Table S1). The most extended period covered here belongs to Sierra Gorda samples from 320 BCE to 1,351 CE (Classic-Postclassic). While the rest of the archaeological sites correspond to Classic (Cañada de la Virgen and Michoacán), and Postclassic (Sierra Tarahumara).

Archaeological context

Sierra Tarahumara (Mummy-F9 and Mummy-MOM6)

These individuals and their archaeological context were previously published in (*1*)

Sierra Gorda, Querétaro (Mummy-P)

Cadereyta cave, Sierra Gorda, Querétaro (P_CCM_b)

In 2002, the discovery of the mummified remains of individual P_CCM_b was reported in the vicinity of the village of Altamira, in the municipality of Cadereyta, Querétaro. The remains were removed and taken to the public prosecutor's offices in the municipal capital. Once the remains were handed over to INAH and archaeologist Elizabeth Mejia Perez Campos, the site was visited and cleaned. Fragments of textiles, organic matter such as bird feathers and wood formed a bracelet, and human hair braids were found. The piece was taken to the Templo Mayor Museum in Mexico City, where computerized axial tomography, necropsy with an endoscope to see internal organs and samples were taken. Likewise, mtDNA studies were performed by PCR amplification of the HVR at the Centro de Investigación y Estudios Avanzados of the Instituto Politécnico Nacional (CINVESTAV - IPN) in Mexico City and the textile and skin were dated by C13 and C14 at Beta Analytical in Florida, USA. The results showed a twenty-months-old infant who lived in 340 BC, shrouded with a cotton blanket with human hair interwoven and with an offering of maguey quills, stork feathers, pine, oak, and sotol leaves (*118*).

Toluquilla, Sierra Gorda, Querétaro

Toluquilla archaeological site is located in the Sierra Gorda, in the State of Querétaro. The occupation of Toluquilla spans ca. 2,000 years, from 400 BCE to 1,550 CE.

Eight individuals, 333A_TOL_a, 333B_TOL_a, 333C_TOL_a, 333O_TOL_b, 333Q_TOL_b, 2417C_TOL_a, 2417Q_TOL_b and, 2417J_TOL_a, from the archaeological site of Toluquilla were sampled for this study. We analyzed samples from individuals from two ritual burials at the site. The burials were analyzed by Dr. Elizabeth Mejia Perez Campos as part of her doctoral thesis (*6*) and were named based on the relevance of the offerings. These burials correspond to the 'Manatee context', and the 'Mirror context'.

Manatee Context (Individuals 333x)

The manatee context has a great quantity, quality, and variety of offerings. It includes shells and snails from both coasts (Gulf of Mexico and the Pacific Ocean), obsidian from the Sierra de las Navajas (Hidalgo, Mexico), local animals such as river fish lynx, armadillo plates, and foreign aquatic species from the deep sea.

Individuals 333A_TOL_a, 333B_TOL_a and 333C_TOL_a 333O_TOL_b, 333Q_TOL_b from this context were sampled for this study. This context is located in building 33 for which dates are available, either by C14 dating, obsidian hydration dating, or the analysis of ceramic materials.

In this context, the more recent burial is dated between 1,300 and 1,400 CE. In this burial individuals 333A_TOL_a, 333B_TOL_a, and 333C_TOL_a were found. Charcoal was deposited among the sediment that covered the bodies and has been C14 dated to 1,350 ± 50 CE.

333A_TOL_a was a 25-30-year-old female. She had a pendant made from a lynx tusk placed as a trousseau on her chest, possibly hung with knots since it has no orifice, in addition to a necklace of Pacific olive shells and a human jaw with a perforation as a pendant. As an offering, a piece in the shape of a star or flower made with a spondylus snail, a stingray caudal spine, a natural shell, a fragment of river fish, an obsidian flake, a yew, and a bean-shaped piece with a mixture of cinnabar and red ocher were deposited under the body (*119*).

333B_TOL_a was an infant individual; he was found carefully arranged, leaving the arm in almost total order, first the humerus, radius, and ulna. However, the latter was placed upside down, and in the place of the wrist, a necklace was placed as a trousseau. The necklace comprises four pieces that form a snake, the first of a cylindrical clay bead, another similar one of bone, a flat and square bead of the mother-of-pearl shell to finish with an elongated piece also of mother-of-pearl shell, which represents the rattlesnake's rattle. It is deduced that this burial is of ritual type, since the coxal bones of the child were placed around the circle, and as an offering, there is a complete shell, an earring made with the incisor of a six or seven-year-old child, and a necklace of shells.

Individual 333C_TOL_a was a woman between 20 and 21 years old. A number of pieces were found with her. Her trousseau was a necklace of 1,150 beads and three snail earrings. Also, as offerings on the body there was a button-like circular piece, an armadillo plaque, an obsidian flake, shuffleboard or ceramic circle, and, next to her, a pectoral made on a manatee rib, which gives its name to this archaeological context.

Other older burials were found in this building. Individual 333O_TOL_b was found next to the platform wall, C14 dated to 700 ± 50 CE. Individual 333Q_TOL_b was found in an older burial C14 dated to 520 ± 40 CE. The platform was to the west of the altar, and the burial was 50 cm deep, it corresponds to the first phase of occupation (*119*).

Mirror context (Individuals 2417x)

Individuals 2417Q_TOL_b, 2417C_TOL_a and 2417J_TOL_a from this context were sampled for this study. This archaeological context is composed of sixteen individuals; it is located in front of the altar of building 24. Among the fill, scattered bones of additional thirty-five people were found. The bones in the fill and the closeness of the primary individuals suggest that the deposition of the primary bodies was at a single time, occupying the bones of previous individuals as fill.

A total of four levels of human body deposition were observed. There is at least one primary body surrounded by secondary bodies, remnants of previous burials, and filler bones in each level. In the first and most superficial level, in front of the altar, a group formed by a woman's primary body (2417C_TOL_a) and the secondary remains of three children (not analyzed in this study) were placed. The remains form a U shape, with the open portion pointing west. The bones of the woman 2417C_TOL_a seem to have received scalding liquids or some other substance, which produces an effect similar to boiling the bones, as they appear crystallized, although additional studies are required to define what was the precise taphonomic process. In addition, she presented pathologies such as spondylitis (the fusion of the last vertebra and the sacrum) and poor dental health, orbital sieve and marks at the insertions of the arms. The woman had as trousseau a necklace formed by

sixteen river mollusks of the genus *Unio*, and as an offerings a snail button and obsidian flake and razor. The offering of individual 2417C_TOL_a is significant for its quality, variety, abundance, and treatment given to the corpse.

Five individuals were found in the third level of the northeast of the assemblage. The most relevant is individual 2417J_TOL_a, who presents pathologies such as spondylitis, evident in the sacrum, iliac and lumbar vertebrae. In addition to poor dental health, he also performed exercises that developed his muscles and left a mark on his arms and legs.

In the fourth and last level, the remains of four more individuals were located, all adults, two secondary, and two remnant primary portions. Individual 2417Q_TOL_b corresponds to a remnant primary portion that was placed in the ventral decubitus position (face down) of a male between forty-five and fifty-five years old. Under individual 2417Q_TOL_b, there were two isolated bones, a coxal, and a vertebrae. Also under this individual, a pyrite mosaic mounted on a ceramic base was found, thought to serve as a mirror. The mirror is a significant offering; it is related to the priests and the god Tezcatlipoca, the deity of the underworld and the night, characterized because it carries a mirror, which is also used in the arts of divination, to obtain omens and to provoke evils (*119*).

Individual 6428A_TOL_b

The body of individual 6428A_TOL_b was found in building 64, being part of burial 28. It corresponds to an adult male in a flexed dorsal decubitus position with a significant offering including a flint bifacial, two jade beads, and ten amulets carved in bone (*119*).

Ranas, Sierra Gorda, Querétaro

Ranas archaeological site is located in the Sierra Gorda, in the State of Querétaro, and the occupation of the site is recorded from 400 BCE to 1,300 CE.

Individual 11R_R_b

Individual 11R_R_b was found in a burial that contained three individuals, one primary (11R_R_b) and two secondary arranged on the sides of the primary. 11R_R_b corresponds to an adult male over 35 years old. He was found deposited in dorsal decubitus, with the trunk slightly flexed upwards. The arms were placed in front of the body and the wrists were touching each other. His hands were on the pelvic region and in the right hand he was holding a ceramic pipe. Two bone circles forming a truncated cone were found in the area of the pubic symphysis. He presented osteological characteristics of repetitive work and high stress to the body. The enteropathies in the long bones of the legs, arm, vertebrae and calcaneus may be directly related to mining. Thus, individual 11R_R_b is inferred to have worked in mining for several years. The measurement of high values of heavy metals in a combination of Pb, Hg, As and Sb indicate chronic exposure to a wide variety of heavy metals present in geological contexts, which were systematically deposited in the bones of the individual (*57*).

The relative chronology of the excavated materials and those directly associated with a more superficial burial context suggested a chronology between 450 and 700 CE. In terms of

Mesoamerican periodicity, it corresponds to the transition between the Classic and Epiclassic periods (*57*).

Individuals 7A_R_b and 3A7I_R_b were found in structure 3A, in burial 7 at the archaeological site of Ranas. Burial 7 is a complex burial, where initially five areas of intrusion are defined, and inside each of them, there is evidence of skeletal remains of one or more individuals.

Individual 7A_R_b

This individual is the most important since it is located just below the altar. It is a direct primary burial with a very flexed seated disposition. The individual was deposited as a mortuary bundle strongly tied (at the height of the shoulders and waist), so when the body decomposed in its taphonomic process, the head fell between the legs. The care taken in the deposition and the fact that limestone slabs covered the individual from its lowest level helped to keep most of the skeleton in its primary position, where the only element of post-depositional alteration was a rodent burrow that altered the metatarsal bones and phalanges of the left foot and altered the distal phalanges of the right foot.

Individual 3A7I_R_b

The case of individual 3A7I_R_b is another burial of an important dignitary since it is the only one deposited in an extended form with the head to the south, its face looking to the east, with the right arm flexed on its chest. The hand was deposited on its left shoulder, while its left arm rested on the belly, and its hand is extended between the right ribs and the right humerus, with the torso slightly leaning forward. The pelvic region was horizontal on the ground while the legs were separated with the right knee to the east, the left knee to the west, and the feet under the coxae. Therefore, this is a primary burial, in dorsal decubitus slightly raised from the back and head, with legs flexed on themselves. As it is a primary individual, the different body portions could be recognized and recovered, although in many cases, they were found in regular and poor preservation.

Cañada de la Virgen, Guanajuato

Individual E2_CdV_b

Individual E2_CdV_b is a direct secondary burial, located in layer III of Table V-4 (although it partially intrudes up to layer IV). The burial presents only proximal fragments of incomplete femurs corresponding to an adult individual of undetermined age. Although it is possible that its partial intrusion into floor 1 and layer IV is accidental, the only difference with respect to the rest of the skeletons from Room 3 is the fact that it does not have contact with floor 2, but is found right at the level of floor 1. It seems more likely that this burial, due to its position and secondary character, functioned as a ritual seal for floor 2, in a termination ritual aimed at "killing the funerary space between the two floors (*120*).

Individual E4_CdV_b

Located in layer IV of Table X-4. It is an indirect primary burial deposited in a cist in extended dorsal decubitus facing east. The estimated age is 5 years +/- 24 months and indeterminate sex. It presents 60% of the skeleton with absence of bones of the feet, patella, sacrum and coccyx (*120*).

Individual E6_CdV_b

Located in layer IV of Table V-3. Indirect secondary burial in cist, corresponding to a male individual of 40-45 years of age. It presents only fragmented iliacs, complete right radius, proximal part of both femurs at 50% and part of the proximal portion of the right tibia. Position and orientation not discernible (*120*).

Individual E7_CdV_b

Located in layer IV of Table X-3. Indirect primary burial deposited in cist in dorsal decubitus flexed oriented to the east. Corresponds to a female of 25 to 30 years of age. Feet and hands are missing (*120*).

Individual E8_CdV_b

Located in layer IV of table U-3. Direct primary burial in flexed dorsal decubitus oriented to the east. It corresponds to a male individual between 35 and 40 years old. He presents 90% of the skeleton with absence of feet (*120*).

Individual E9_CdV_b

Located in layer IV of Table T-4. It is an indirect primary burial deposited in flexed dorsal decubitus with orientation to the east. It corresponds to a male subject of 20 to 25 years of age. It presents 80% of the skeleton with absence of feet (*120*).

Individual E10_CdV_b

Located in layer IV of box V-3. This is a direct secondary burial with no discernible anatomical position. It consists only of a left femur diaphysis and a mandible (*120*).

Individual E11_CdV_b

Located in layer IV of squares U/V-3 and 4. It is a direct secondary burial formed by three fragments of long bones, fragments of ribs, a clavicle, some phalanges, fragments of pelvis and skull, as well as some molars (*120*).

Individual E14_CdV_b

Located in complex C, which is a circular structure, with an association to the wind and related to the North Star and Ursa Minor, which is linked to Tezcatlipoca. This burial was found at the foot of the access ramp to the second structure (*120*).

Individual E15_CdV_b

This individual was called the "girl of the rain". This subadult individual, through osteological method was estimated to be approximately 7 years old and possible a female, but we take this with caution since the individual is a subadult. She was found in complex B, and linked to the pluvial drainage, which is large and well-constructed with superimposed slabs. She was 1.10 m tall and is thought to be part of the ancestral veneration practiced in the place, since she was accompanied by a coyote and placed in such a way that it was necessary to decompose her body in order to accommodate her seated and compressed. Possible postmortem treatment in the temazcal that is at the top of the pyramidal base. The “girl of the rain”, carried in her neck a small necklace of beads that includes a bead in the form of butterflies in the central part of the necklace. The offering is completed with a ceramic spindle and a fragment of marine shell. Next to the girl and as a companion and vestige of the practices of ancestral veneration, a small coyote was placed -a complete specimen of *Canis latrans*, an animal of totemic importance in the region. Forensic studies performed on the girl of the rain provided her height, and some pathologies associated with infectious processes such as periostitis or infection of the periosteum, as well as cranial lesions probably diagnostic of tuberculosis (*120*).

Individual E18_CdV_b

"The decapitated one". Part of his burial offering consisted of an elongated handle smoker. He was found associated with the enclosed patio located on the elevated platform on one side of the pyramidal base of Complex B. The individual is male, aged between 20 and 30 years old and with an unusual height of 1.73 centimeters - the left femur itself measured 45 centimeters. Forensic analysis indicates evidence of disease, specifically ankylosing arthritis, as well as apparent postmortem decapitation (*120*).

Individual E19_CdV_b

According to the skeleton, the burial of E19_CdV_b corresponds to a female individual, 20-25 years old. It was identified as an indirect primary burial in a dorsal decubitus position (slightly to the right) flexed. This burial presented a series of alterations regarding the original position, and the hands were absent. Because there are no marks of violence or dismemberment, archaeologists suggest that the most likely hypothesis is that these elements were removed during some looting and were subsequently lost by agricultural work or by the action of rodents. The remains of this burial were found under a layer of killed pottery and associated remains of a canid (*Canis familiaris*) (*121*).

Michoacán

Eight individuals were retrieved from the state of Michoacán. Individuals E1A_Mich_b, E2_Mich_b and E4_Mich_b from the site La Mina; individual M1_Mich from the site Tanhuato and Individuals M2_Mich_b, M3_Mich_b, M4_Mich_b and M5_Mich_b from the site Zaragoza.

La Mina, Michoacán

At the end of December 2014, the INAH Michoacán Center was notified of the discovery of a series of skeletons located in the construction of a perimeter wall in the Telesencundaria #133 school, in the town of La Mina, municipality of Álvaro Obregón, Michoacán.

Upon arriving at Telesecundaria School #133, the archaeologists detected that the funerary context had been seriously altered due to the construction of a trench for the foundation of the perimeter wall. From the initial reports obtained from the masons, it was possible to determine that the bodies were concentrated in the same sector. During the first visit, towards the end of December 2014, a collection of those osteological elements extracted due to the construction work was carried out.

So far, the extension of the site of La Mina is unknown due to the lack of monumental architecture and the growth of the current settlement, which occupies much of the area with vestiges. Excavations have shown that La Mina appears to be on what was once the shores of a village founded on the southwestern shore of Lake Cuitzeo.

The absence of architectural elements on the surface, in the intervened area, and the surroundings, suggests an architecture based on perishable materials. The bioarchaeological remains were located in the northern sector of an embankment covering an area of approximately 1200 m2. The ceramic materials on the surface indicate that the settlement was occupied between 200 B.C. and 900 A.D. Although this data indicates an extensive temporal extension, the stratigraphic unit indicates a single moment of occupation. The excavations carried out at this site were only aimed at recovering the bioarchaeological materials affected by the construction work, so the archaeologists do not rule out the existence of more burials at the site.

E1A_Mich_b

Individual E1A_Mich_b was in a poor state of preservation. Only part of the lower limbs, two-thirds of the mandible, and the upper dental arcade were located. This individual corresponded to a male young adult placed in a right lateral decubitus position and was oriented in a west-east direction.

The grave goods consisted of a vessel of the Paso Ancho/graffito type placed at the level of the lumbar vertebrae and showed clear traces of use. Two reused sherds deposited in front of the face completed the offering. The intentional placement of this ceramic fragment leads us to suppose that it is an object used in the development of the funerary ritual and later placed as part of the trousseau. These two ceramic fragments were dated together with another sample from another burial using the archaeomagnetic technique. The results indicated that the burial corresponded to 647-825 CE (*122*) during the so-called Epiclassic period. A third element that made up the offering consisted of a dog in its juvenile development phase. Despite its poor state of preservation, it was possible to identify that the animal was not more than eight months old, considering the process of ossification of the epiphyses (*123*).

E2_Mich_b

Individual E2_Mich_b corresponds to an early infant who was approximately six months old. This infant did not present any artifact or other element that could be identified as part of an offering.

The infant was buried at a greater depth than the rest of the individuals, approximately 1m, even surpassing the tezontle layer that formed the leveling embankment.

E4_Mich_b

The individual E4_Mich_b corresponds to a skull collected at the site without context because it was removed by personnel building the school's perimeter fence. This skull was placed next to the burial of individual E1A_Mich_b.

Tanhuato, Michoacán

M1_Mich corresponds to the burial excavated in a rescue excavation in the town of Tanhuato, Michoacán, during the construction of a canal in a street inside the town. Tanhuato is located in the municipality of the same name in the state of Michoacán. The burial was found at 1,543 meters above sea level without a greater context, since it is located in an urbanized area. The surrounding area is almost flat and there is a large amount of cultivated land. The burial was mixed, and the inhumations were made in an enclosure dug in the soil matrix, whose bottom is one meter deep with respect to the level of the street; it covers an area of about 5 m^2. Primary and secondary burials were found in the pit and two incomplete individuals, two ossuaries and a funerary pot could be identified. Unfortunately, both bones and materials had already been removed by construction workers, so the context was very altered. Absolute dates for this individual are not yet available.

Zaragoza, Michoacán

The Zaragoza site is located in the municipality of La Piedad, on a flat-topped, isolated elevation called Mesa Acuitzio. This site presents a large rocky front that the settlers of Zaragoza took advantage of as a support for their settlement. The monumental area of the site was built on a relatively flat part located more or less halfway up the mesa, around 1,753 meters above sea level, while the living and cultivation area is distributed downwards, both to the east and north, along the slopes through artificial terraces.

Samples M2_Mich_b and M3_Mich_b come from a mixed burial, excavated in 2004 by Anyul Cuellar and Diana Bustos, called Burial 1. The space was used in successive events, that is, not all the bodies were deposited there simultaneously. It is located outside, on the east side, of an architectural unit that we have identified as a possible temascal (low heat sweat lodge), located at the southern end of one of the main group of structures. Among the buildings that make up this complex, a ball court stands out, almost 57 meters long (including the headers) and 14 meters wide.

A total of 14 individuals were located in the funerary context, identified from the number of skulls, i.e., not necessarily complete bodies were found. The remains were in a very poor state of preservation, but at least 5 primary deposits were identified. M2_Mich_b comes from individual 7, located at the western end of the pit, next to the wall that delimits the east side of the temascal. M3_Mich_b corresponds to a second molar of skull 13. Both skulls were found without anatomical connection, although they were part of the group of bodies deposited in the burial space.

M4_Mich_b and M5_Mich_b, also from the Zaragoza site, but from Burial 2, were also mixed and excavated in 2012 by Alfredo Salas. This burial is located to the south of structure 1, which is part of the same complex as the court, to the east of it. The pit, of 24 m^2, presents a total of 8 or 9 individuals, which are not specified yet. The samples correspond to a first molar of M4_Mich_b and a third molar of M5_Mich_b, respectively. Individual M4_Mich_b belongs to an individual placed in extended dorsal decubitus with a south-north axis, slightly deviated from 6 to 8 degrees to the east; it is important to note that this is the only individual with this orientation inside the pit. This individual had an associated offering composed of a cajete (ceramic bowl), a pot and a "Capiral" lid. Individual M5_Mich_b was the most complete skeleton, although in a poor state of preservation; M5_Mich_b was found in an extended dorsal decubitus position, with arms flexed and with an east-west orientation with a 287° deviation with respect to north.

Extended Results

Sample selection and sequencing strategy

We screened the DNA content of archeological samples from 39 pre-Hispanic individuals (Table S1). The central Mexico sites include three in the Sierra Gorda (SG) in Querétaro state, namely Toluquilla (TOL, n=9), Ranas (R, n=3) and a cave in Cadereyta de Montes (CCM, n=1); Cañada de la Virgen (CdV) in Guanajuato state (n=16); and the Zaragoza (n=4), Tanhuato (n=1) and La Mina (n=3) sites in Michoacán state (Mich). Furthermore, we produced additional genome-wide data for two mummies from the Sierra Tarahumara in northern Mexico (Table S1), which were sequenced at lower depths in a previous study (*78*) (Fig. 1).

