



**HAL**  
open science

## 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.

► **To cite this version:**

Viridiana Villa-Islas, Alan Izarraras-Gomez, Maximilian Larena, Elizabeth Mejía Perez Campos, Marcela Sandoval-Velasco, et al.. Demographic history and genetic structure in pre-Hispanic Central Mexico. *Science*, 2023, 380 (6645), pp.eadd6142. 10.1126/science.add6142 . hal-04244008

**HAL Id: hal-04244008**

**<https://hal.science/hal-04244008v1>**

Submitted on 20 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Demographic history and genetic structure in pre-Hispanic Central Mexico

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

## Affiliations:

<sup>1</sup>International Laboratory for Human Genome Research, Universidad Nacional Autónoma de México (UNAM), Querétaro, México

<sup>2</sup>Department of Organismal Biology, Uppsala University, Uppsala, Sweden

<sup>3</sup>National Institute of Anthropology and History, Querétaro, Mexico

<sup>4</sup>Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark

<sup>5</sup>Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA

<sup>6</sup>Departamento de Ecología Evolutiva, Instituto de Ecología, Universidad Nacional Autónoma de México, CDMX, Mexico

<sup>7</sup>Centro de Geociencias, UNAM Juriquilla

<sup>8</sup>Department of Biochemistry, Microbiology, Cell Biology and Genetics, Universidad de La Laguna, San Cristóbal de La Laguna, Spain

<sup>9</sup>Center for Computational Molecular Biology, Brown University, Providence, RI, USA

<sup>10</sup>Institute for Medical Microbiology and Hospital Epidemiology, Hannover Medical School, 30625 Hannover, Germany.

<sup>11</sup>Equity and Gender Office of the Centre for Research and Advanced Studies (CODIGO-C), Cinvestav, Mexico City, Mexico

<sup>12</sup>Unit of Advanced Genomics, National Laboratory of Genomics for Biodiversity (LANGEBIO), CINVESTAV, Irapuato, Guanajuato, Mexico

<sup>13</sup>Department of Anthropology, University of Minnesota Twin Cities, USA

<sup>14</sup>National Institute of Anthropology and History, Guanajuato, Mexico

<sup>15</sup>Department of Anthropology, University of Colorado Boulder, USA

<sup>16</sup>National Institute of Anthropology and History, Michoacán, Mexico

<sup>17</sup>Universidad de Burgos, Departamento de Historia, Geografía y Comunicaciones, Burgos, Spain

38 <sup>18</sup>La Trobe University, Department of History and Archaeology, Melbourne, Australia.

39 <sup>19</sup>Institute for Computational and Mathematical Engineering, Stanford University, USA

40 <sup>20</sup>Université Paris-Saclay, CNRS, INRIA, Laboratoire Interdisciplinaire des Sciences du  
41 Numérique, 91400, Orsay, France

42 <sup>21</sup>Lundbeck Foundation GeoGenetics Centre, GLOBE Institute, University of Copenhagen,  
43 Copenhagen, Denmark

44 <sup>22</sup>Computational Genomics, Instituto Nacional de Medicina Genómica, Ciudad de México,  
45 Mexico

46 \*Correspondence to: [mavila@liigh.unam.mx](mailto:mavila@liigh.unam.mx)

47

48

49

## 50 **Abstract**

51

52 Aridoamerica and Mesoamerica are two distinct cultural areas in northern and central Mexico,  
53 respectively, that hosted numerous pre-Hispanic civilizations between 2,500 BCE and 1,521 CE.  
54 The division between these regions shifted southward due to severe droughts ca. 1,100 years ago,  
55 allegedly driving a population replacement in central Mexico by Aridoamerican peoples. Here, we  
56 present shotgun genome-wide data from 12 individuals and 27 mitochondrial genomes from eight  
57 pre-Hispanic archaeological sites across Mexico, including two at the shifting border of  
58 Aridoamerica and Mesoamerica. We find population continuity spanning the climate change  
59 episode and a broad preservation of the genetic structure across present-day Mexico for the last  
60 2,300 years. Lastly, we identify a contribution to pre-Hispanic populations of northern and central  
61 Mexico from two ancient unsampled ‘ghost’ populations.