For each sample, we evaluated the extent of endogenous ancient DNA (aDNA) via shotgun sequencing. Out of the 39 samples, only three had endogenous content above 10% (13 – 39%) (Table S2). We then generated additional sequence data for these three samples using shotgun sequencing, while for the remaining libraries, we performed whole-genome or mitochondrial DNA (mtDNA) capture (*124*, *125*) (Table S2). For all of these, we verified the presence of the characteristic misincorporation patterns of aDNA using mapDamage (*71*) (Fig. S1-S4). We obtained informative amounts of genome-wide data for 12 individuals (0.01 – 4.7 x). Ten were from Mesoamerica: Sierra Gorda (n=4), Cañada de la Virgen (n=4), and Michoacán (n=2); and two from the Sierra Tarahumara in Aridoamerica (Table S1). We integrated these data with previously reported low-depth genomes (0.09 – 0.3 x) from three ancient Pericúes from Aridoamerica reported in (*78*, *79*), resulting in a total of 15 pre-Hispanic individuals from Mexico for which we had genome-wide data. Our analysis also included ancient genomic data from previous studies for 21 individuals from across the Americas (Table S3)(*46*, *78*, *79*, *105–109*). Additionally, we reconstructed the mitochondrial genomes for 27 individuals across Mexico (3.2 – 1284.8x), spanning 1,600 years (320 BCE to 1,351 CE). Of these, 25 were from Mesoamerica: SG (n=13), CdV (n=7), and Michoacán (n=5); and two individuals were from Aridoamerica: Sierra Tarahumara (n=2) (*78*) (Table S1 and S2). These mitochondrial genomes were analyzed along with previously published mitochondrial genomes from Mexico and South America (*26*, *78*, *80*, *81*, *83–85*, *87–91*, *108*, *126*) (*30*, *78*, *79*) (Table S4 and S6).

Mitochondrial haplogroup structure and diversity

We observed mitochondrial DNA (mtDNA) haplogroup C to be predominant in Aridoamerica in the Sierra Tarahumara, with decreasing frequency toward Mesoamerica. This gradient has been

previously observed in ancient and contemporary populations from Mexico (*26–32*, *80*). Specifically for central Mexico, we found the presence of all four autochthonous haplogroups, A, B, C, and D, similar to previous observations on ancient and present-day populations inhabiting this region (*20*, *27*, *28*, *51*, *127*, *128*). This highlights the greater genetic diversity and heterogeneity in this particular area (*44*, *129*), and is in agreement with larger Pre-Hispanic agriculture-based civilizations and greater population mobility compared to Aridoamerica. Furthermore, we detected unreported variants in the mtDNA PhyloTree (*33*) for seven individuals assigned to the sub-haplogroups A2d (n=4), B2c (n=2), and B2l (n=1). The absence of these variants could reflect a poor sampling of Native American mitochondrial genomes in the PhyloTree database, or a loss of these haplotypes because of drift. Accordingly, we propose adding these new unreported variants to the PhyloTree database (Table S5).
In Sierra Tarahumara, we found two sub-lineages of the mtDNA haplogroup C (C 100%), as previously reported. In contrast, Central Mexico shows a higher diversity of mtDNA haplogroups. Cañada de la Virgen was the site with the highest number of sub-haplogroups, distributed in B 42.3%, A 28.6%, C 14.3%, and D 14.3%. Michoacán presented four different sub-haplogroups (A 40%, D 40%, and C 20%). Sierra Gorda also presented four different sub-haplogroups but corresponding to major lineages A 46.15%, B 46.15%, and D 7.7%, respectively.

In Sierra Gorda, out of six A2d individuals four belong to the pre-drought period according to dating based either on C14 or the archaeological context, and two to the drought period. The haplotype median-joining network for the haplogroup A2 shows a clustering of five Sierra Gorda individuals carrying the sub-haplogroup A2d (the sixth A2d individual, 333Q_TOLb, was not included in the network as it did not pass the quality filters for inclusion) despite a time transect of 1,480 years (Fig. 2A). One of the five individuals, 37A1_R_b, harbors several polymorphisms with respect to the other four. Unfortunately, this individual barely passed our filtering criteria (high presence of 'N's' and excess C > T mutations). Because haplotype networks can be biased by sequence uncertainty, most of this individual's polymorphisms are likely from sequencing errors and/or ancient damage. This makes this individual difficult to place and compare with confidence. However, the other four individuals (7A_R_b, 2417C_TOL_a, 2417J_TOL_a and P_CCM_b) form a unique clade with each other showing genetic continuity. The clade that they form is striking given their high sequence similarity, location and time span. This observation supports a continuity of this sub-haplogroup, despite the severe droughts between 900 – 1,300 CE in this region of the Northern frontier of Mesoamerica. In contrast, the haplotype network for haplogroup B, shows three pre-drought individuals sharing sub-haplogroup B2I, while the two post-drought are found in the same node representing sub-haplogroup B2c. No individuals from Sierra Gorda carried C haplogroups and only one individual (11R_R_b) had a D sub-haplogroup (D1) (Fig. S5).

We estimated the diversity of the mitochondrial genomes per pre-Hispanic site and period using the 27 mtDNA genomes, along with eight public mitochondrial genomes from Mexico (Table S4). We calculated nucleotide diversity Pi through the genetic distance method of Tajima and Nei, with the software Arlequin v. 3.5.1.3 (*77*). As expected, we find that the pre-Hispanic populations from Mesoamerica have higher Pi values (Sierra Gorda pre-drought: 0.003022, Sierra Gorda post-drought: 0.003219, Cañada de la Virgen: 0.005609, Michoacán: 0.004394) than those from Aridoamerica (Sierra Tarahumara: 0.001895, Cueva Candelaria: 0.000413 , Paquimé: 0.002118).

Past female effective population size inference

To estimate the past female effective population size, we merged our dataset of pre-Hispanic mtDNA sequences from (n=13, selected after alignments, see methods) with 164 publicly available ancient (n=8) (Table S4), modern (n=232) mtDNA sequences from Mexico and ancient from South America (n= 129) (Table S6).

When performing EBSP on the Pre-Hispanic mitogenome data, we faced a lack of resolution due to the small sample size. However, when using present-day mitogenomes from Mexico, the EBSP showed a marked population expansion ca. 5 kybp likely related to the domestication and propagation of maize cultivation in Mexico (*130*). Notably, we were not able to recapitulate with the modern mitogenome data the drastic population bottleneck experienced ca. 500 ybp after European colonization. Some possible explanations are: 1) Lack of power to infer coalescent events in recent times, 2) Need of a larger sample size, 3) European colonization affected the female lineage in a lower extent than in other regions in the Americas, 4) a combination of all the three. This is intriguing considering previous reports of massive Indigenous population decline in New Spain during European conquest (*131*, *132*). However, these estimates were based primarily on historical accounts of the early colonial Mexican population through taxation lists of most settlements across Mexico but not including all of them (*132*). Our results suggest that the demographic impact of European colonization on the female population was variable across Mexico, and less drastic than in insular and other isolated regions of the Americas where a bottleneck in ca. 500 ybp is clearly observed in the EBSPs (*54*, *56*, *81*). When using mtDNA from South America in the EBSP we could not improve the resolution at the haplogroup level, except for haplogroup C, in which we could observe a population decline ca. 5 kybp., in agreement with a previous study analyzing haplogroup C in South America.

Sex determination and Y chromosome haplogroups

We could assign a sex for 16/39 individuals (females=3, males= 13). Determination of biological sex was made using the tool reported in (*98*). This approach computes the proportion of reads mapped to the Y chromosome with respect to the reads mapping to the X chromosome and Y chromosomes (Ry). According to the method, $Ry > 0.075$ corresponds to XY, while $Ry < 0.016$ corresponds to XX. For another four individuals, although the Ry values were above the required threshold for XX assignment, they were very close (E2_Mich_b: 0.0218, E4_Mich_b: 0.022, E7_CdV_b: 0.0339 and E19_CdV_b: 0.0312). We followed the recommendation by Skoglund et al. (*98*) of restricting the sex estimation to the sequences with evidence of the characteristic aDNA damage patterns when deviations from expected Ry values are observed. Consequently, sex assignment for these individuals was corrected after applying *post-mortem* damage filters with pmdtools (*133*) using thresholds of 0 – 3 (see Table S1). After pmdtools, all these four individuals were assigned as XX (Table S1). We speculate that the whole-genome in-solution enrichment (WGE kit by Arbor Biosciences) performed in these female individuals biased the proportion of reads in the sex chromosomes, as the baits are built from male individuals (as described by the company Arbor Biosciences). Even with the low contamination estimates (<3% in mtDNA), we assume that the few contaminating Y chromosome sequences are captured and enriched when using a WGE kit that includes Y chromosome probes. Moreover, according to a previous study, captured libraries tend to be enriched in pseudo-autosomal and repetitive regions, also due to the use of DNA of male individuals for preparing the baits (*134*).

After applying the filters, a total of 20 individuals got a molecular sex assignment; nine agreed with the previous morphological assignment, one was not in agreement (individual 2417J_TOL_a), and ten corresponded to individuals for which there was not a morphological assignment due to poor preservation of the skull or youngness of the individual (Table S1).

The identification of Y-DNA haplogroups was only possible in five out of the thirteen individuals genetically assigned as males, due to the low coverage (Table S1). All individuals possessed the Native American Q lineage as expected and in agreement with previous studies on ancient and present-day Indigenous Mexican individuals (*20–23*, *80*, *128*). Specifically, we identified the sub-haplogroups Q1a2a1-L54 in the Sierra Tarahumara (n=1) and Ranas (n=1) and Q1a2a1a1a1-M3 in the Sierra Tarahumara (n=1), Toluquilla (n=1) and Cañada de la Virgen (n=1), with no apparent differences between Aridoamerica and Mesoamerica (Table S1). The information on the paternal lineage was limited, since, despite confirming the presence of the Native American chromosomal haplogroup Q in the pre-Hispanic males, there is no clear differential spatial distribution of the two identified sub-haplogroups, Q1a2a1-L54 and Q1a2a1a1a1-M3, in both Aridoamerica and Mesoamerica. Few studies have recovered information of ancient Y chromosome haplogroups across the Americas and, given the sex-biased admixture events after European colonization, the frequency of haplogroup Q in the present-day population is low. Thus, further studies on more ancient individuals from the Americas at higher coverage will likely shed light on the demography of the male lineages.

Autosomal genetic structure

All pre-Hispanic individuals from Mexico, as well as all ancient individuals from California (*79*), Belize (*108*) and Patagonia (*109*) clustered together in PCA with present-day Indigenous populations from Mexico (Fig. S7) and share similar genetic composition between themselves as evidenced in ADMIXTURE analysis (Fig. S8). The same analyses were repeated including only the pre-Hispanic individuals and the present-day Indigenous populations from Mexico (*23*, *78*, *80*, *103*) as a reference.

We applied a novel approach named “missing DNA” PCA (mdPCA) on the 15 pre-Hispanic individuals from Mexico, 12 from this study (Table S1) and the 3 Pericúes previously published (Table S3). Ancient individuals clustered closely with present-day populations from the same geographical region. Individual F9_ST_a from Aridoamerica clusters closely with present-day Rarámuri populations from north Mexico. Notably, MOM6_ST_a, also from Aridoamerica, clustered closely with present-day Indigenous populations from central-west Mexico (Purépecha and Nahuas from Jalisco). Pericúes from Baja California (*78*, *79*) cluster closely between Wixárika, and central west Mexico Nahuas from Jalisco and Purépechas. Regarding the ancient individuals from central Mexico (Sierra Gorda, Cañada de la Virgen, and Michoacán), these cluster closely with present-day Indigenous populations from this same region (Nahua from Jalisco, Purépecha, Totonac, Nahua from Puebla, and Nahua) (Fig. 3A). Both pre-, and post-drought Sierra Gorda individuals cluster together with central Mexico present-day populations, contrary to what would be expected if a population replacement from Aridoamerica in the Sierra Gorda would have taken place (Fig. 3A). A similar structure was observed with standard PCA, except that here F9_ST_a clusters closer with Wixárika than with Rarámuri (Fig S9).

We contrasted the placement of individuals in mdPCA and standard space with that obtained when applying a Temporal Factor Analysis (TFA) (*39*) to correct the ancestral relationships using the

dates of the pre-Hispanic individuals. For this, we included only four pre-Hispanic individuals with >1x genome-wide coverage (F9_ST_a, 2417Q_TOL_b, 2417J_TOL_a, and 333B_TOL_a). This analysis confirmed the closer genetic affinity of individual F9_ST_b with populations from Aridoamerica, while Sierra Gorda individuals from Toluquilla were slightly shifted towards the central-east populations in contrast with their position in the mdPCA between central-west and central-east populations. Most notably, despite the different periods of the individuals from Toluquilla, they cluster together with all three dimension reduction approaches.

In agreement, ADMIXTURE analysis (K=6) reveals that pre-Hispanic individuals from Sierra Tarahumara and Pericúes have a bigger proportion of genetic ancestry shared with present-day northern populations Akimel O'odham and Rarámuri (blue component in Fig. 3B). Pre-Hispanic individuals from central Mexico also display a similar genetic composition to present-day central-west and central-east Mexico populations (Nahua from Jalisco, Purépecha, Totonac and, Nahua from Puebla). Interestingly, individuals from Sierra Gorda show a homogeneous genetic composition independently of the period in which they lived. The genetic composition of individuals from Cañada de la Virgen and Sierra Gorda is similar, but Cañada de la Virgen individuals have a slightly higher proportion of the blue component shared with northern populations. In addition, the two individuals from Michoacán have a different proportion of the components shared with northern populations, E2_Mich_b presents the orange and blue components while E4_Mich_b presents orange and green. However, due to the low-coverage of these individuals (0.03x and 0.04x, respectively), we take this observation with caution.

Outgroup-f3 and D-statistics

We computed outgroup-f3 statistics of the form f3(Test, Source1; YRI), where the Test refers to the pre-Hispanic individual under analysis, and Source1 being a present-day Indigenous population. D-statistics were of the form D(Pop1, Pop2; Test, YRI), where Test refers to the pre-Hispanic individual under analysis, while Pop1 and Pop2 were all possible combinations of present-day Indigenous populations. For both analysis we considered only the combinations that have >1,000 overlapping Single Nucleotide Variants (SNVs), being 1,181 the lowest amount of SNVs plotted.

We found F9_ST_a to share higher genetic drift with present-day northern Mexico populations: Rarámuri, Akimel O'odham and Konkaak (0.291, 0.282, 0.277), than with any other present-day population (Fig. S11) (Table S9). Consistently, D-statistics showed F9_ST_a to be significantly more related to the present-day Rarámuri than to any other population (Pop1) when tested in the form D(Pop1, Rarámuri; F9_ST_a, YRI) (D= -0.0763 to -0.0210, $|Z| = 6.913$ to 20.281) (Fig. S12) (Table S10).

In contrast, the genetic ancestries of pre-Hispanic individuals from central Mexico are more complex. They do not seem to have a higher shared genetic drift with any present-day Indigenous population according to the f3 outgroup values, where almost all standard errors overlap. Furthermore, in D-statistics analysis, we find mostly D=0 in all possible combinations of Pop1 and Pop2 (Fig. S13 – S18) (Table S9-S10). These results are consistent with populations from central Mexico having extensive gene flow between them and not completely diverged from each other.

Conditional Nucleotide Diversity (CND) and Runs of Homozygosity (ROH)

CND, a relative measure of genetic diversity, was estimated by genetic pairwise comparison of the genomic positions that are variable in African individuals (*41*). For these comparisons we grouped the individuals per archeological site and calculated CND by pairs of individuals. This analysis included 16 pre-Hispanic individuals (Pericúes, Sierra Tarahuamara, Sierra Gorda, Cañada de la Virgen and Patagonia) with at least 10,000 SNVs overlapping between each pair of individuals (Table S11 and S12). We found similar values of CND in pairs of pre-Hispanic individuals to those observed in present-day Indigenous populations from Mexico, except for Pericúes from Baja California, who had the lowest CND values (Fig. S19) (Table S12). This is in line with Pericúes being hunter gatherers, likely with small population sizes and living in isolation in Baja California (*135*).

We calculated ROH in six pre-Hispanic individuals with the highest coverage; namely two from Aridoamerica (the Pericú B03 and F9_ST_a from Sierra Tarahumara) and four from central Mexico (three from Toluquilla and one from Cañada de la Virgen) (Table S13). Of the three Toluquilla individuals included in this analysis, one was dated pre-drought (2417Q_TOL_b), and two post-drought (2417J_TOL_a and 333B_TOL_a). Pre-Hispanic individuals display fewer segments in ROH than some present-day Indigenous individuals, especially present-day individuals from northern Mexico. The segment size distribution of ROHs suggests that the pre-Hispanic individuals studied here belonged to populations with small effective population (Ne) sizes (2Ne= 1,600-6,400) (Fig. S20), which is similar to Ne previously calculated for present-day northern (Rarámuri, Ne= 2,419; Wixárika, Ne= 2,193) and southern (Triqui, Ne= 2,407; Maya, Ne= 2,750) populations (*44*). We contrasted the CND values and ROH estimations of Pericúes from Mexico with another isolated pre-Hispanic population from Patagonian hunter gatherers (1,275 – 895 BP) (*109*), expecting similarly low levels of genetic diversity. We found that two pairs of Pericúes had similar CND levels as Patagonia individuals (~0.16-0.175), and the third pair of Pericúes presented a lower CND value (~0.14). However, in ROH analysis, we found that the four individuals from Patagonia have a higher sum of segments in ROH than the B03 Pericú individual, suggesting that the Pericú population had a bigger population size than that of South American Patagonian individuals (Fig. S20). In summary, the CND and ROH information showed that pre-Hispanic and present-day individuals from northern Mexico (Pericúes and Akimel O'odham, respectively) present the lowest CND values and the highest sum of inferred ROH >4 cM compared to other populations of their respective period. Individuals from the Sierra Tarahumara, Toluquilla and Cañada de la Virgen show similar values of genetic diversity and sum of inferred ROH segments. Furthermore, in the case of Toluquilla, where we have three individuals from different dates, we found that the CND value increases with the date difference between pairs of individuals, probably reflective of the accumulation of new mutations or genetic drift through time during the 489-680 years of difference between 2417Q_TOL_b and the individuals of a later period (2417J_TOL_a and 333B_TOL_a).

Genetic continuity from 1,100 to the present in Sierra Tarahumara, northern Mexico

We decided to evaluate the possible scenario of population continuity in the Sierra Tarahumara from northern Mexico, considering that the individual F9_ST_a was significantly more related to present-day Rarámuri than to any other indigenous population from Mexico in f3 and D statistics. We used the population continuity test described in (*40*), with the null hypothesis of population continuity tested by using read counts in the ancient individual and allele frequencies in the

reference population (Rarámuri). Using this method, we found no population continuity between F9_ST_a and present-day Rarámuri and rejected the null hypothesis with a p-value of 10^-499.3.

The northern region of Mexico was one of the most difficult to access and the last areas to be invaded by the Europeans in the seventeenth century, due to its geography and the resistance movements by the local people. Rarámuri still preserve their pre-Hispanic cultural elements today in language, music, dresses, and handcraft. We hypothesize that this genetic population discontinuity observed from this analysis can result from the changes in allele frequencies derived from the admixture events, population bottleneck, and isolation experienced by Rarámuri since pre-Hispanic times and after the European colonization (*49*).

Interestingly, Sierra Tarahumara mummy MOM6_ST_a has a higher shared genetic drift with F9_ST_a, as evidenced by outgroup-f3 tests, but mdPCA and ADMIXTURE analyses show that he has also genetic ancestry components found in central west present-day populations. Using D statistics, we found MOM6_ST_a to be equally related to all individuals from Aridoamerica and Mesoamerica, supporting the idea of gene flow between these two areas.

Ghost population contribution to pre-Hispanic Mexico

We tested the presence of UpopA in the pre-Hispanic individuals, using an admixture graph model with the same topological framework but some modifications compared to (*136*). We used Mbuti as an outgroup, Ami as the East Asian, MA1 as the ancient north Eurasian, USR1 as the ancient Beringian, Anzick-1 (*46*) and Spirit Cave (*136*) as the SNA, and Athabascan as the NNA. Interestingly, the Mummy F9_ST_a from the north and the individuals from Cañada de la Virgen from central Mexico show genetic ancestry from a 'ghost' population, at 28% and 17%, respectively (|Z-score|: 2.142 and 1.655, respectively) (Fig. S27A-B), consistent with the "ghost" UpopA genetic ancestry reported in Mixe. Models including individuals from other archaeological sites had |Z-score| values over 3 and branches with inner zeros, thus we rejected them. We repeated the test with the Mixe population to verify the contribution of the ghost population in our tested admixture graph structures, and found Mixe to have a genetic ancestry coming from an unsampled UpopA population at 13% (|Z-score|: 2.34) (Fig. S27C), which is similar to previous estimates of 11% (*136*). To confirm these results, we tested whether the ghost genetic ancestry observed in pre-Hispanic individuals was the same found in Mixe, by modeling the pre-Hispanic individuals along with Mixe, applying two different scenarios: 1) pre-Hispanic modeled in a same clade with Mixe, receiving UpopA genetic ancestry, and 2) pre-Hispanic and Mixe receiving UpopA genetic ancestry, with an extra contribution of NNA in the pre-Hispanic. Interestingly, Results show that Sierra Tarahumara and Mixe share the same ghost genetic ancestry UPopA (with F9_ST_a better modeled when having a higher contribution of NNA genetic ancestry) (Fig. S28), while in both models Cañada de la Virgen does not share the same UPopA contribution (|Z| > 3) (Fig. S29). Then we tested a third model with Mixe and Cañada de la Virgen to see whether the genetic ancestry in the latter, comes from a secondary unsampled population (UPopA2) and not the previously observed in Mixe (UPopA1). Results show Cañada de la Virgen can be modeled receiving genetic ancestry from a different unsampled population "UPopA2" suggesting the existence of another ghost population not sampled yet (Fig S29). We then modeled the two pre-Hispanic, Mixe and the two unsampled populations within the same model to confirm and summarize the results (Fig. 5). We attempted different combinations for the placement of UpopA, as well as adding and removing USR1 to confirm whether Mixe and Cañada de la Virgen could be

modeled together with only one UpopA contribution, but all these models were rejected (Fig. S30-S33).

Extended Discussion

Population continuity in Sierra Gorda

The observed population continuity in the Sierra Gorda might have been supported by the favorable climatic conditions of the northern Sierra Gorda Mountain range. The landscape could have helped maintain higher humidity conditions than in other regions of the northern frontier of Mesoamerica. In this part of the Sierra, the humid winds from the Gulf of Mexico course through. The area is characterized by coniferous and mixed forests, where the average maximum temperature is 25ºC, and the minimum temperature is 6-7ºC (*137*). This landscape is very different from the conditions found in Cañada de la Virgen and Guanajuato, where the climate is considerably more arid. Cañada de la Virgen is located north of the Lerma River, where rainfall is known to be scarce. The inhabitants of that region had to develop systems for the capture and maintenance of water needed for cultivation, which probably became impossible after the prolonged droughts. This forced them to abandon the site during 1,000-1,100 CE and migrate to areas with better climatic conditions suitable for agriculture, which was the core of their primary subsistence activity (*58*). In contrast, the main subsistence strategy in Toluquilla and Ranas was mining. Residents at these sites exploited cinnabar for trade purposes with other villages and metropolises (*57*, *61*). Cinnabar was of great sacred value to pre-Hispanic cultures. Some ancient communities used it to anoint their dead, especially in cases of highly regarded individuals, and as part of ritual offerings (*61*, *62*). We hypothesize that the cinnabar trade and the landscape of the Sierra Gorda allowed the peoples of Toluquilla and Ranas to subsist despite low rainfall conditions during the drought period as their economy did not depend strongly on agriculture.

Demography of central Mexico populations

We were also interested in obtaining insights into the demography of central Mexico populations since previous attempts of demographic modeling have been hampered by the high genetic heterogeneity observed in these populations (*44*, *52*). Our outgroup-f3 and D statistics results show that, although pre-Hispanic individuals from central Mexico share higher genetic drift with present-day populations from that region, none of these relationships are statistically significant. We thus leveraged the ancient genome data to propose an admixture model for central Mexico populations. Using qpGraph we found that the pre-Hispanic populations from central Mexico, all have different genetic ancestry sharing with the NNA and SNA branches. This was expected given previous studies on ancient genomes from Central and South America reporting multiple admixture events between these two branches after their split ca. 15,000 years ago (*49*, *79*). Together, these observations point to a scenario in which populations from central Mexico have not completely diverged from one another, possibly due to extensive gene flow as expected based on the active commercial exchange between different Mesoamerican populations for centuries (*63*). This interaction was mainly through trade routes and alliances between different nations (*63*), as observed in present-day Indigenous populations who share identity-by-descent (IBD) segments (*37*, *49*) and through the analysis of STR loci (*138*). There was also gene flow between Mesoamerican and Aridoamerican populations in pre-Hispanic times, as evidenced by the study of IBD segments in present-day Indigenous populations from Mexico (*49*). However, this gene

flow occurred less frequently than within Mesoamerica. The individual MOM6_ST_a from Aridoamerica may have been the result of such admixture between Mesoamerican and Aridoamerican ancestors. This individual has a higher shared genetic drift with F9_ST_a in outgroup-f3 tests. However, ADMIXTURE analysis show that MOM6_ST_a has genetic ancestry components found in central west present-day populations. Also, using D statistics, we found MOM6_ST_a to be equally related to all individuals from Aridoamerica and Mesoamerica, supporting the idea of gene flow between these two areas.

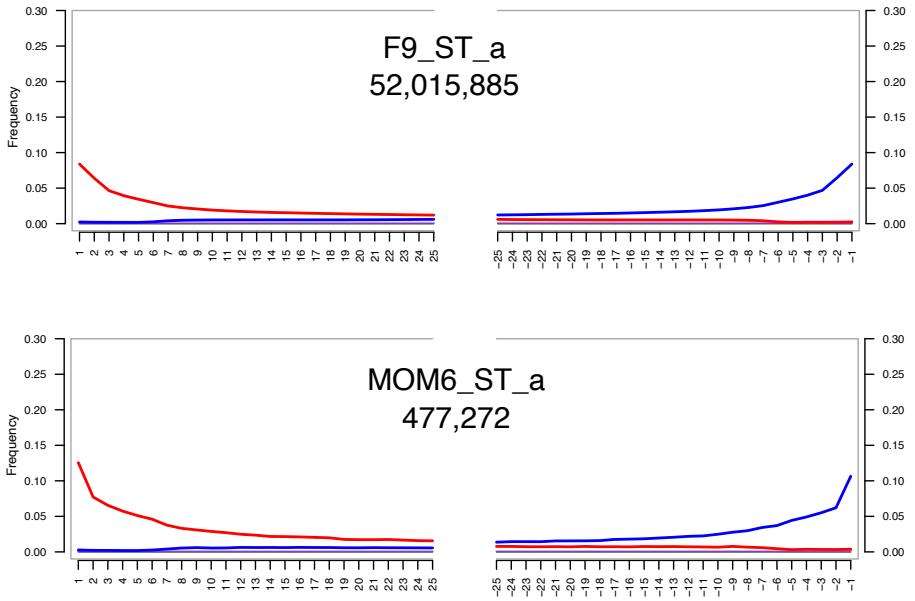

Fig S1. Damage patterns of the genomic data from Sierra Tarahumara pre-Hispanic individuals. The number of reads included in the analysis is indicated below the individual's ID. The red color indicates the frequency of cytosine to thymine changes, while the blue color indicates guanine to adenine changes. The x-axis shows the position of the base within the read, in the direction 5' to 3'. The y-axis shows the frequency of such changes.

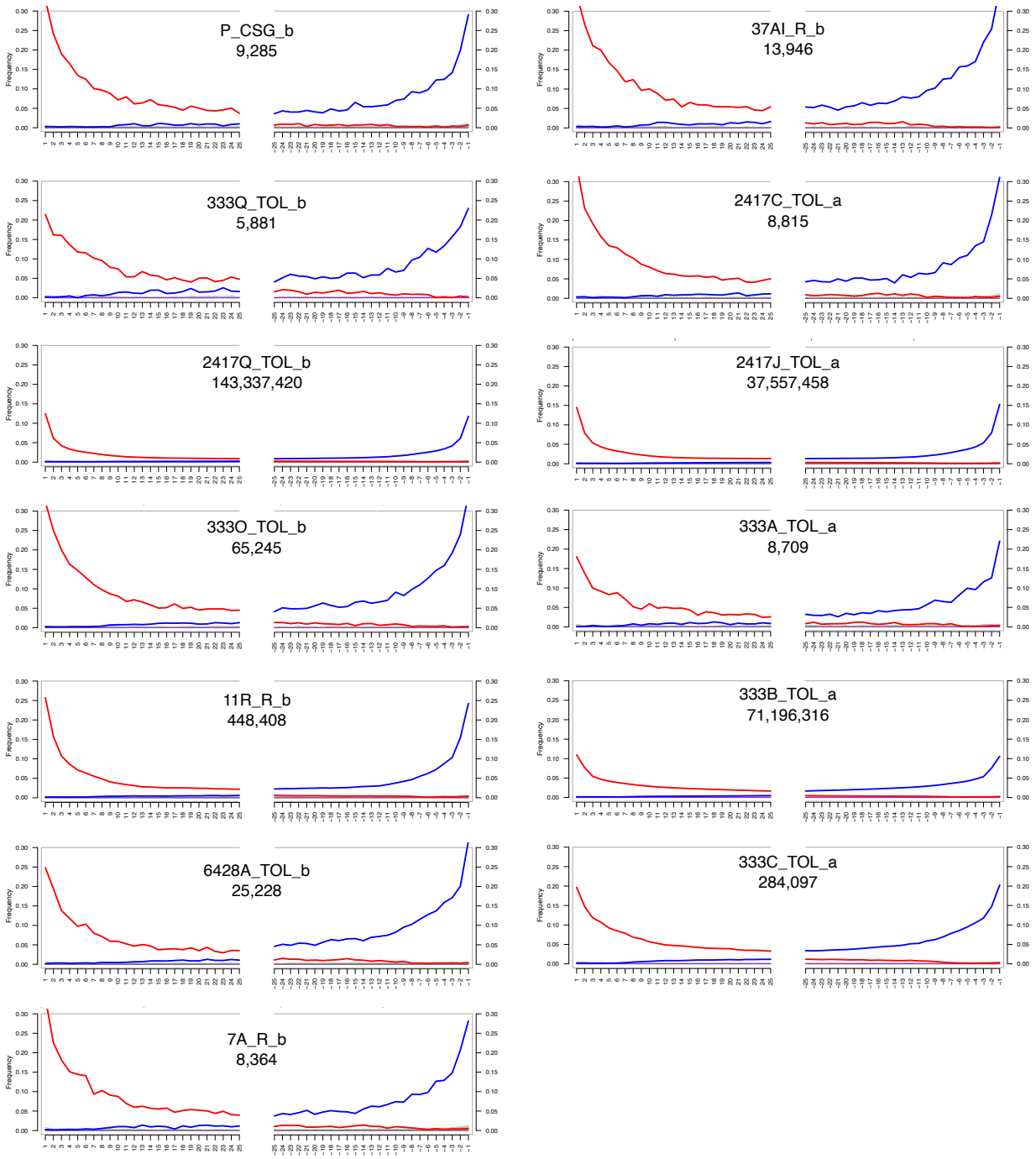

Fig S2. Damage patterns of the genomic data from Sierra Gorda pre-Hispanic individuals. The number of reads included in the analysis is indicated below the individual's ID. The red color indicates the frequency of cytosine to thymine changes, while the blue color indicates guanine to adenine changes. The x-axis shows the position of the base within the read, in the direction 5' to 3'. The y-axis shows the frequency of such changes.

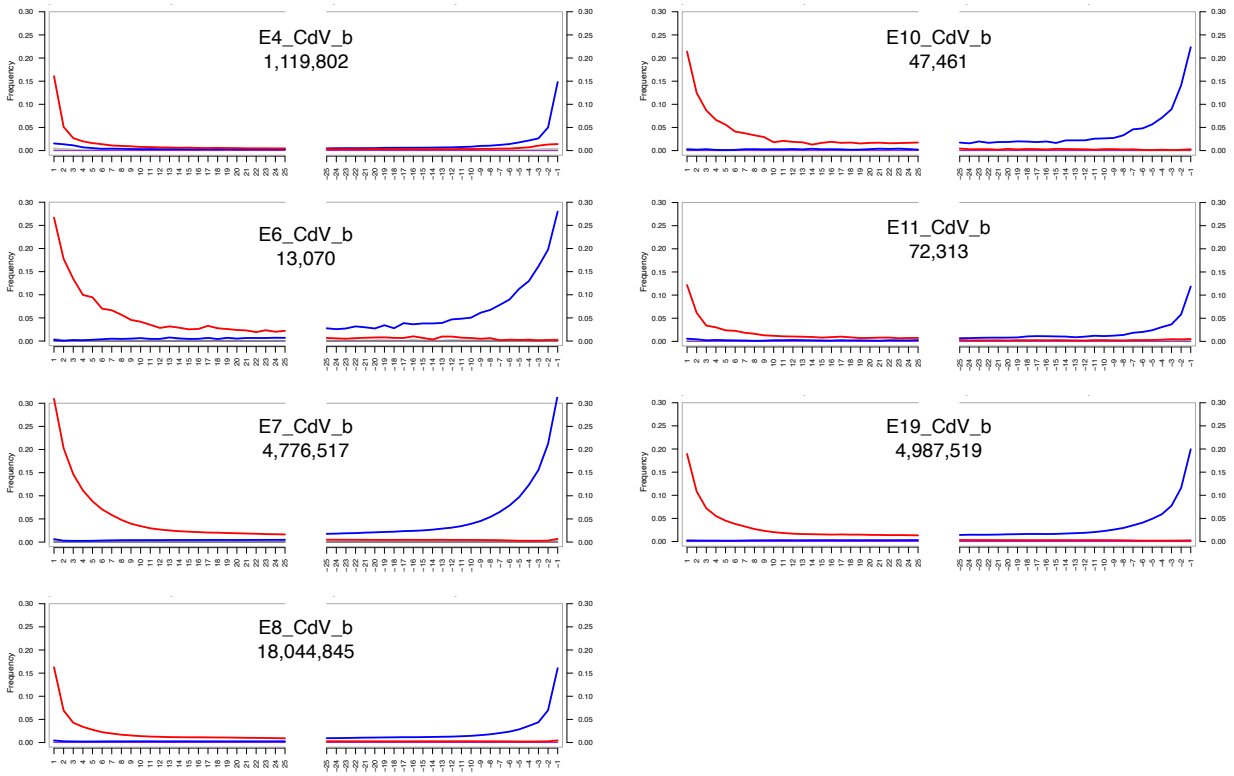

Fig S3. Damage patterns of the genomic data from Cañada de la Virgen pre-Hispanic individuals. The number of reads included in the analysis is indicated below the individual's ID. The red color indicates the frequency of cytosine to thymine changes, while the blue color indicates guanine to adenine changes. The x-axis shows the position of the base within the read, in the direction 5' to 3'. The y-axis shows the frequency of such changes.

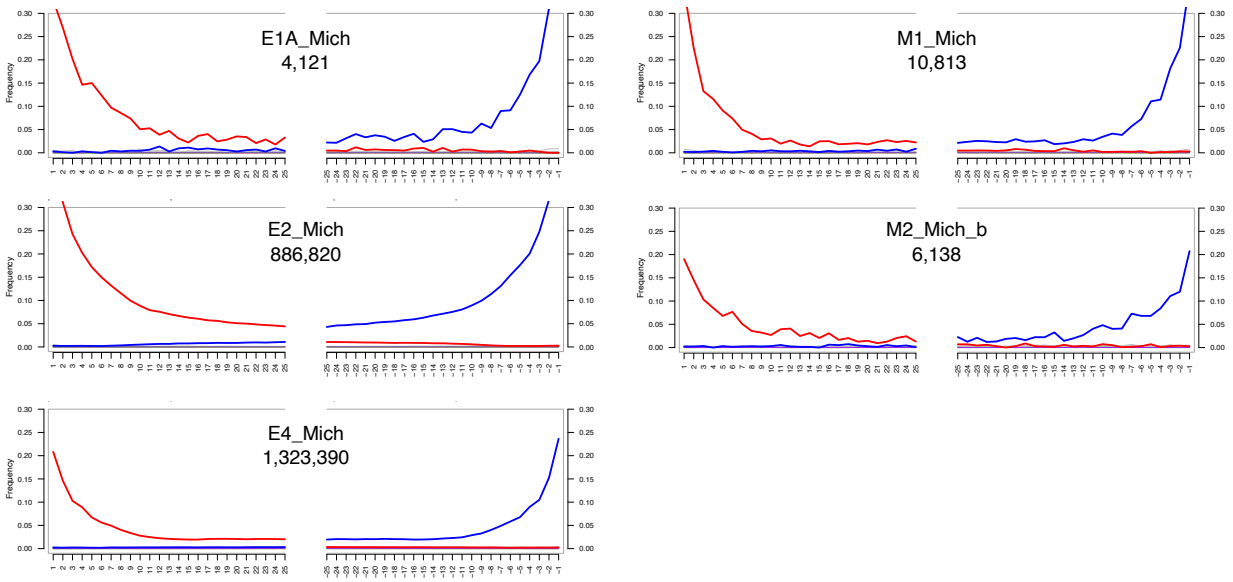

Fig S4. Damage patterns of the genomic data from Michoacán pre-Hispanic individuals. The number of reads included in the analysis is indicated below the individual's ID. The red color indicates the frequency of cytosine to thymine changes, while the blue color indicates guanine to adenine changes. The x-axis shows the position of the base within the read, in the direction 5' to 3'. The y-axis shows the frequency of such changes.

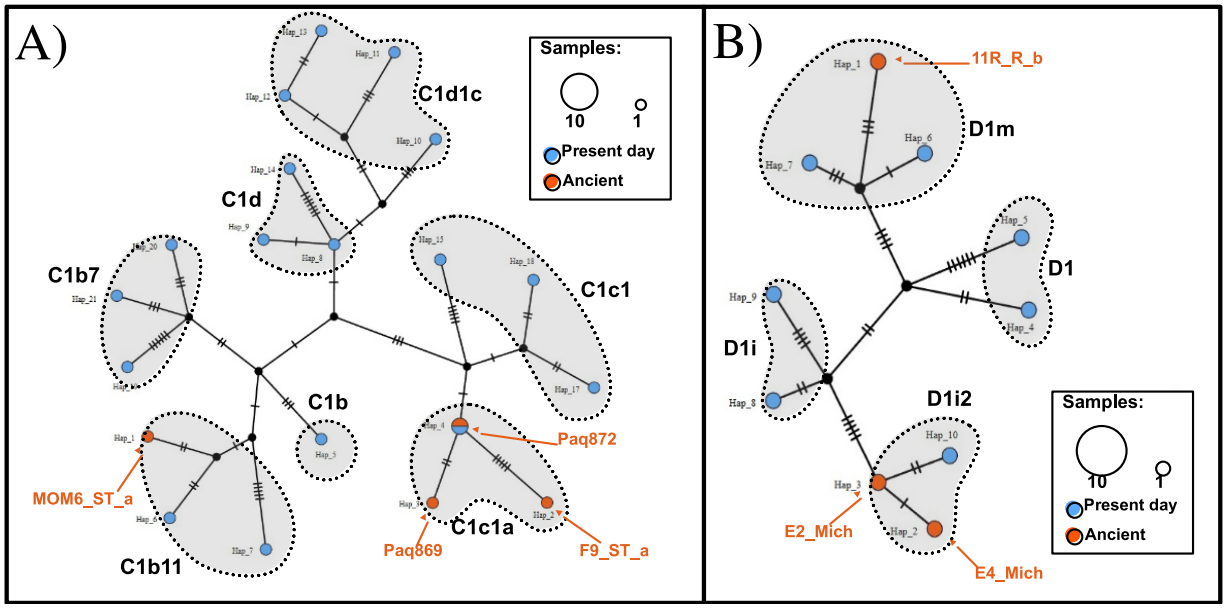

Fig S5. Haplotype Networks for mitochondrial haplogroups of pre-Hispanic individuals from Mexico. (A) Haplogroup C, (B) Haplogroup D. Sub-haplogroups are shaded with gray. Orange arrows point to ancient individuals from the present study and public available data used for the analysis (Table S4).

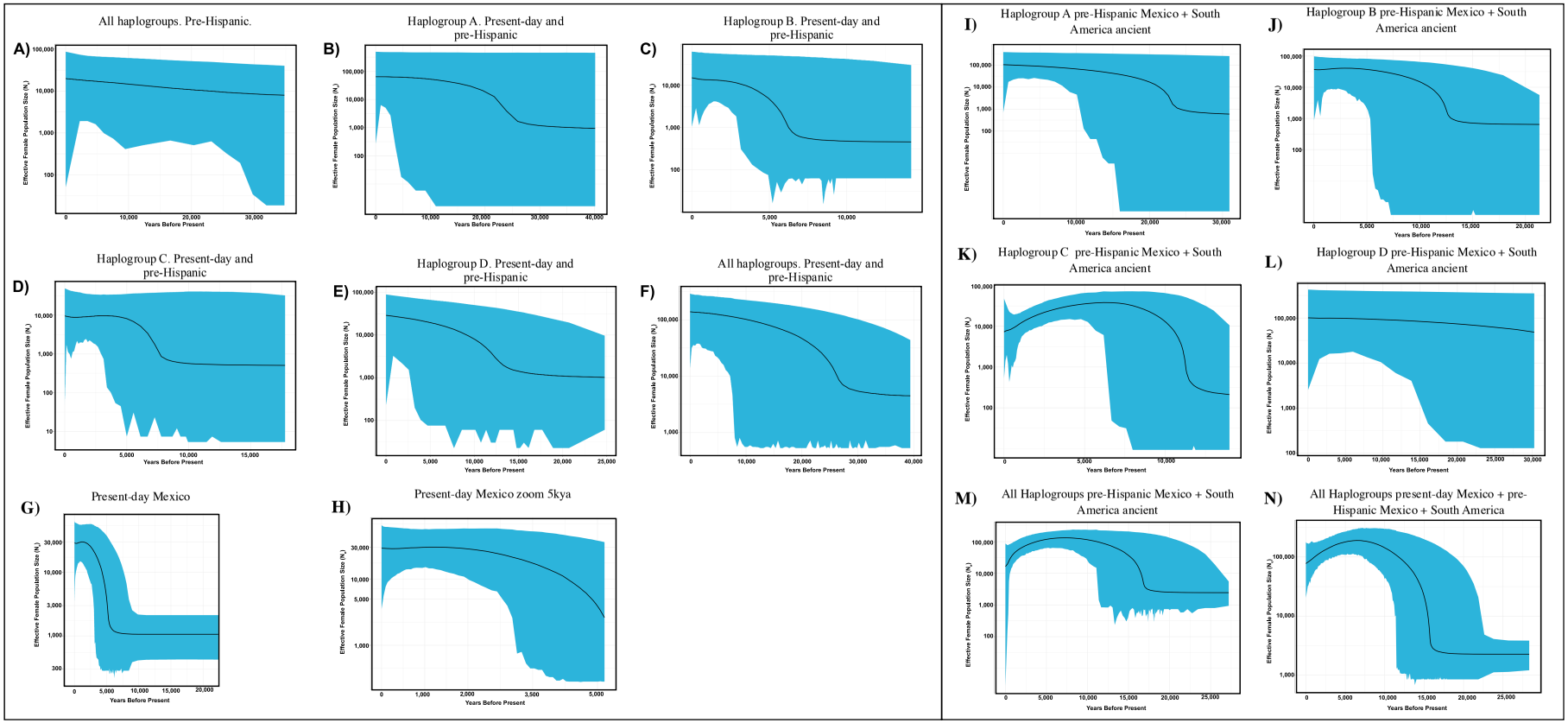

Fig S6. Inference of past effective female population size through Extended Bayesian skyline plots (EBSP). The EBSPs were constructed with ancient and present-day mitochondrial DNA. The median is shown as a black line, and the credibility interval of 95% is shown in blue.