62

63

## 64 **Introduction**

65

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

79

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

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

89  
90 Here, we report the most extensive set of complete mitogenomes and shotgun genome-wide data  
91 to date from pre-Hispanic individuals from Mexico. We analyze our data jointly with publicly  
92 available datasets of other ancient Native Americans and present-day Indigenous populations from  
93 Mexico. This large compendium of ancient genomic data allows us to study the pre-Hispanic  
94 genetic structure in the territory occupied by Mexico, contributing to longstanding questions  
95 regarding population dynamics at the northern frontier of Mesoamerica, disentangling the complex  
96 structure of central Mexico populations, and revealing a previously unknown ancient contribution  
97 of unsampled genetic lineages to Mexican populations.  
98

## 99 **Results**

### 100 **Sampling and pre-Hispanic genomic data generation**

101 We screened archeological samples from 37 pre-Hispanic individuals excavated at seven sites  
102 from within Mesoamerica in central Mexico. Three are in the Sierra Gorda in Querétaro state  
103 (Toluquilla and Ranas sites, and a cave in Cadereyta de Montes), three in Michoacán state  
104 (Zaragoza, Tanhuato, and La Mina sites) and one in Guanajuato state (Cañada de la Virgen site)  
105 (Fig. 1). Furthermore, we produced additional shotgun data for two mummies from the Sierra  
106 Tarahumara from Aridoamerica, northern Mexico who were sequenced at lower depths in a  
107 previous study (16) (Fig. 1). The sampling of the archeological human remains was made  
108 following approval by the Archaeology Council of the Instituto Nacional de Antropología e  
109 Historia with permit numbers 401.3S.16-2017/990 and 401.3S.16-2019/222. We sampled the  
110 minimum possible amount of tissue to avoid unnecessary destruction. For teeth, we tried to  
111 separate the root without damaging the crown and returned all leftover material to the archeologists  
112 responsible for each collection. Although consultation with Indigenous populations for destructive  
113 analysis of human archeological material is not a requirement by law in Mexico and the studied  
114 sites are not located within Indigenous communities, we acknowledge the need to involve  
115 Indigenous perspectives in delineating the regulations for these types of studies (see (17) for a  
116 deeper discussion on the subject). To engage the neighboring communities, we delivered public  
117 lectures to communicate the genetic results to the village closest to the sites in Sierra Gorda.  
118

119  
120 After processing and sequencing samples, we obtained informative amounts of shotgun sequence  
121 data (0.01x – 4.7 x genomes) for 12 individuals spanning a time transect of 750 years (600 – 1351  
122 CE). These include ten low-depth genomes from within Mesoamerica in central Mexico, and two  
123 from Aridoamerica, in the Sierra Tarahumara (Table S1). In addition, we reconstructed the  
124 mitochondrial genomes for 27 individuals across these sites (5.7 – 1284.8x), spanning a time  
125 transect of 1,600 years (320 BCE to 1,351 CE). (Table S1 and S2). We included in the dataset  
126 three previously published ancient genomes (Table S3) as well as nine ancient mitogenomes from  
127 northern Mexico (Table S4), yielding a total of 15 ancient genomes and 36 ancient mitogenomes  
128 from the Mexican territory.



129  
130 We named each sample according to burial and individual identifiers provided by archaeologists  
131 and added two additional labels. The first label includes 1-3 letters that refer to the archaeological  
132 site, while the second refers to whether the individual dates to before (label 'b') or after (label 'a')  
133 the drought period. The three pieces of information are divided by an underscore character '\_'.  
134 Throughout the text we refer to all individuals using these IDs.  
135

### 136 **Genetic structure and diversity in uniparental markers**

137 All chromosomal sex assignments matched the morphological sex except one (XX chromosomally  
138 and male based on morphology) (Table S1) (18). It was possible to assign Y-DNA haplogroups to  
139 five out of the thirteen individuals assigned as XY, all of which had a Native American Q lineage.  
140 This is in agreement with previous studies on ancient and present-day Indigenous Mexican  
141 individuals (19–24). At the sub-haplogroup level, we found Q1a2a1-L54 and Q1a2a1a1a1-M3,  
142 with no apparent differences between Aridoamerica and Mesoamerica (Table S1).  
143