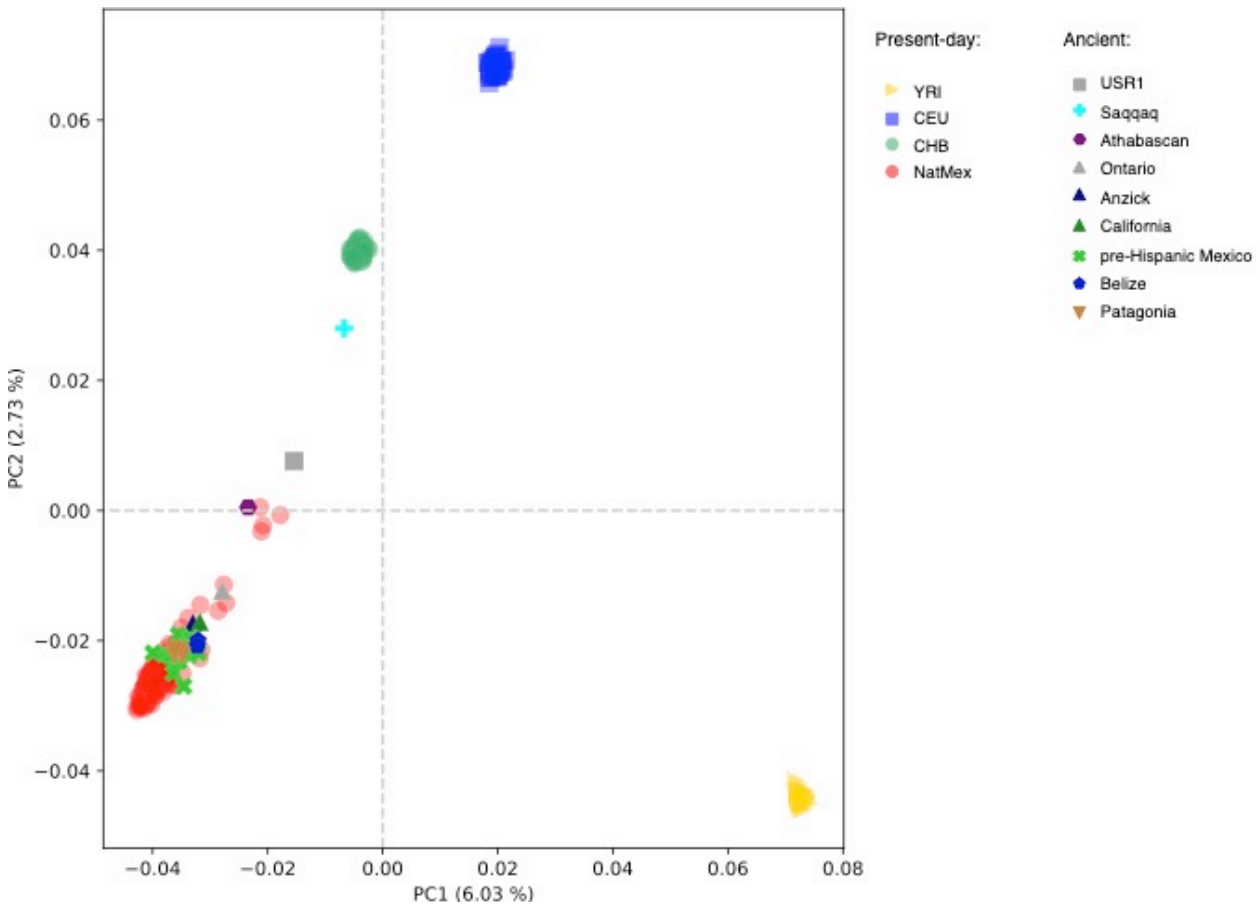

Fig. S7. Global Principal Components Analysis (PCA) including pre-Hispanic individuals from this study. The PCA includes ancient individuals Mexico from this and other studies, as well as ancient individuals from the Americas (Supplementary Table S1 and S3), Indigenous populations from Mexico (NatMex), and continental reference populations from Africa (YRI), Europe (CEU), and Asia (CHB) from 1000 Genomes.

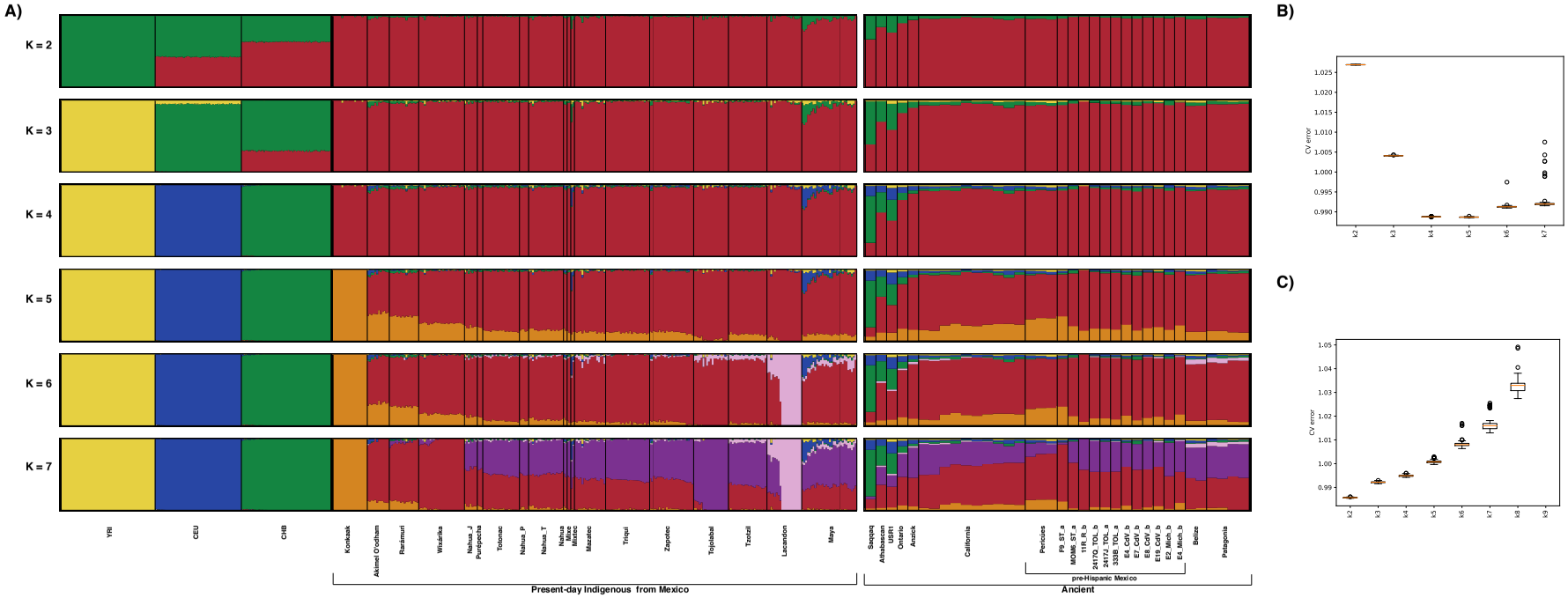

Fig. S8. ADMIXTURE analysis including pre-Hispanic individuals from Mexico. (A) Continental ADMIXTURE plot shown for K=2 through K=7. Individuals include continental references from Africa (YRI), Europe (CEU), and Asia (CHB) from 1000 Genomes as well as present-day Indigenous populations from Mexico, and ancient individuals from the Americas. Ancient populations are sorted from north to south. (B) Cross validation (cv) error values for the Continental ADMIXTURE. (C) Cross validation (cv) error values for the ADMIXTURE analysis with the reference panel including only Indigenous individuals from Mexico.

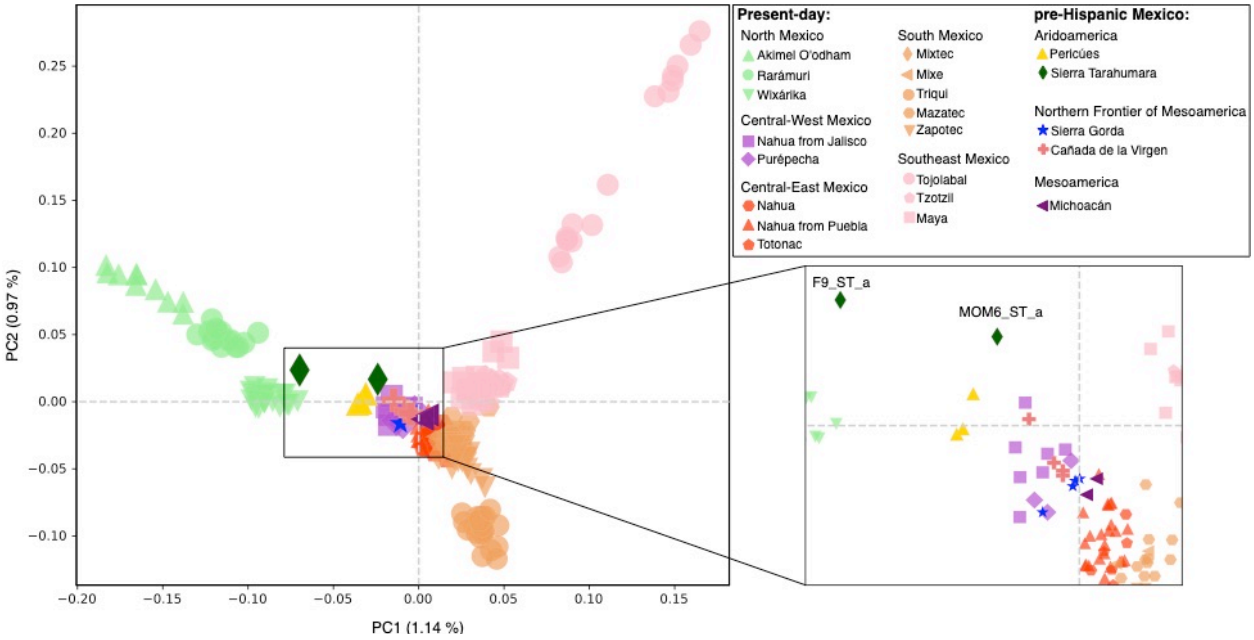

Fig. S9. PCA of populations from Mexico. The analysis includes the reference panel of present-day Indigenous Populations (see methods) and the genomic data of the pre-Hispanic individuals from Mexico. Individual F9_ST_a from Sierra Tarahumara clusters with present-day populations from northern Mexico, while individuals from Cañada de la Virgen and MOM6 cluster close to central-west present-day populations. Sierra Gorda and Michoacán individuals cluster between central-west and central-east present-day populations.

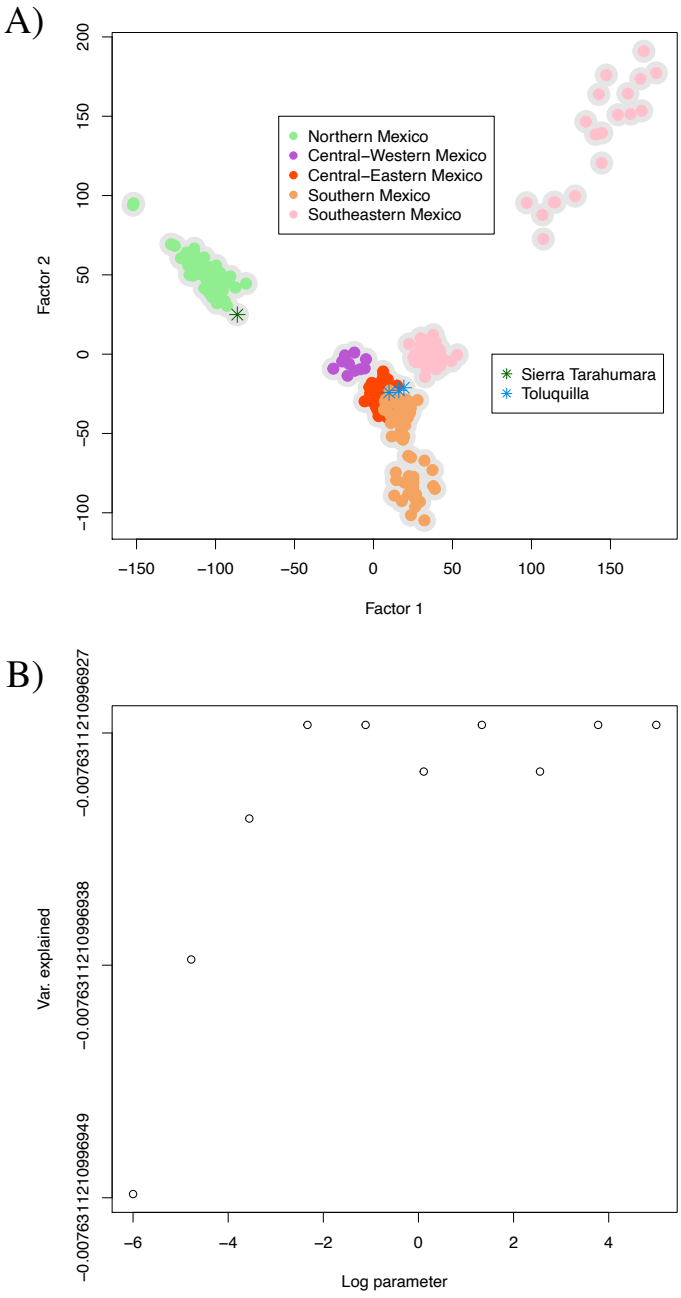

Fig. S10. Temporal factor analysis (TFA) of ancient and present-day Indigenous individuals from Mexico. (A) TFA includes the four pre-Hispanic individuals, F9_ST_a from Sierra Tarahumara and 2417Q_TOL_b, 2417J_TOL_a 333B_TOL_a from Toluquilla, Sierra Gorda. (B) Percentage of variance explained in the y-axis by different lambda values. The x-axis shows the lambda values in a log scale. A decrease in variance from right to left represents the lambda value used to correct temporality.

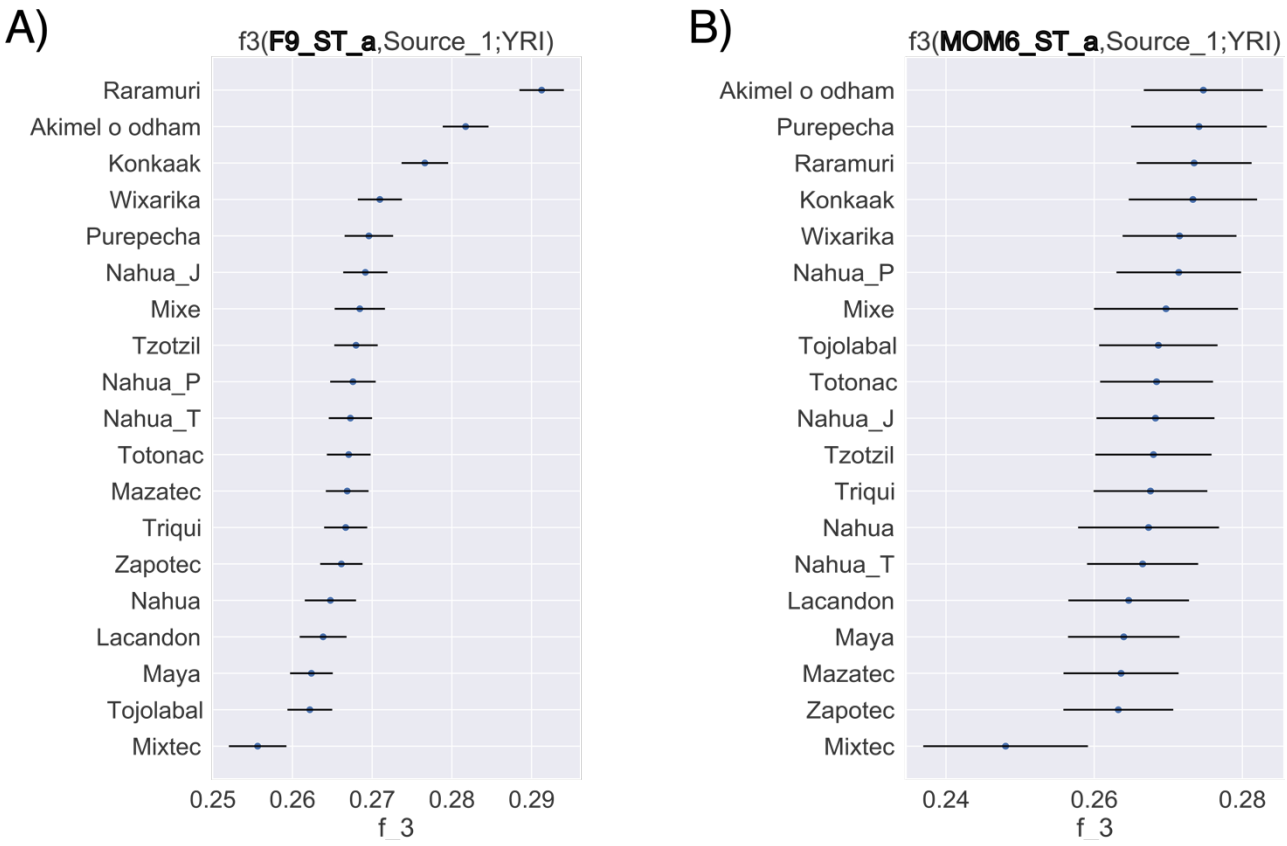

Fig. S11. Shared generic drift between Sierra Tarahumara mummies and other present-day populations of Mexico. Outgroups f3 statistics are shown for (A) Mummy F9_ST_a and (B) Mummy MOM6_ST_a. Higher values of f3 indicate higher shared genetic drift. Point estimates and one standard error are shown. The individual F9_ST_a have a higher genetic drift shared with present-day Rarámuri population from Northern Mexico. While individual MOM6_ST_a have a similar genetic drift shared with all present-day Indigenous populations from Mexico except with Mixtec.

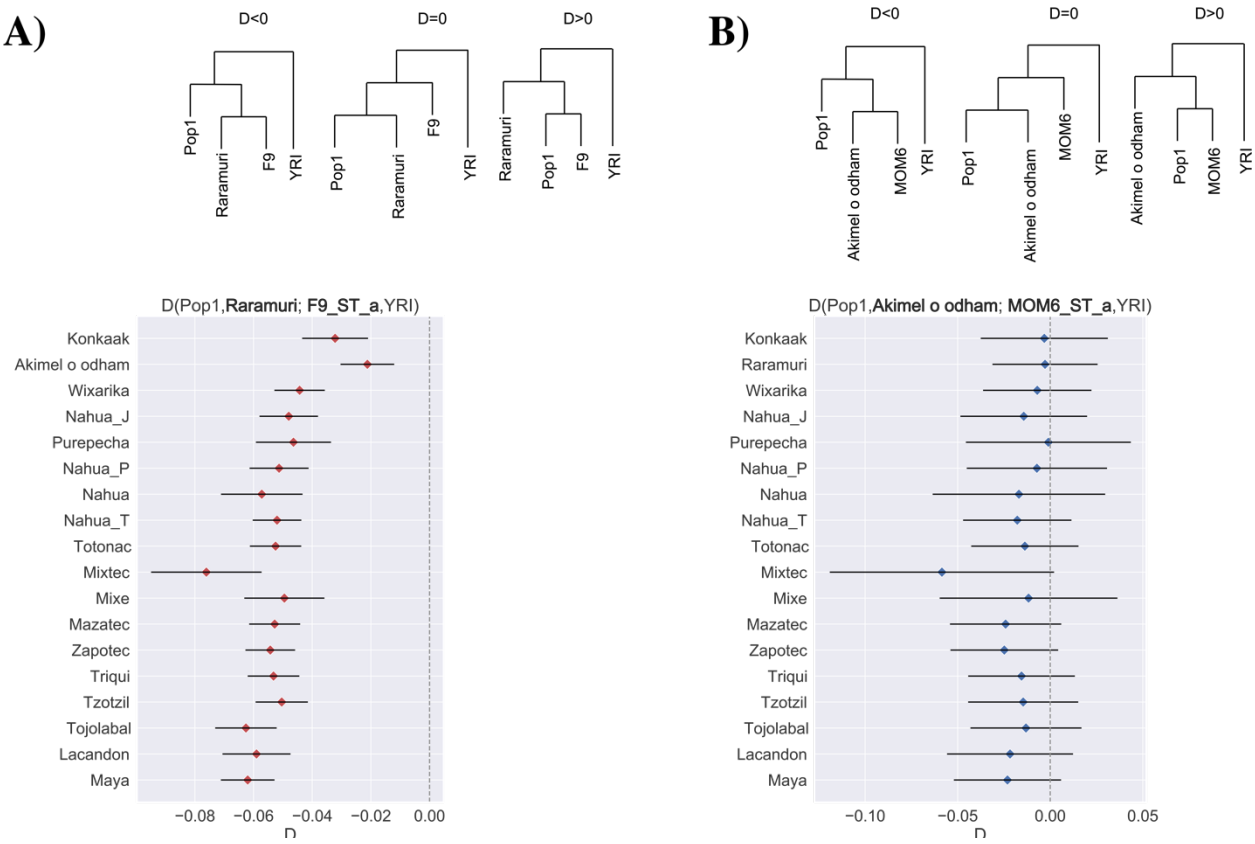

Fig. S12. D statistics for pre-Hispanic individuals from Sierra Tarahumara and other present-day populations. (A) D statistics in the form D(Pop1, Rarámuri; F9_ST_a, YRI), (B) D statistics in the form D(Pop1, Akimel o'odham; MOM6_ST_a, YRI). Expected tree topologies according to D value are drawn on the top of the plot, individual IDs in the trees are indicated with no suffixes. Red dots indicate significant deviations from D=0 (|Z|>3). The individual F9_ST_a shows a significantly higher relationship with present-day Raramuri then with any other present-day population. The individual MOM6_ST_a seems to be equally related to all present-day Indigenous populations from Mexico.

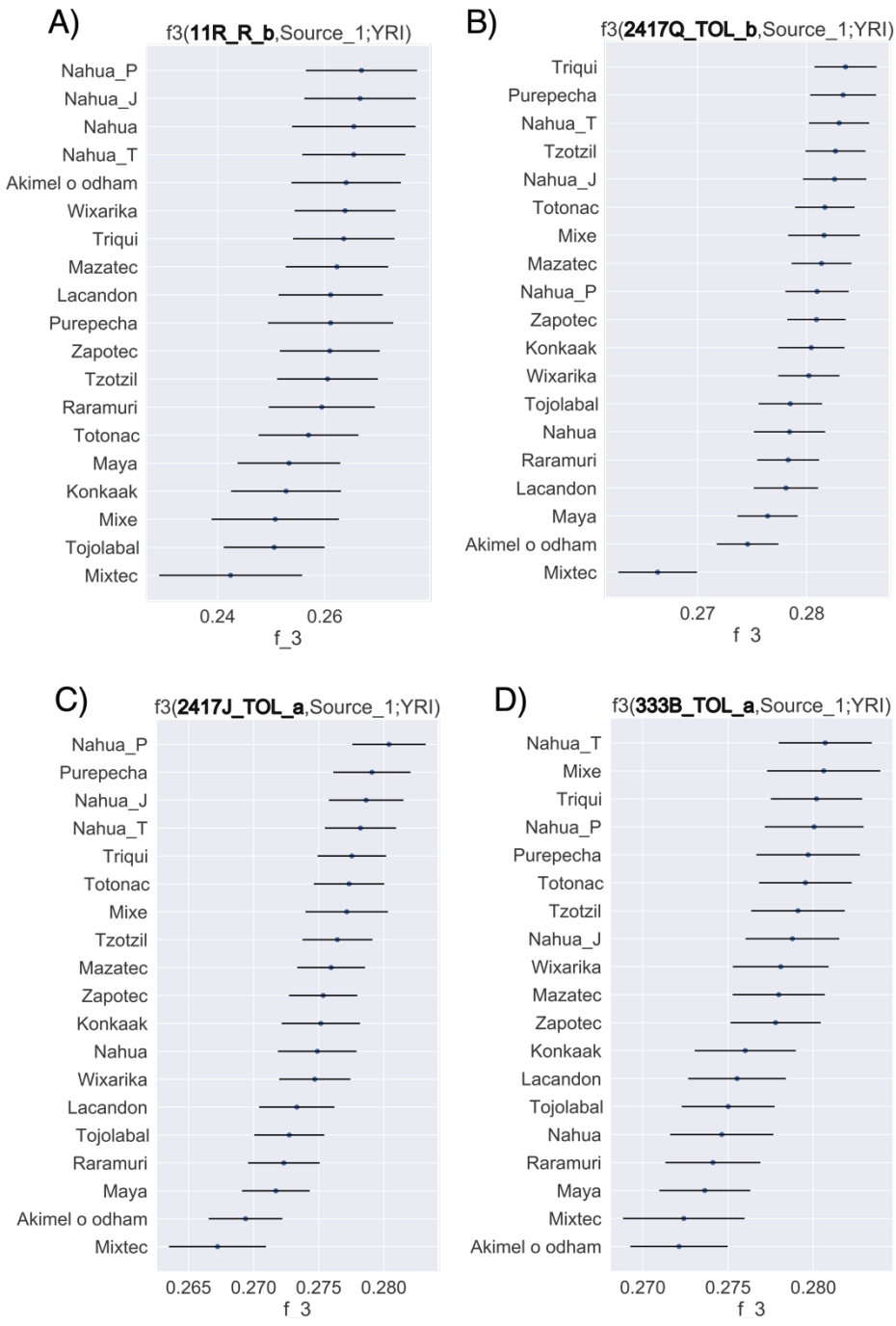

Fig. S13. Outgroup f3 statistics for pre-Hispanic individuals from Sierra Gorda and present-day Indigenous populations. (A) 11R_R_b, (B) 2417Q_TOL_b, (C) 2417J_TOL_a , and (D) 333B_TOL_a. Higher values of f3 indicate higher shared genetic drift. Point estimates and one standard error are shown. All individuals show a similar genetic drift shared with all present-day Indigenous populations from Mexico.

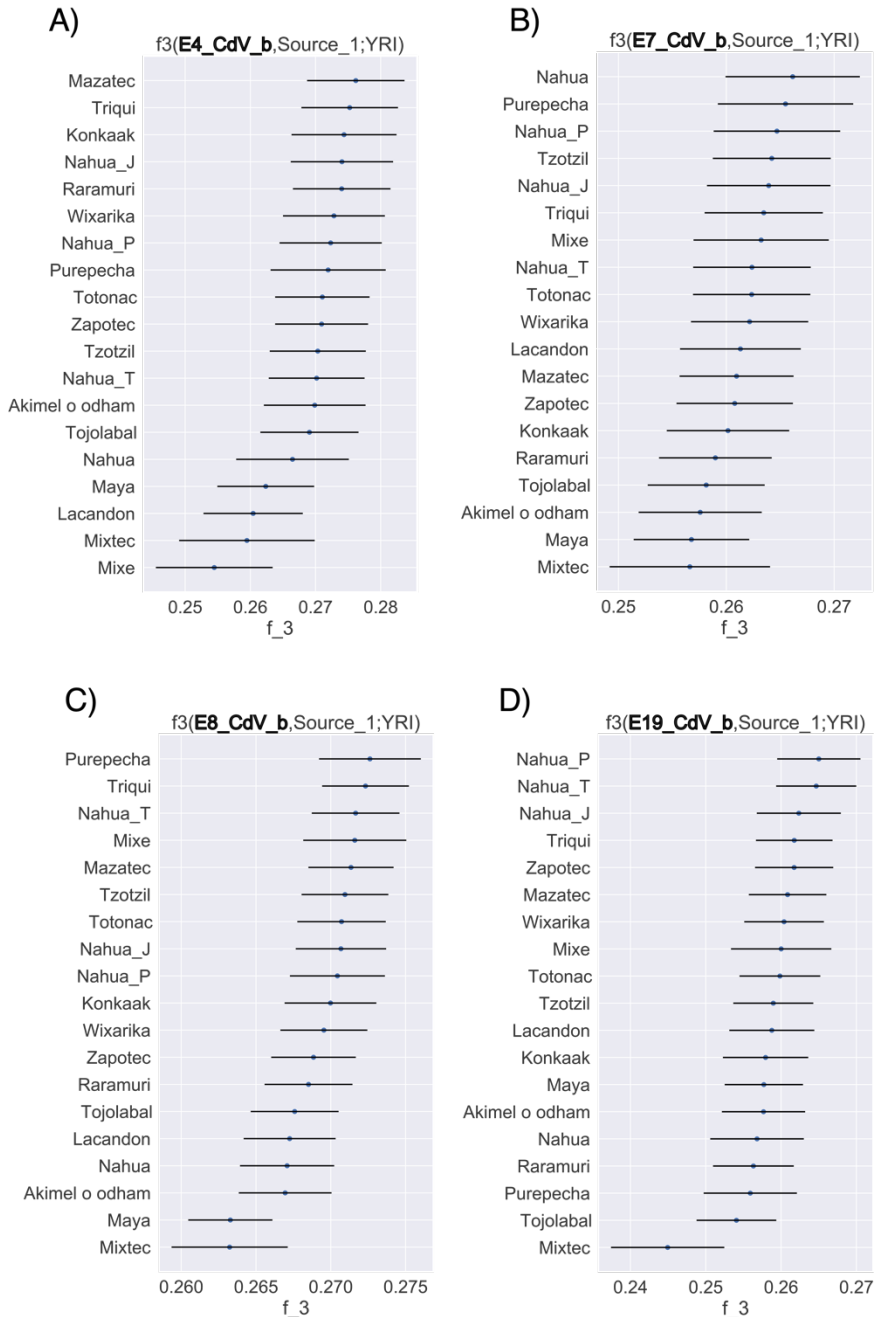

Fig. S14. Outgroups f3 statistics for pre-Hispanic individuals from Cañada de la Virgen (CdV) and present-day Indigenous populations. (A) E4_CdV_b, (B) E7_CdV_b, (C) E8_CdV_b, and (D) E19_CdV_b. Higher values of f3 indicate higher shared genetic drift. Point estimates and one standard error are shown. All pre-Hispanic CdV individuals show a similar genetic drift shared with all present-day Indigenous populations from Mexico, except with Mixtec.

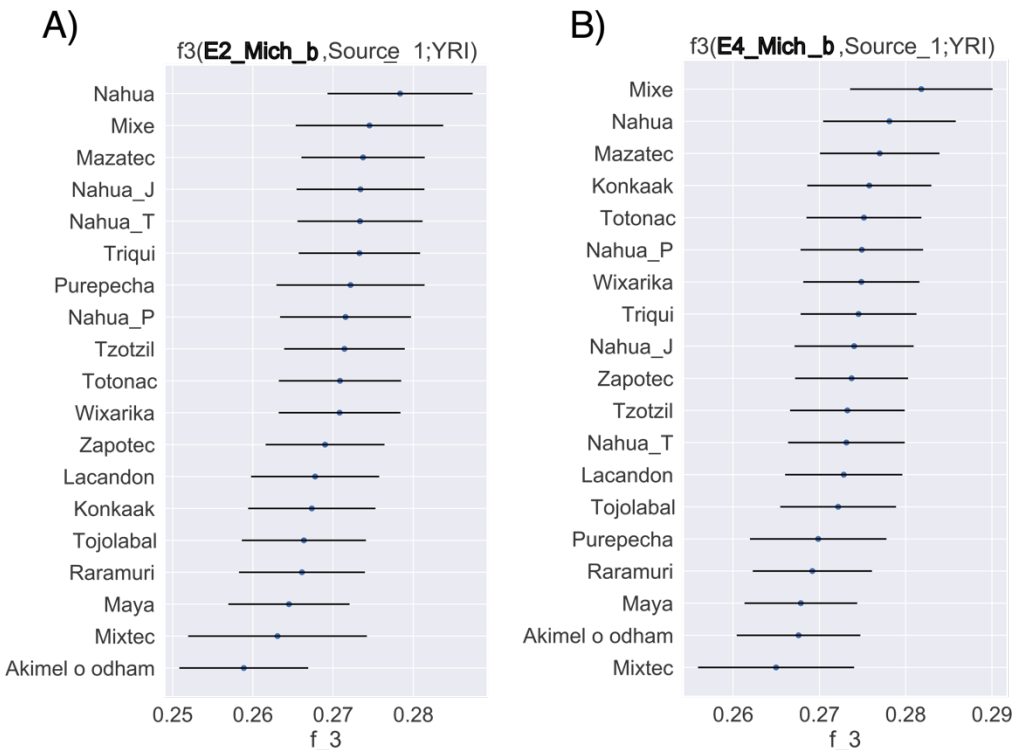

Fig. S15. Outgroups f3 statistics for pre-Hispanic individuals from Michoacán and present-day Indigenous populations. (A) E2_Mich_b, and (B) E4_Mich_b. Higher values of f3 indicate higher shared genetic drift. Point estimates and one standard error are shown. Both individuals show a similar genetic drift shared with all present-day Indigenous populations from Mexico.

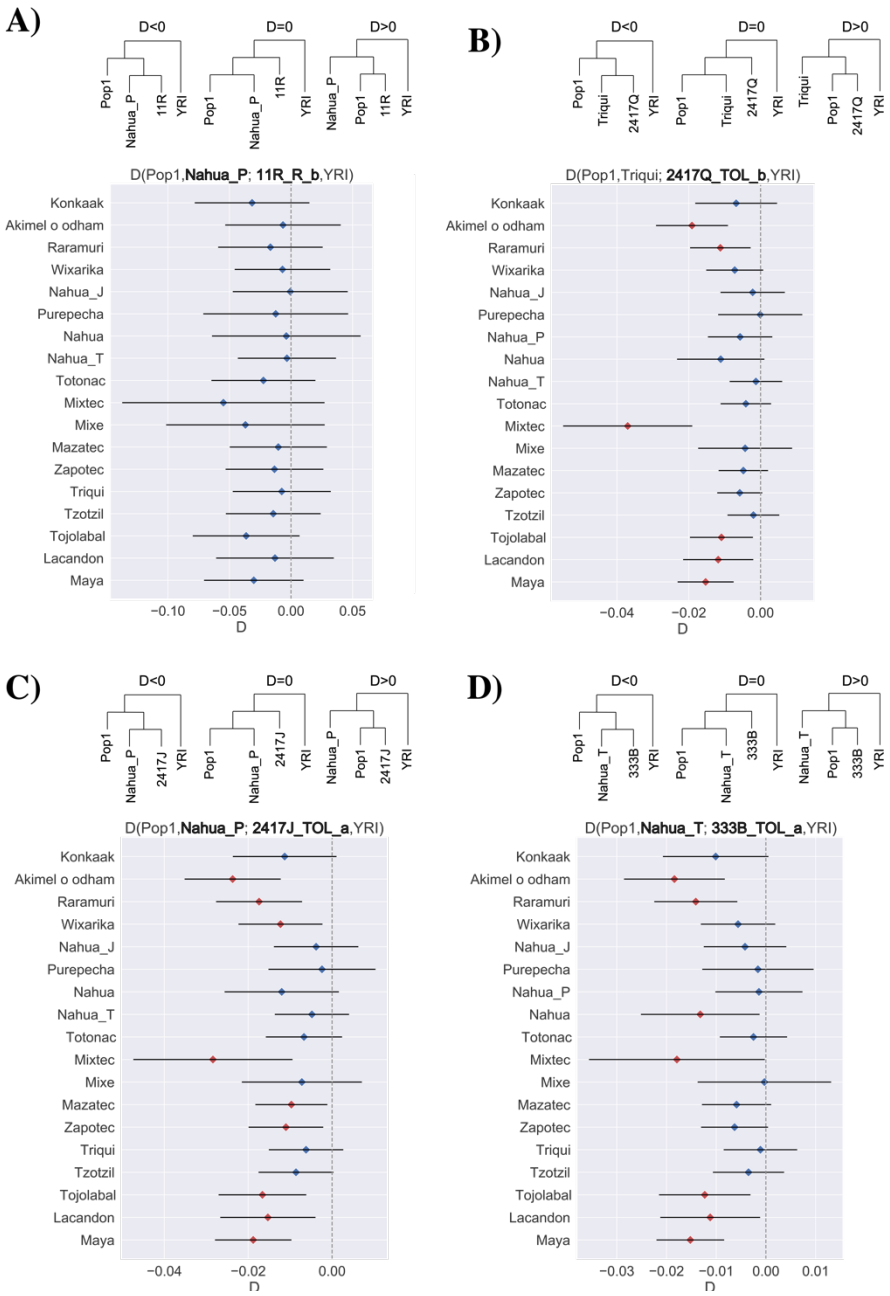

Fig. S16. D statistics for pre-Hispanic individuals from Sierra Gorda and present-day Indigenous populations. (A) D(Pop1, Nahua_P; 11R_R_b, YRI), (B) D(Pop1, Triqui; 2417Q_TOL_b, YRI), (C) D(Pop1, Nahua_P; 2417J_TOL_a, YRI) and (D) D(Pop1, Nahua_T; 333B_TOL_a, YRI) . Expected tree topologies according to D values are drawn on the top of the plot, individual IDs in the trees are indicated with no suffixes. Red dots indicate significant deviations from D=0 (|Z|>3). Individuals from Sierra Gorda tend to have a significantly higher relationship with the present-day population used as Pop2 when Pop1 is a population from Northern or Southeast Mexico, but not when they are compared with other present-day populations from Central Mexico.

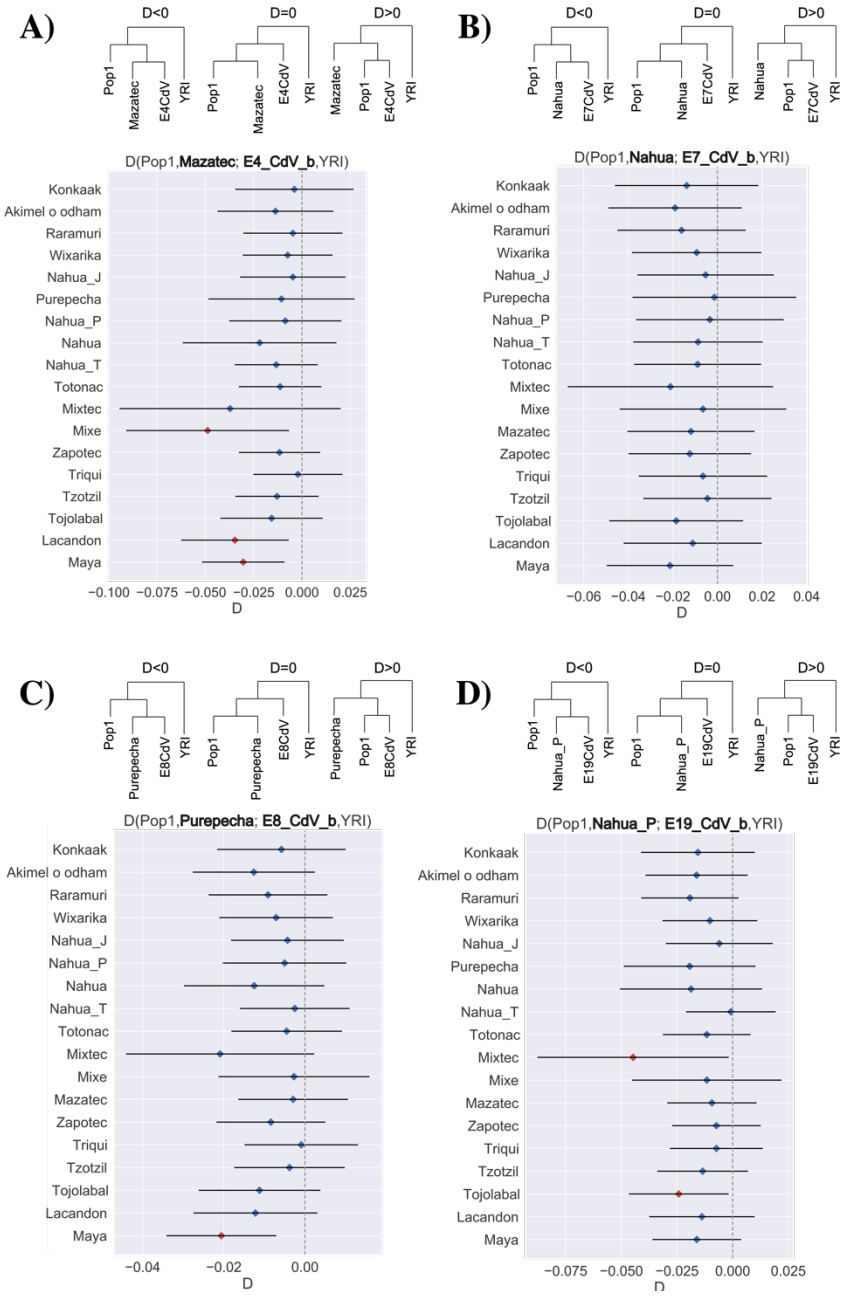

Fig. S17. D statistics for pre-Hispanic individuals from Cañada de la Virgen and present-day Indigenous populations. (A) D(Pop1, Mazatec; E4_CdV_b, YRI), (B) D(Pop1, Nahua; E7_CdV_b, YRI), (C) D(Pop1, Purepecha; E8_CdV_b, YRI) and (D) D(Pop1, Nahua_P; E19_CdV, YRI) . Expected tree topologies according to D value are drawn on the top of the plot, individual IDs in the trees are indicated with no suffixes. Red dots indicate significant deviations from D=0 (|Z|>3). Individuals from Cañada de la Virgen seem to be equally related to all present-day Indigenous populations except when Pop1 is Mixtec, Mixe or Southeastern populations, where they show a significant higher relationship with Pop2.