144 All 27 reconstructed mitochondrial genomes carried one of the mtDNA haplogroups found in  
145 Indigenous populations of the Americas: A (n=10), B (n=9), C (n=4), and D (n=4) (25) (Table S1,  
146 and S4) (18). After merging with nine additional pre-Hispanic mitogenomes from Mexico from  
147 other studies (15, 16, 26), we found all four haplogroups in Mesoamerica (n=25), and only C and  
148 B in Aridoamerica (n=11) (Table S4). Consistently, we observed higher nucleotide diversity  $\pi$   
149 values in Mesoamerican populations compared to Aridoamerica (18).  
150

151 The pre-Hispanic mtDNA haplogroup distribution resembles the one in present-day individuals  
152 (24, 26, 27), showing an overall continuity of the matrilineal genetic structure for at least 2,300  
153 years. Of the available complete ancient mtDNA genomes from Aridoamerica, 36% harbor  
154 haplogroup C, decreasing in frequency to 8% in Mesoamerica, while the opposite occurs for  
155 haplogroup A, with 40% of presence in Mesoamerica and 0% in Aridoamerica (Table S4). This  
156 gradient has been previously observed in ancient and contemporary populations from Mexico (19,  
157 20, 24, 26–32). Moreover, we detected unreported variants in the mtDNA PhyloTree (33) for seven  
158 individuals assigned to the sub-haplogroups A2d, B2c, and B2l (Table S5) (18).

159 In Sierra Gorda, we identified sub-haplogroup A2d in six individuals, four pre-drought and two  
160 post-drought. After quality filtering (18), five of them were included in a median-joining network  
161 (Table S4). Four (two pre- and two post- drought) clustered together in a single clade, despite  
162 spanning a time transect of 1,480 years, while the fifth individual remained in a different clade  
163 within A2d (Fig. 2A). Remarkably, P\_CCM\_b and 2417J\_TOL\_a are found in the same node  
164 within A2d. This reflects a continuity of this maternal lineage, despite the severe droughts in Sierra  
165 Gorda. In contrast, the haplotype network for haplogroup B including five Sierra Gorda  
166 individuals, shows that the three pre-drought individuals share sub-haplogroup B2l, while the two  
167 post-drought are related and found in a node representing sub-haplogroup B2c. Since mtDNA is  
168 subject to genetic drift that leads to lineage loss, it cannot accurately inform about genetic structure  
169 or admixture events; the discrepant patterns between haplogroups A and B could be explained by  
170 a number of scenarios, which are difficult to test due to the small sample size. This highlights the  
171 need to explore autosomal data to test the population replacement hypothesis reliably.

172

173 We further leveraged the mtDNA data to estimate the past female effective population size ( $N_e$ )  
174 using Extended Bayesian Skyline Plots (EBSP) in pre-Hispanic ( $n=29$ ) and present-day Mexico  
175 ( $n=232$ ) (Table S4 and S6), per haplogroup and all haplogroups together (Fig. S6A-H). All EBSPs  
176 runs (Table S7) showed wide (95%) Confidence Intervals (CI) except in the EBSP of only present-  
177 day Mexico, where we observed a population expansion ca. 5 kybp and a population decline ca.  
178 500 ybp, though the median values do not support the latter (Fig S6G-H). Despite the wide CI,  
179 there is a clear distinction between the estimated female  $N_e$  per haplogroup, with haplogroup A  
180 reaching the highest present-day values (104,048), followed by D (100,812) and decreasing almost  
181 an order of magnitude for haplogroups B and C (37,155 and 7,226, respectively). To improve the  
182 estimations and get insights about the past female  $N_e$  in a wider geographical range, we merged  
183 our dataset with available ancient mtDNA sequences from South America ( $n=137$ ) (Table S6).  
184 The EBSPs for haplogroups A, B, and D, respectively also show wide CI (Fig. S6I-J, N), while  
185 haplogroup C shows a clearer pattern of population expansion starting ca. 15 kybp and a population  
186 decline starting ca. 5 kybp (Fig. S6K). Consistent with our previous observation, the estimated  
187 female  $N_e$  is notably higher for haplogroup A and D, than for B and C in the merged dataset (Fig.  
188 S6I-L).