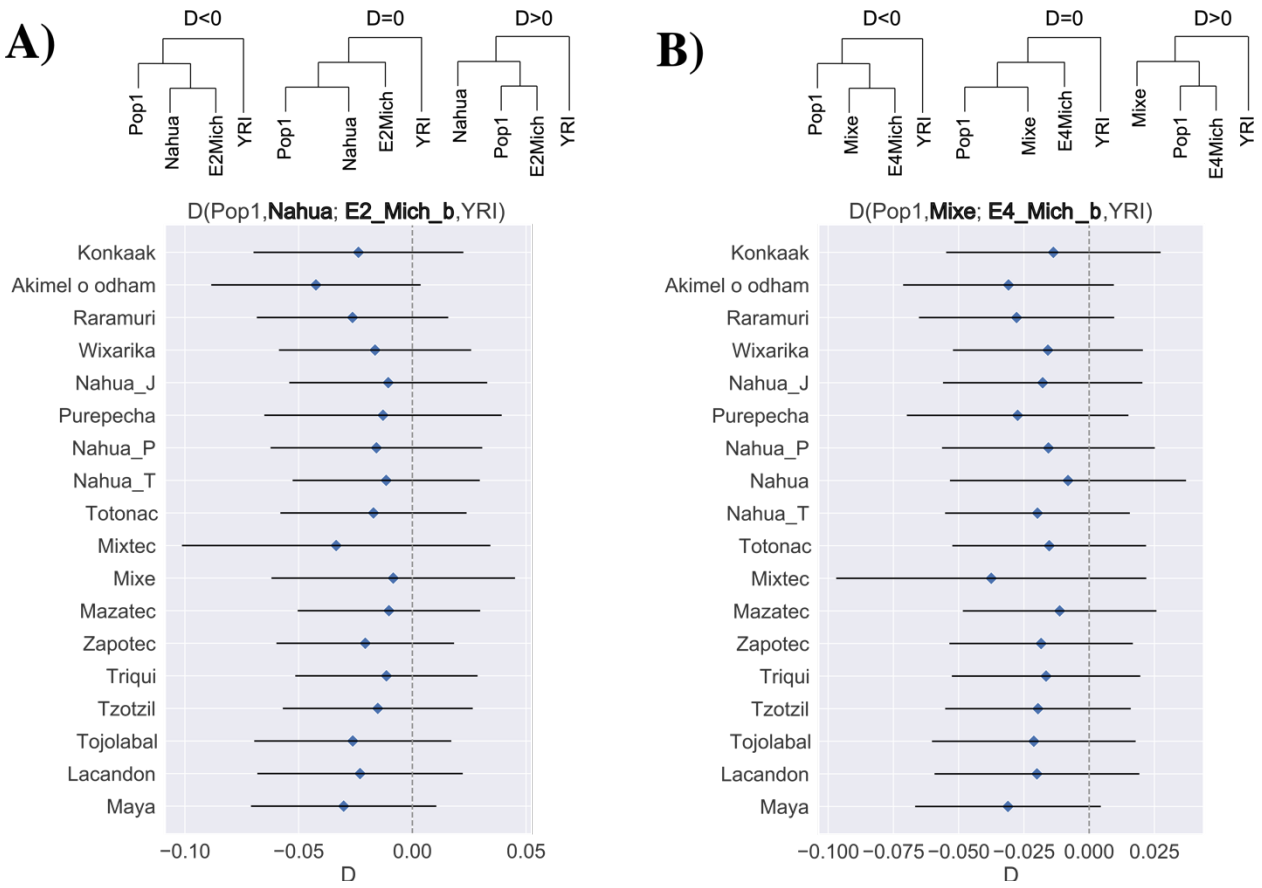

Fig. S18. D statistics for pre-Hispanic individuals from Michoacán and present-day Indigenous populations. (A) D(Pop1, Nahua; E2_Mich_b, YRI), (B) D(Pop1, Mixe; E4_Mich_b, YRI). Expected tree topologies according to D value are drawn on the top of the plot, individual IDs in the trees are indicated with no suffixes. Red dots indicate significant deviations from D=0 (|Z|>3). Individuals from Michoacán seem to be equally related to all present-day Indigenous populations.

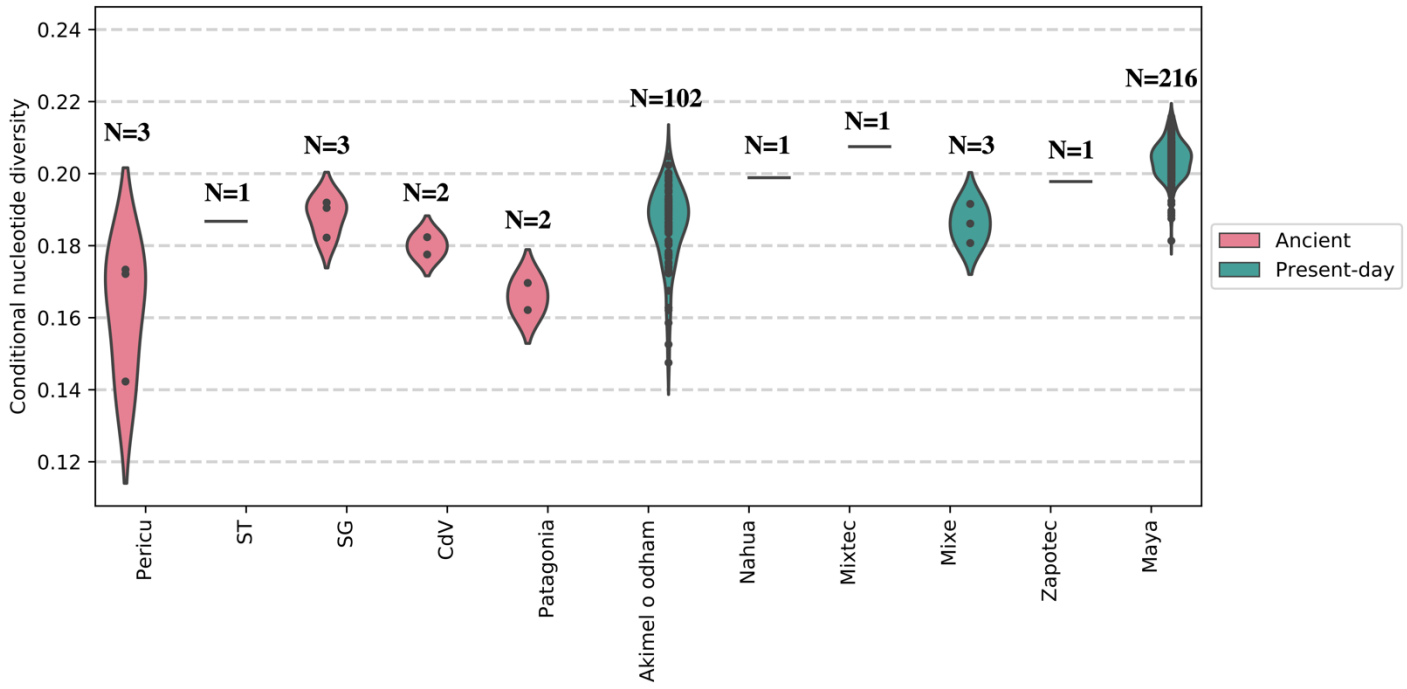

Fig. S19. Conditional Nucleotide Diversity of pairs of ancient pre-Hispanic individuals and present-day Indigenous populations from Mexico. Violin plot indicating the CND values estimated by pairs of individuals from the same archaeological site or population. The N value above each violin indicates the number of pairwise comparisons. Ancient pre-Hispanic populations are indicated in pink while present-day populations are in green. Up to each violin plot are indicated the number of comparisons (CND values estimated) for each population or archaeological site. ST (Sierra Tarahumara), SG (Sierra Gorda) and CdV (Cañada de la Virgen).

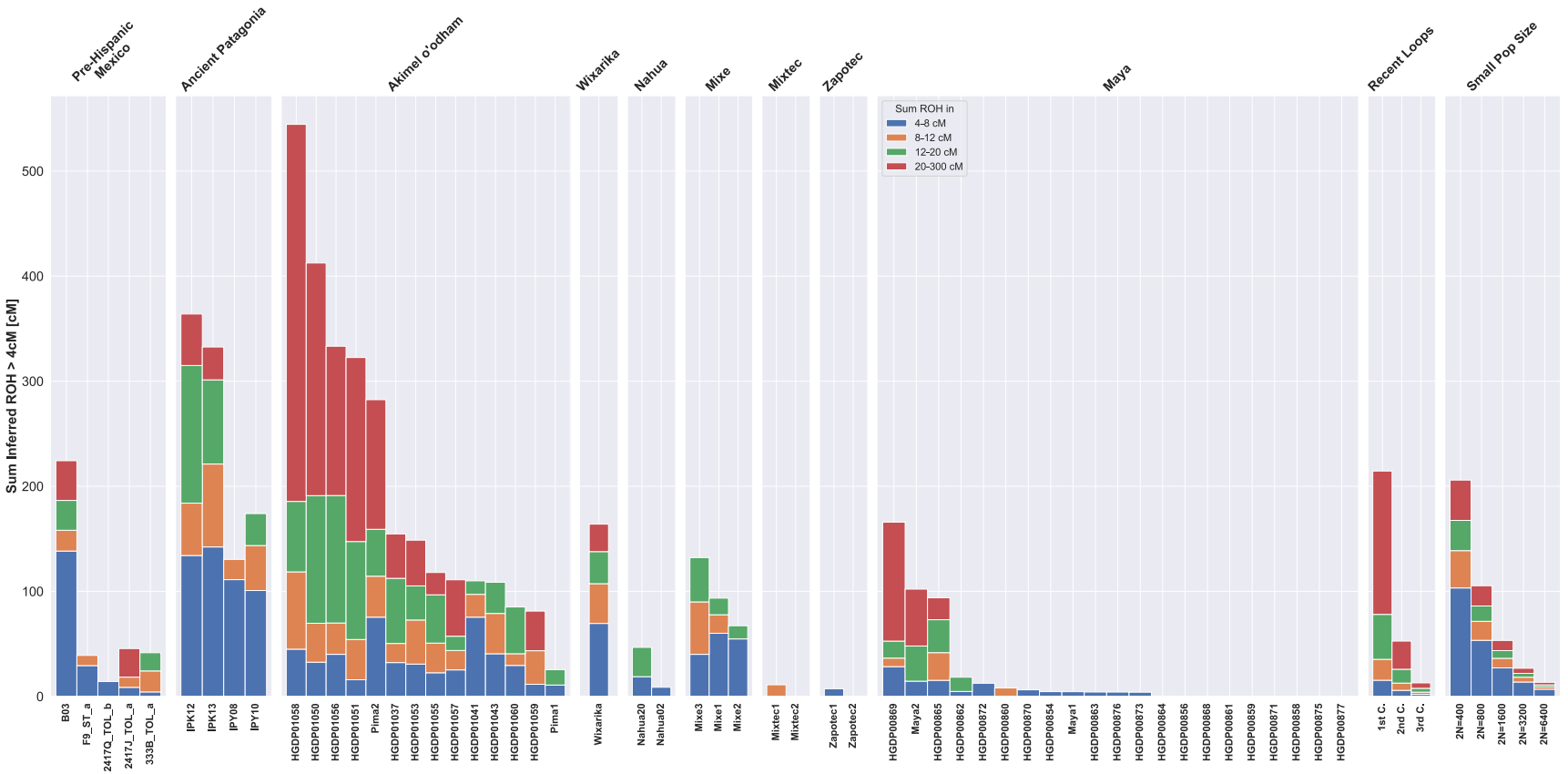

Fig. S20. Runs of homozygosity in pre-Hispanic and present-day Indigenous individuals from Mexico. Four pre-Hispanic individuals from Patagonia were analyzed for comparison. Length of ROH are shown in different colors. Pre-Hispanic individual B03 (Pericú) and the present-day Akimel O'odham population from Northern Mexico tend have the longer ROH from their time. Pre-Hispanic individuals from Patagonia have longer ROH than pre-Hispanic individuals from Mexico. Recent loops is an estimation of the ROH expected when an individual is a descendant of 1st cousins, 2nd cousins and 3rd cousins. Small Pop Size refers to the expected values in individuals from small populations.

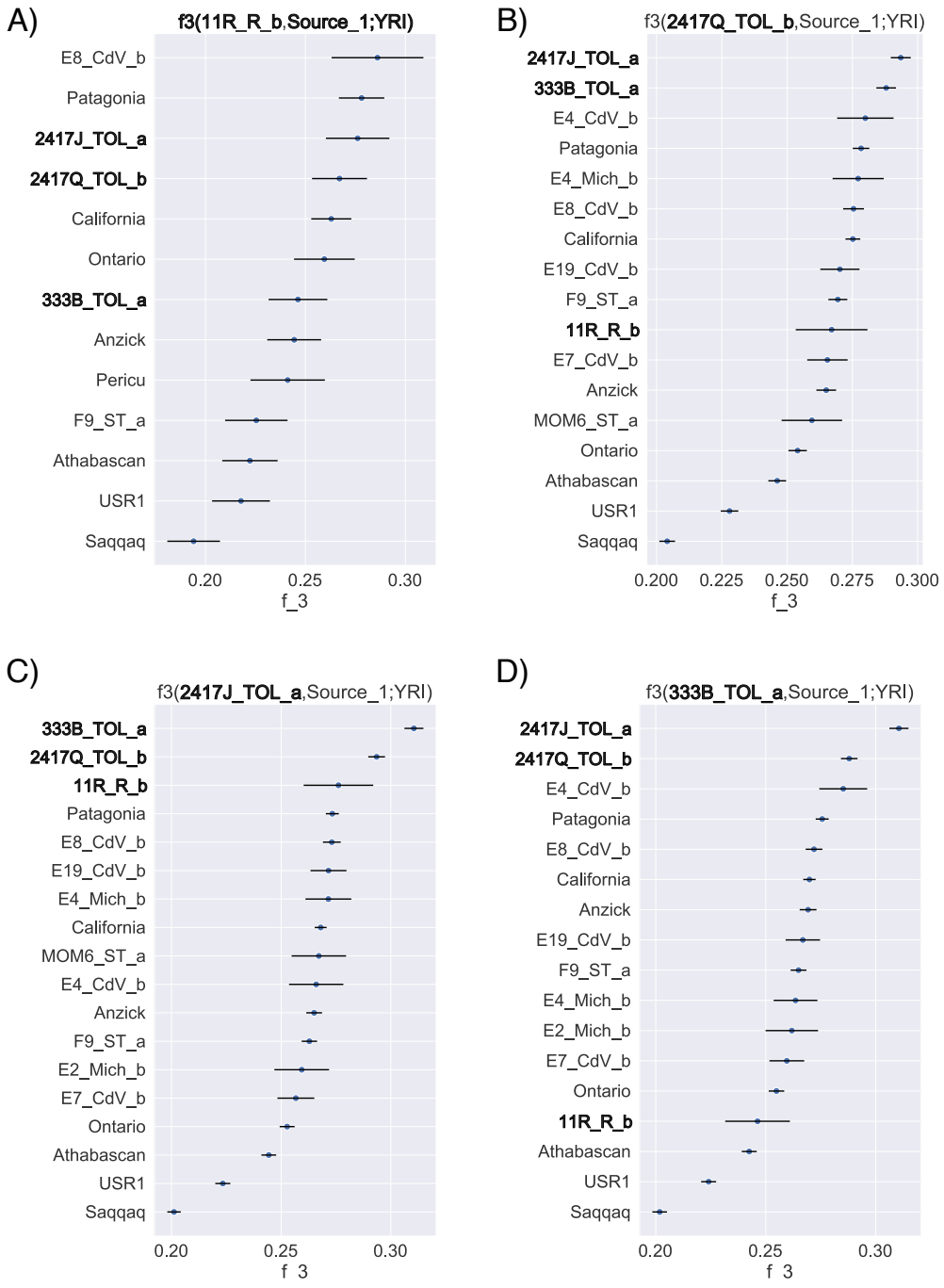

Fig. S21. Outgroup f3 statistics for pre-Hispanic individuals from Sierra Gorda. (A) 11R_R_b, (B) 2417Q_TOL_b, (C) 2417J_TOL_a, and (D) 333B_TOL_a compared with ancient individuals from Mexico and the Americas. Sierra Gorda individuals are shown in bold. Higher values of f3 indicate higher shared genetic drift. Point estimates and one standard error are shown.

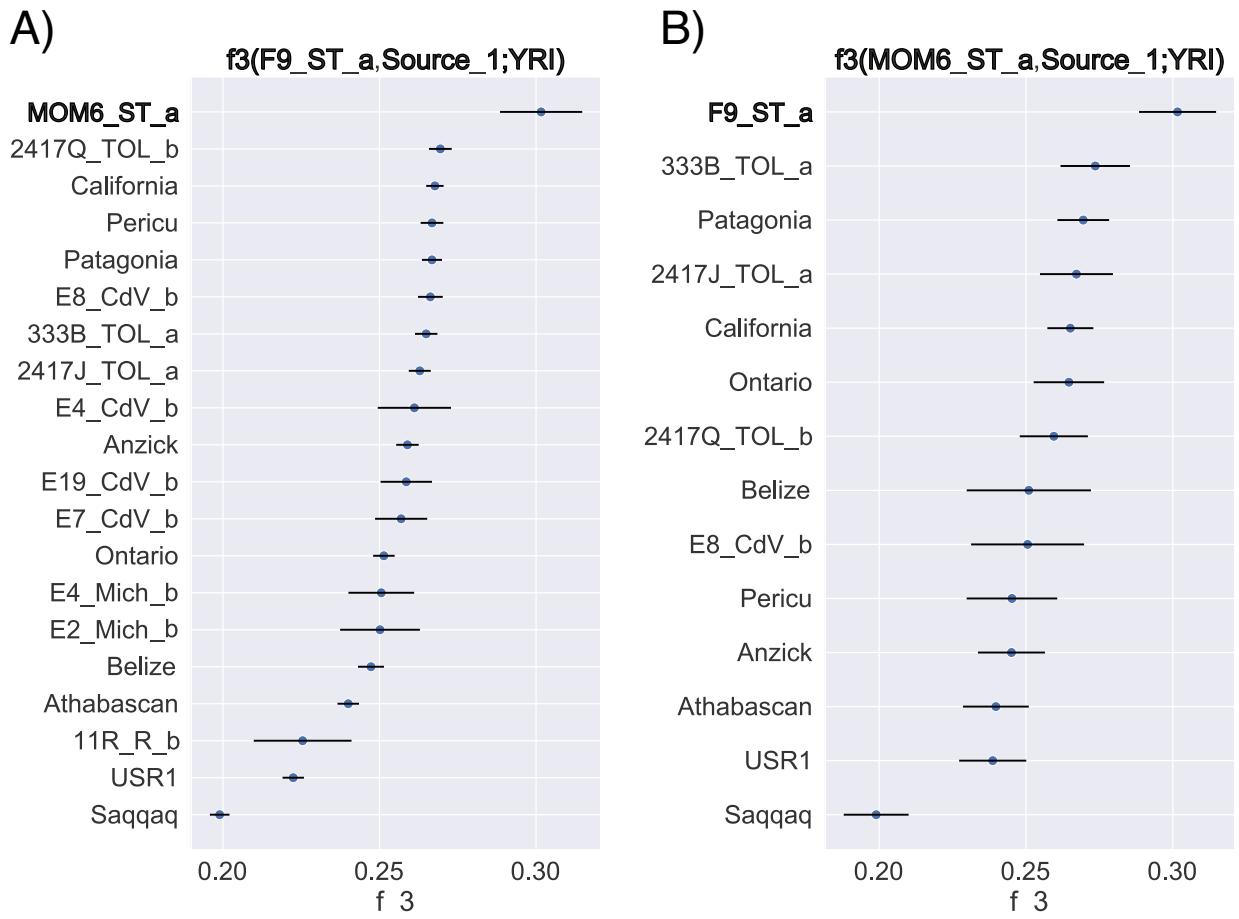

Fig. S22. Outgroup f3 statistics for pre-Hispanic individuals from Sierra Tarahumara and other ancient Native Americans. (A) Individual F9_ST_a and (B) Mummy MOM6_ST_a compared with pre-Hispanic individuals from Mexico and the Americas. Individuals from Sierra Tarahumara are highlighted in bold. Higher values of f3 indicate higher shared genetic drift.