189

### 190 **Autosomal genetic structure in pre-Hispanic Mexico**

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

199 We projected the pre-Hispanic individuals onto a principal components space of only pre-Hispanic  
200 and present-day Indigenous populations from Mexico (16, 19, 23, 37) using a novel approach  
201 named “Missing DNA” PCA (mdPCA) that corrects for the high missing fraction of low-depth  
202 ancient genomes (see Materials and Methods). All ancient individuals cluster closely in the  
203 mdPCA with present-day populations from the same geographical region, except for individual  
204 MOM6\_ST\_a which is intermediate between northern and central present-day populations. Both  
205 pre-, and post- drought Sierra Gorda individuals cluster together with present-day Nahua from  
206 Jalisco, Purépecha, Totonac, Nahua from Puebla, and Nahua, all from central Mexico (Fig. 3A).  
207 This pattern was replicated when using traditional PCA (Fig. S9) and a Temporal Factor Analysis  
208 approach (39), which corrects the ancestral relationships using the dates of the pre-Hispanic  
209 individuals (Fig. S10) (18). In agreement, the ADMIXTURE analysis assuming six ancestral  
210 components reveals that pre-Hispanic individuals show similar genetic ancestry proportions to  
211 those observed in present-day populations from the same region (Fig. 3B) and that individuals  
212 from Sierra Gorda show a homogeneous genetic composition independently of the period in which  
213 they lived (Fig. 3B).

214 We then performed several combinations of D- and Outgroup-f3 statistics to further explore the  
215 genetic structure of ancient and present-day populations (18). Outgroup-f3 statistics were of the  
216 form  $f_3(\text{Test}, \text{Source1}; \text{YRI})$ , where the Test is the pre-Hispanic individual under analysis, and

217 Source1 is a present-day Indigenous population. D-statistics were of the form D(Pop1, Pop2; Test,  
218 YRI), where Test refers to the pre-Hispanic individual under analysis, Pop1 and Pop2 are all  
219 possible combinations of present-day Indigenous populations. YRI represents Yoruba as the  
220 outgroup population in both tests. We found F9\_ST\_a to share higher genetic drift with present-  
221 day northern Mexico populations than with any other present-day population (Fig. S11) (Table S9)  
222 and to be significantly more related to the present-day Rarámuri than to any other population when  
223 tested in the form D(Pop1, Rarámuri; F9\_ST\_a, YRI) (Fig. S12) (Table S10). However, a  
224 population continuity test that considers the read counts in the ancient individual and allele  
225 frequencies in the present-day population (40) rejected the null hypothesis of population continuity  
226 between F9\_ST\_a and present-day Rarámuri a p-value of  $10^{-499.3}$  (18).

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

233

#### 234 **Genetic diversity in pre-Hispanic Mexico**

235 We estimated the conditional nucleotide diversity (CND) (41) and runs of homozygosity (ROH)  
236 (42) in pre-Hispanic and present-day Indigenous individuals from Mexico (19, 23) to make  
237 inferences about past genetic diversity and effective population sizes and investigate changes in  
238 patterns of genetic variation that may have arisen after the drought period (18). Both CND values  
239 and ROH distribution show that pre-Hispanic and present-day individuals from northern Mexico  
240 (Pericúes and Akimel O’odham) have the lowest CND values and the highest sum of inferred ROH  
241  $>4$  cM compared to other populations of their respective period (Table S11, Table S12, Table S13).  
242 Pericú hunter gatherers had the lowest CND values (Fig. S19) (Table S12) but the ROH proportion  
243 revealed a larger population size than another ancient hunter gatherer population from Patagonia  
244 (Fig. S19) (Table S12) (18). These results are in agreement with previous studies reporting lower  
245 genetic diversity and longer ROH in hunter-gatherer populations from South America than those  
246 observed in other regions of the Americas (43).