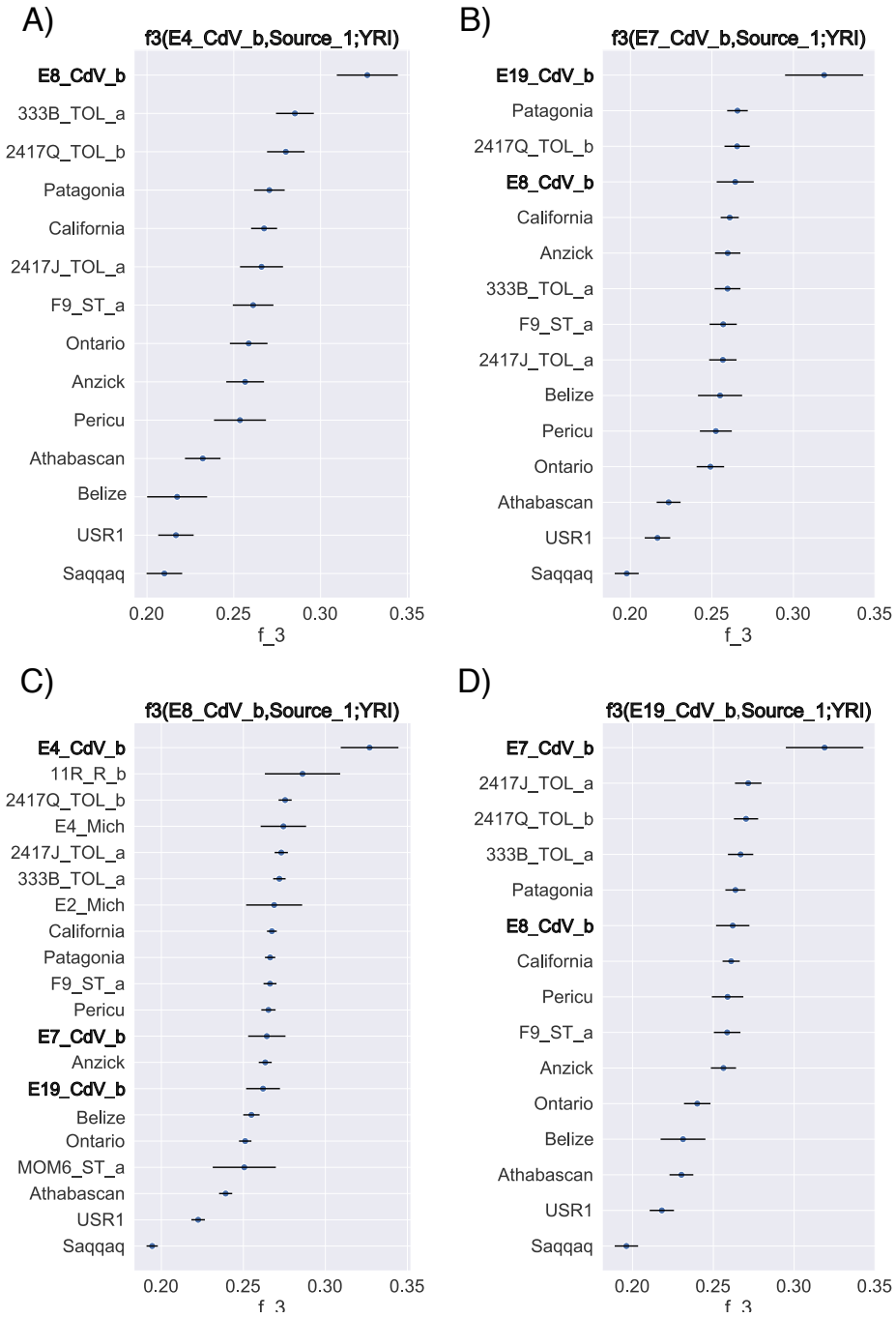

Fig. S23. Outgroup f3 statistics for pre-Hispanic individuals from Cañada de la Virgen. (A) E4_CdV_b, (B) E7_CdV_b, (C) E8_CdV_b, and (D) E19_CdV_b compared with ancient individuals from Mexico and the Americas. Individuals from Cañada de la Virgen are highlighted in bold. Higher values of f3 indicate higher shared genetic drift. Point estimates and one standard error are shown.

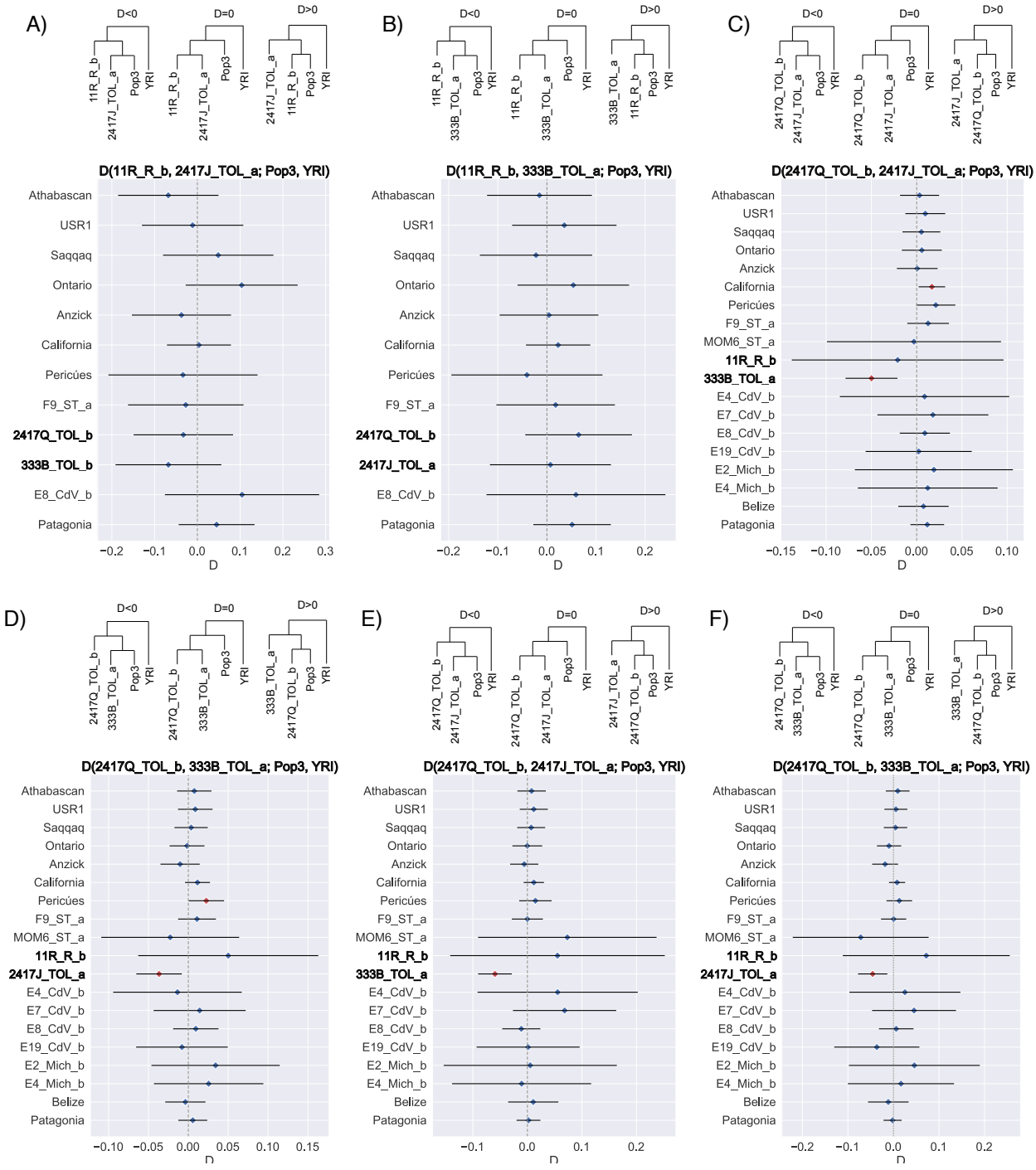

Fig. S24. **D statistic values between the pre-Hispanic individuals from Sierra Gorda in the form D (SG_b, SG_a; Pop3, YRI).** SG_b is a pre-drought pre-Hispanic individual from Sierra Gorda, SG_a is a post-drought pre-Hispanic individual from Sierra Gorda. Pop3 is any ancient individual from Mexico or the Americas and are shown in the Y axis. (A) Dstats in the form D (11R_R_b, 2417J_TOL_a; Pop3, YRI). (B) Dstats in the form D (11R_R_b, 333B_TOL_a; Pop3, YRI). (C) Dstats in the form D (2417Q_TOL_b, 2417J_TOL_a; Pop3, YRI). (D) Dstats in the

form D (2417Q_TOL_b, 333B_TOL_a; Pop3, YRI). (E) Dstats in the form D (2417Q_TOL_b, 2417J_TOL_a; Pop3, YRI), transversions only. (F) Dstats in the form D (2417Q_TOL_b, 333B_TOL_a; Pop3, YRI), transversions only. Expected tree topologies according to D value are drawn on the top of the plot. Red dots indicate significant deviations from D=0 (|Z|>3). Individuals from Sierra Gorda are highlighted in bold.

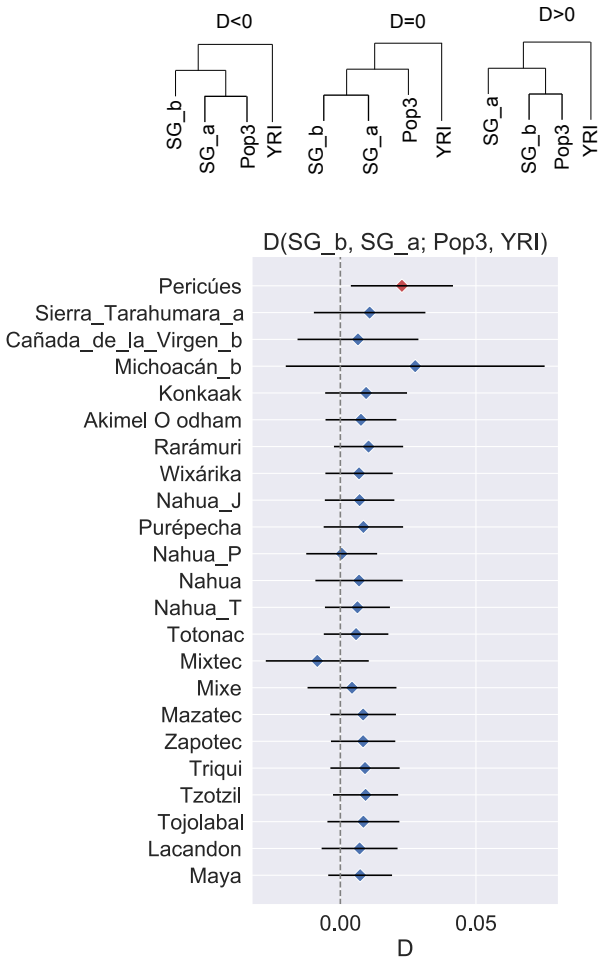

Fig. S25. D statistic values between the pre-Hispanic individuals from Sierra Gorda and pre-Hispanic/present-day Indigenous populations from Mexico. D statistics in the form D(SG_b, SG_a; Pop3,YRI). SG_b includes the pre-Hispanic individuals 2417Q_TOL_b and 11R_R_b, while SG_a include the pre-Hispanic individuals 2417J_TOL_b and 333B_TOL_a, Pop3 is any of the other pre-Hispanic or present-day populations from Mexico and is shown in the Y axis. YRI as outgroup corresponds to Yoruba from Africa from the 1000 Genomes Project (*38*, *101*). Expected tree topologies according to different D values are drawn on the top of the plot. Red dot indicates significant deviation from D=0 (|Z|>3).

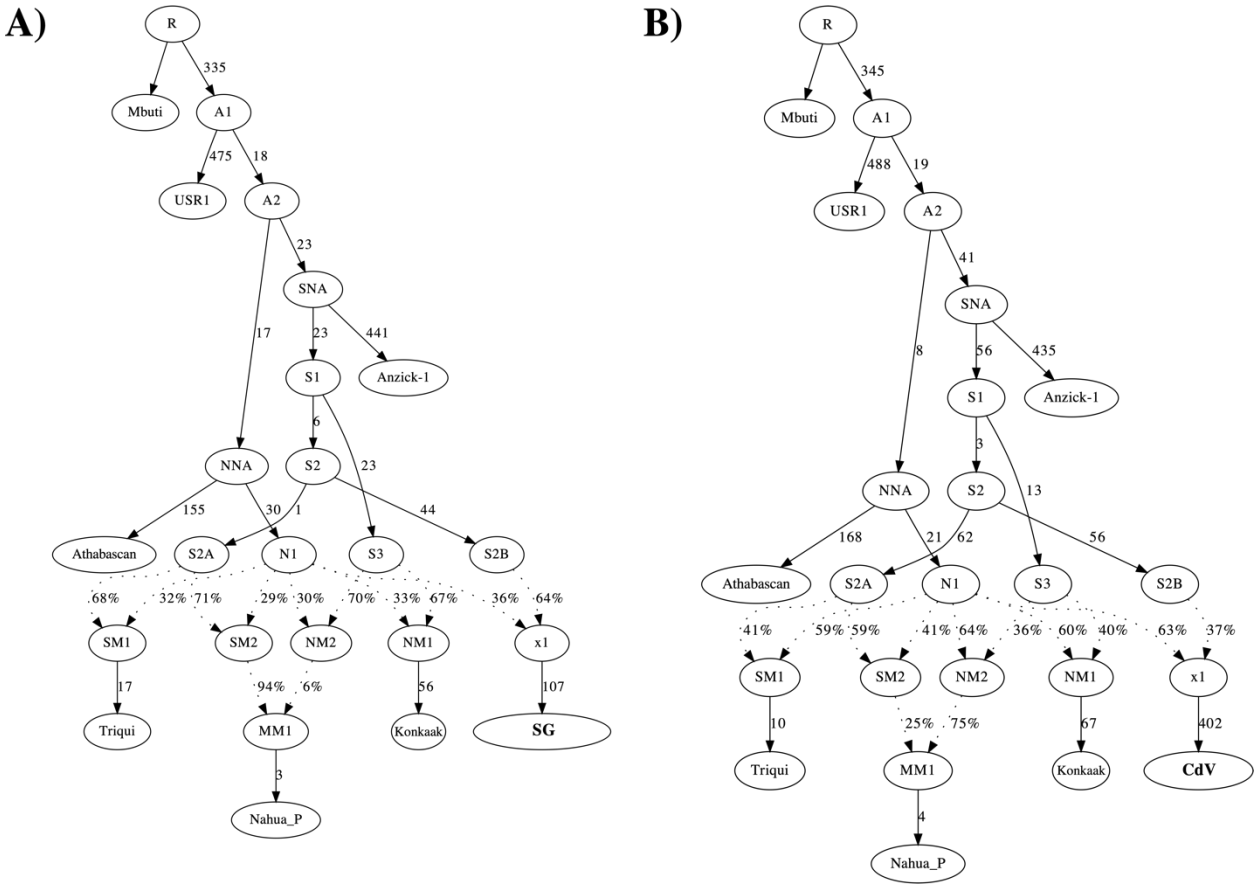

Fig. S26. Genetic ancestry proportions for the individuals from Central Mexico, inferred with qpGraph. Base model includes Anzick as a reference for the SNA and Athabascan as a reference for the NNA branch. Konkaak was used a proxy for Northern Indigenous population from Mexico, Triqui as South proxy and Nahua_P as Central Mexico proxy. (A) Model for SG individuals, it includes the genetic information of 2417Q_TOL_b, 2417J_TOL_a and, 333B_TOL_a. The tree has a |Z-score| of 1.319. (B) Model for CdV individuals, it includes the genetic information of E7_CdV_b, E8_CdV_b and E19_CdV_b. Individuals from Sierra Gorda (SG) and Cañada de la Virgen (CdV) show very different proportions of S2B and N1 ancestries. Tree with |Z-score| of 2.821. All f statistics are within 1.75 standard error.

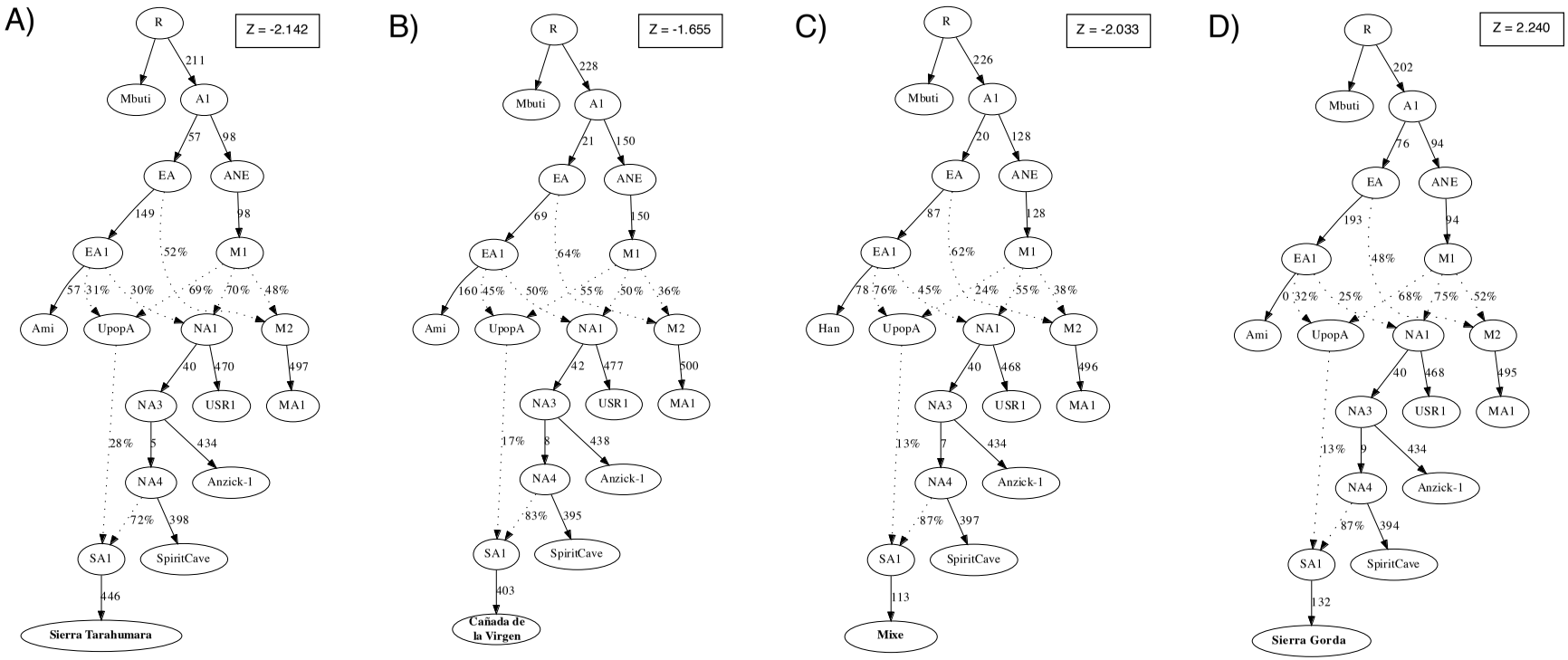

Fig. S27. Presence of ghost population genetic ancestry in pre-Hispanic from Sierra Tarahumara and Cañada de la Virgen, and present-day Mixe population inferred with qpGraph. Base model includes Mbuti as an outgroup, Ami as the East Asian, MA1 as the ancient North Eurasian, USR1 as the ancient Beringian, Anzick-1 and SpiritCave as the NNA. (A) Model with UpopA contribution to Sierra Tarahumara. (B) Model with UpopA contribution to Cañada de la Virgen. (C) Model with UpopA contribution to present-day Mixe. (D) Model rejected of UpopA contribution to present-day Sierra Gorda, with inner zeros in the branch to Ami.

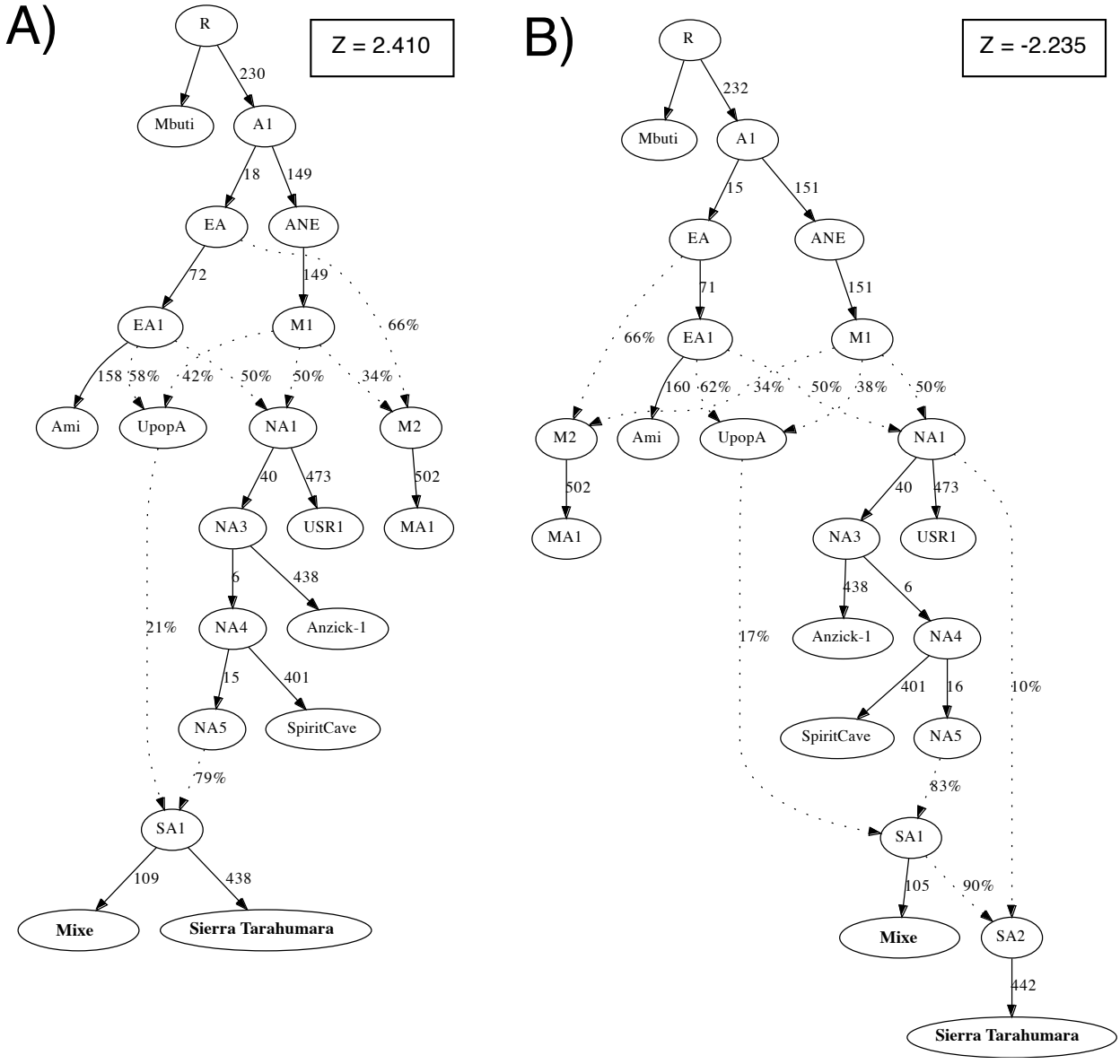

Fig. S28. Presence of ghost population genetic ancestry in pre-Hispanic from Sierra Tarahumara and present-day Mixe. Base model includes Mbuti as an outgroup, Ami as the East Asian, MA1 as the ancient North Eurasian, USR1 as the ancient Beringian, Anzick-1 and Spirit Cave as the NNA. (A) Model with UpopA contribution to present-day Mixe and Sierra Tarahumara in a single cladee. (B) Model with UpopA contribution to both present-day Mixe and Sierra Tarahumara in a different clade.

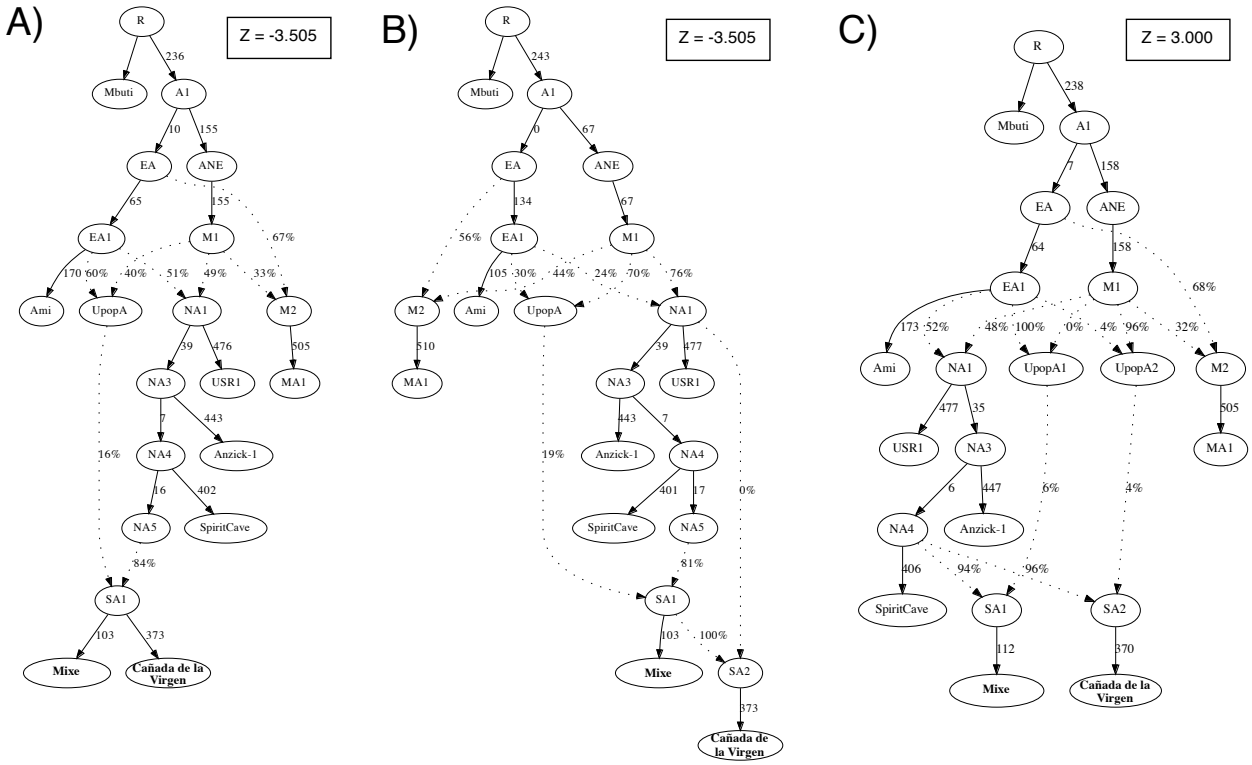

Fig. S29. Presence of ghost population genetic ancestry in pre-Hispanic from Cañada de la Virgen. Base model includes Mbuti as an outgroup, Ami as the East Asian, MA1 as the ancient North Eurasian, USR1 as the ancient Beringian, Anzick-1 and Spirit Cave as the NNA. (A) The model with UpopA contribution to present-day Mixe and Cañada de la Virgen in a single clade is rejected. (B) The model with UpopA contribution to both present-day Mixe and Cañada de la virgen in a different clade is also rejected. (C) The model with UpopA contributing to Mixe and a second "ghost" population (UpopA2) contributing to Cañada de la Virgen is accepted.

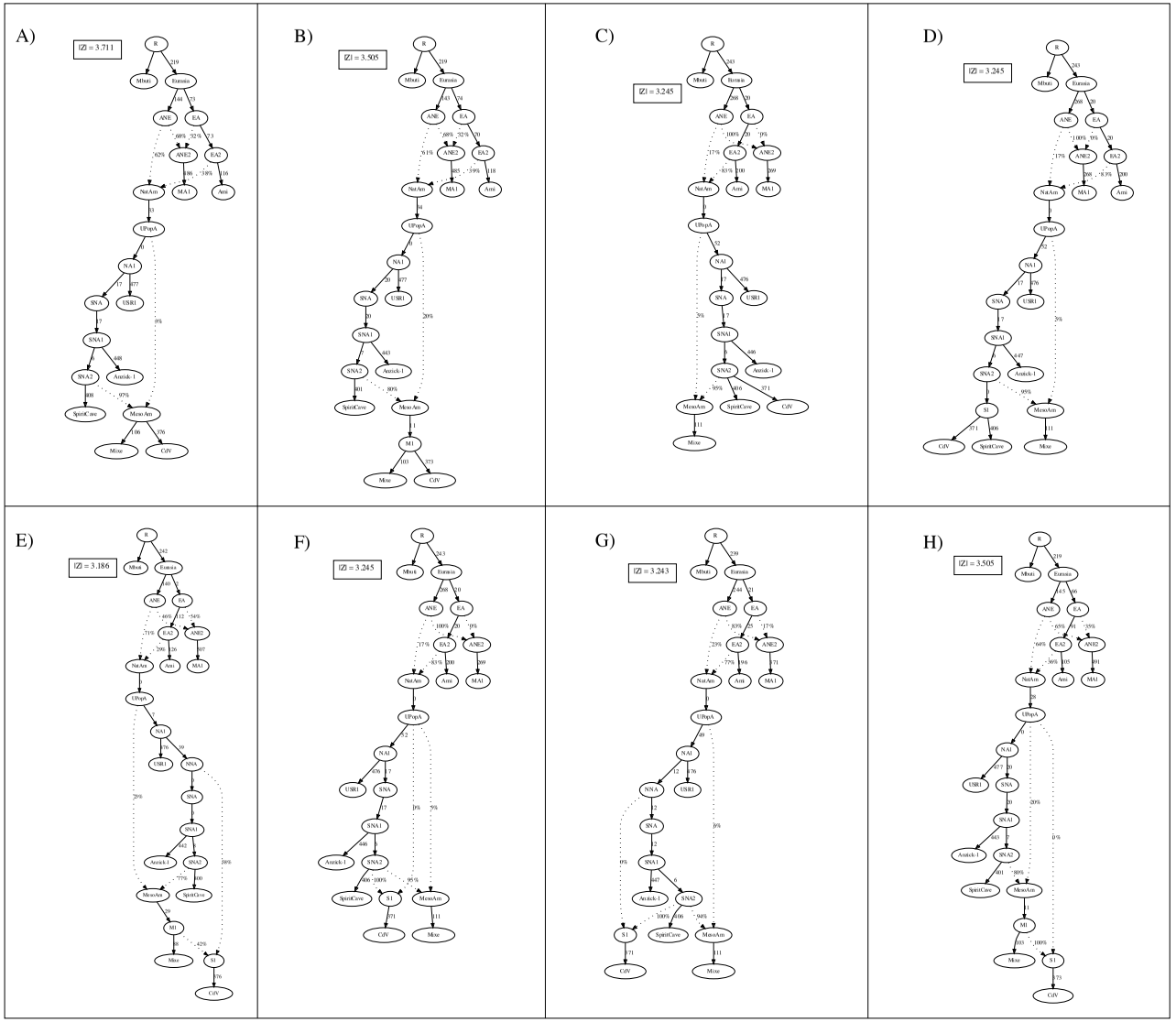

Fig. S30. qpGraphs modeling the ghost genetic ancestry in Cañada de la Virgen (CdV) and UPopA modeled with Native American clade as its ancestor. (A) Model with Cañada de la Virgen and Mixe in the same clade. (B) Same as A but including an intermediate lineage 'M1' between Mesoamerica, Mixe and Cañada de la Virgen. (C) Model with Cañada de la Virgen and SpirtCave in the same clade. (D) Same as C but including an intermediate

lineage ‘S1’ between SNA2, SpiritCave and Cañada de la Virgen. (E) Model with Cañada de la Virgen coming from ‘M1’ with a contribution of NNA. (F) Model with Cañada de la Virgen and SpirtCave in the same clade with a contribution of UPopA. (G) Model with Cañada de la Virgen and SpirtCave in the same clade with a contribution of NNA. (H) Model with Cañada de la Virgen coming from the ‘M1’ lineage with an intermediate ‘S1’ and receiving contribution of UPopA. All models are rejected.

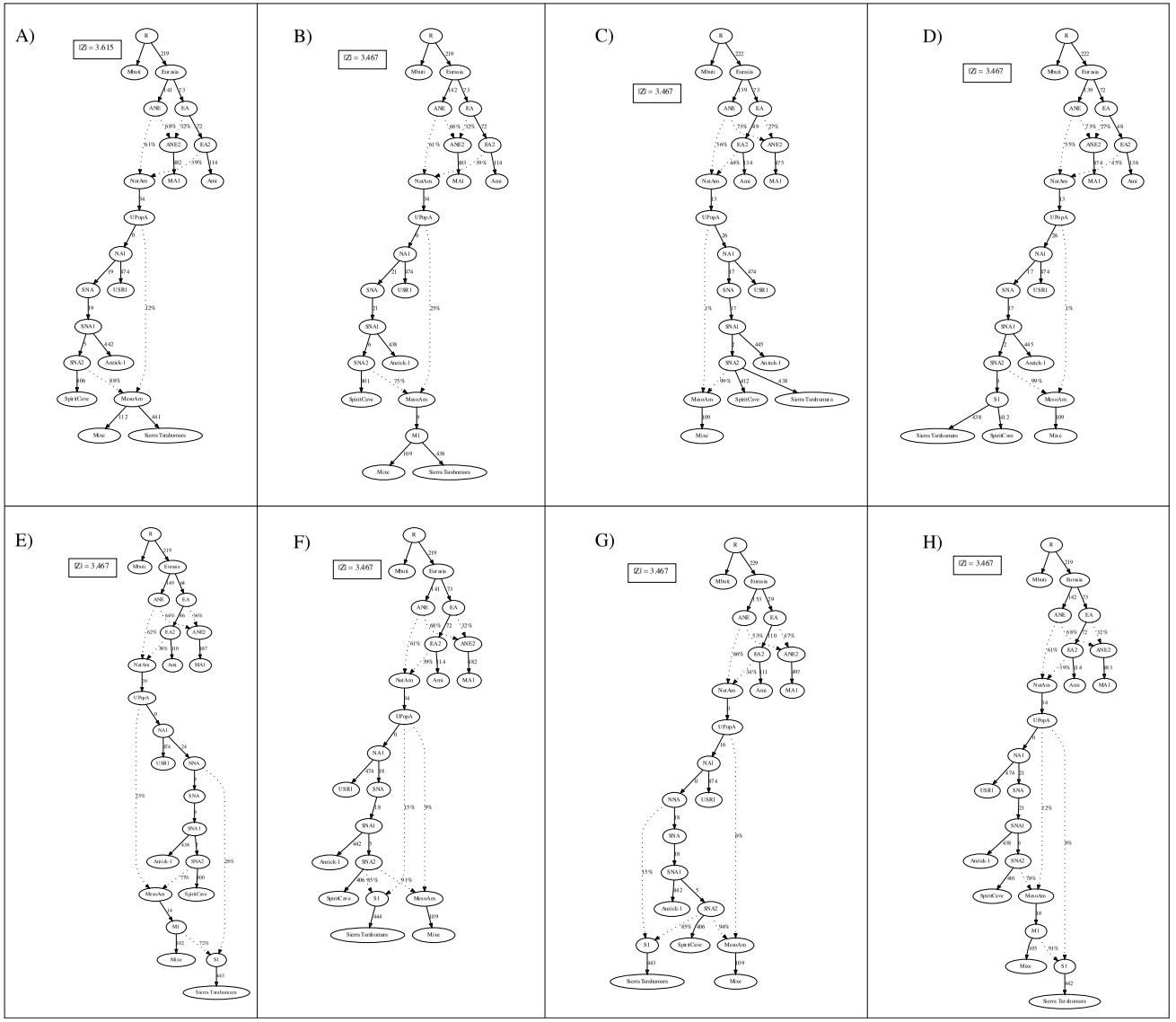

Fig.S31. **qpGraphs modeling the ghost genetic ancestry in Sierra Tarahumara and UPopA modeled with Native American clade as its ancestor.** (A) Model with Sierra Tarahumara and Mixe in the same clade. (B) Same as A but including an intermediate lineage 'M1' between Mesoamerica, Mixe and Sierra Tarahumara. (C) Model with

Sierra Tarahumara and SpirtCave in the same clade. (D) Same as C but including an intermediate lineage ‘S1’ between SNA2, SpiritCave and Sierra Tarahumara. (E) Model with Sierra Tarahumara coming from ‘M1’ with a contribution of NNA. (F) Model with Sierra Tarahumara and SpirtCave in the same clade with a contribution of UPopA. (G) Model with Sierra Tarahumara and SpirtCave in the same clade with a contribution of NNA. (H) Model with Sierra Tarahumara coming from the ‘M1’ lineage with an intermediate ‘S1’ and receiving contribution of UPopA. All models are rejected.

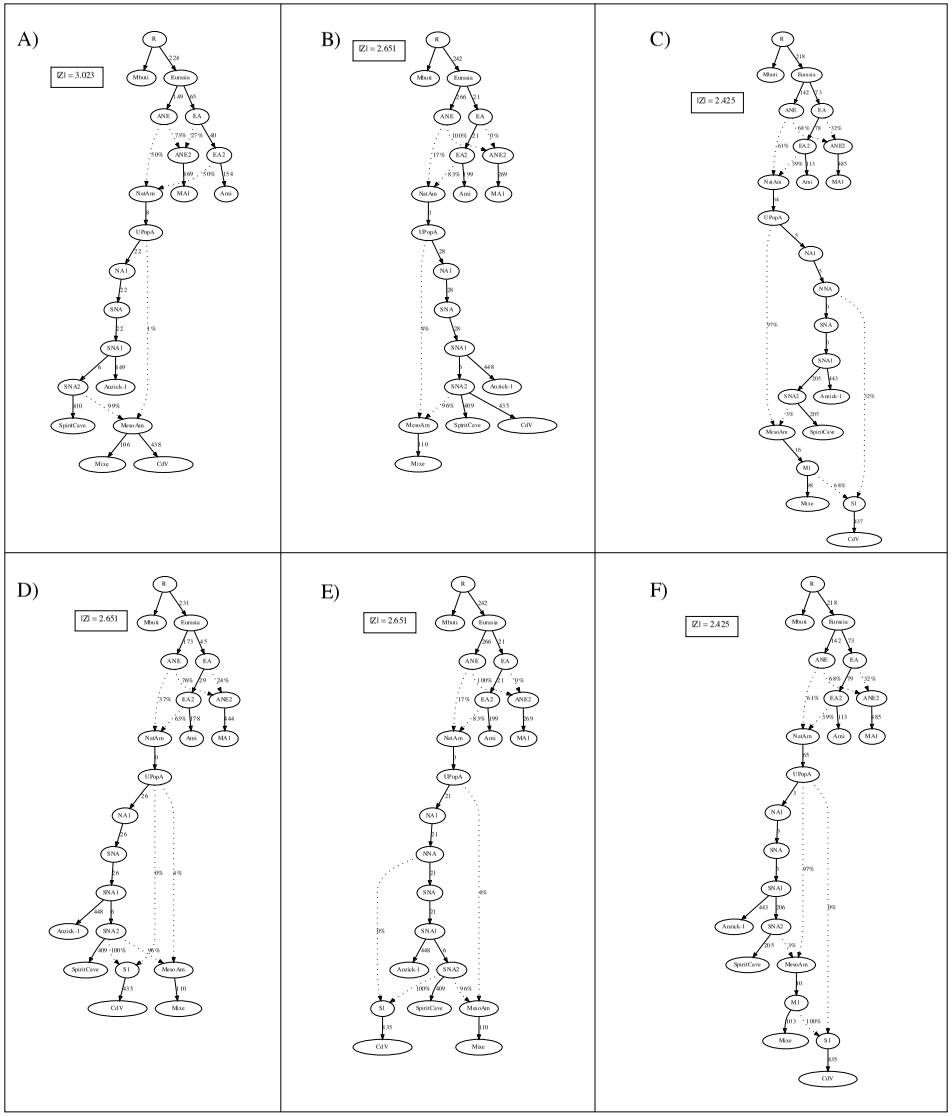

Fig. S32. **qpGraphs modeling the ghost genetic ancestry in Cañada de la Virgen (CdV), UPopA modeled with Native American clade as its ancestor and excluding USR1**. (A) Cañada de la Virgen and Mixe in the same clade. (B) Cañada de la Virgen and SpirtCave in the same clade. (C) Cañada de la Virgen coming from 'M1' with a

contribution of NNA. (D) Cañada de la Virgen and SpirtCave in the same clade with a contribution of UPopA. (E) Cañada de la Virgen and SpirtCave in the same clade with a contribution of NNA. (F) Cañada de la Virgen coming from the ‘M1’ lineage with an intermediate ‘S1’ and receiving contribution of UPopA.

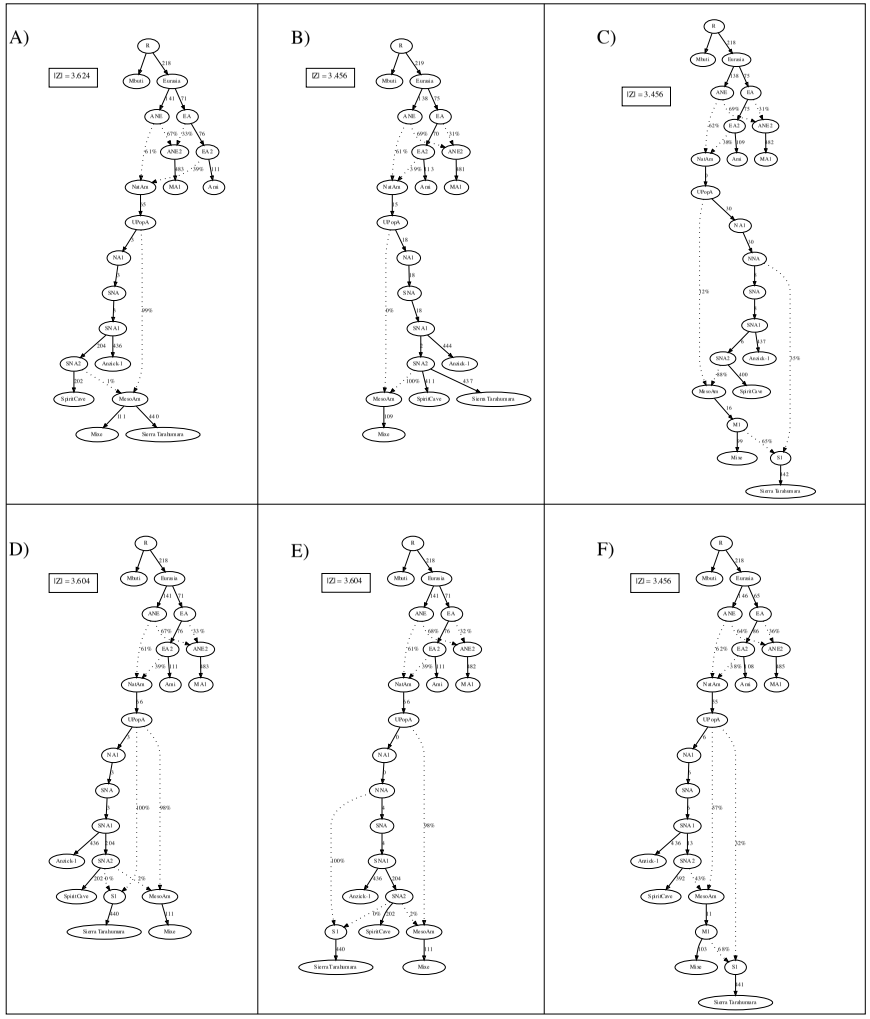

Fig. S33. qpGraphs modeling the ghost genetic ancestry in Sierra Tarahumara, UPopA modeled with Native American clade as its ancestor and excluding USR1. (A) Sierra Tarahumara and Mixe in the same clade. (B) Sierra Tarahumara and SpirtCave in the same clade. (C) Sierra Tarahumara coming from 'M1' with a contribution of NNA. (D) Sierra Tarahumara and SpirtCave in the same clade with a contribution of UPopA. (E) Sierra Tarahumara and SpirtCave in the same clade with a contribution of NNA. (F) Sierra Tarahumara coming from the 'M1' lineage with an intermediate 'S1' and receiving contribution of UPopA.