247 Pre-Hispanic individuals from Sierra Tarahumara, Toluquilla and Cañada de la Virgen show  
248 similar values of CND and sum of inferred ROH segments. In Toluquilla we found that the CND  
249 value increases with the date difference between pairs of individuals, probably reflective of the  
250 accumulation of new mutations during the 489-680 years of difference between pre-  
251 (2417Q\_TOL\_b) and post- (2417J\_TOL\_a and 333B\_TOL\_a) drought individuals. Lastly, the  
252 segment size distribution of ROHs suggests that the pre-Hispanic individuals studied here  
253 belonged to populations with small effective population ( $N_e$ ) sizes ( $2N_e= 1,600-6,400$ ) (Fig. S20)  
254 (18), which is in agreement with  $N_e$  previously calculated for present-day northern and southern  
255 populations (44).

256

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

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

260 statistics. We used individuals from Sierra Gorda from the pre- (TOL\_b and R\_b) or post- (TOL\_a)  
261 drought period as the test population in an outgroup-f3 of the form  $f_3(\text{Test}, \text{Source}; \text{YRI})$ , and  
262 Source representing another ancient individual from Mexico or ancient representatives of other  
263 regions in the Americas (North, Central and South America) (Table S3).

264 Under a population replacement scenario, we would expect the Toluquillan individuals who lived  
265 after the drought to be more closely related to pre-Hispanic individuals with higher proportions of  
266 Aridoamerican genetic ancestry (Pericúes and F9\_ST\_a) than to Toluquillan individuals who lived  
267 before the drought (2417Q\_TOL\_b). Instead, what we observe is that the pre-Hispanic individuals  
268 from Sierra Gorda before and after the climate change episode share higher genetic drift between  
269 them than with any other pre-Hispanic individual (Fig. 4A), except for the individual 11R\_R\_b,  
270 who shows higher shared genetic drift with an individual from Cañada de la Virgen (Fig. S21)  
271 (Table S14). However, standard errors for 11R\_R\_b are high, due to its low coverage (Fig. S21).  
272 In the case of the two individuals from Sierra Tarahumara, the highest value of genetic drift is  
273 between them (Fig. S22). In contrast, outgroup-f3 tests with individuals from Cañada de la Virgen  
274 (CdV) as the Test population in  $f_3(\text{CdV}, \text{Source}; \text{YRI})$  (Fig. S23), have the highest outgroup-f3  
275 values when a related individual (inferred using a relatedness analysis, (18)) from the same site is  
276 used as the source, but the Source with the second highest value does not belong to Cañada de la  
277 Virgen.

278 We performed D-statistic tests in the form  $D(\text{SG}_b, \text{SG}_a; \text{Pop3}, \text{YRI})$ , with the individuals from  
279 Sierra Gorda of each period, first without collapsing them per period and then merging the  
280 individuals in their respective group, and Pop3 being any pre-Hispanic or present-day individual  
281 or population from Mexico. SG\_b represents the pre-drought individuals 2417Q\_TOL\_b and  
282 11R\_R\_b, while SG\_a includes the two post-drought individuals 2417J\_TOL\_a and  
283 333B\_TOL\_a. Notably, 2417Q\_TOL\_b shared more alleles with California or Pericú individuals  
284 than the post-drought individuals, but this sharing was not observed when only using transversions  
285 (Fig. S24) (Table S15). When merging the individuals per period, none of the comparison under  
286 this model deviates significantly from  $D=0$ , meaning that SG\_b and SG\_a form a clade to the  
287 exclusion of other individuals and are more closely related to each other than with any other pre-  
288 Hispanic or present-day population. The only exception was when Pericúes were used as Pop3,  
289 where  $D>0$  with a  $|Z|=3.625$ , indicating that the pre-drought individuals are more closely related  
290 to Pericúes (Fig. S25) (Table S16). However, this significant deviation disappears when repeating  
291 the analysis using only transversion sites (Fig. 4B) (Table S17).

292 We then used f-statistic-based admixture graphs (qpGraph) (45) to further evaluate the population  
293 continuity scenario in Sierra Gorda by testing the fit of models that grouped individuals from  
294 before and after the drought period in the same clade. For the model, we used the 11,500-kyo  
295 USR1 (46) as the surrogate for ancient Beringian, 10,700-kyo Anzick (47) individual as the  
296 surrogate for ancestral Southern Native American (SNA) (16) and the 4,200-kyo Ancient  
297 Southwestern Ontario individual (ASO) (15) as the surrogate for Northern Native American  
298 (NNA) (16). In the analysis, we included the pre-Hispanic individuals with the highest coverage  
299 2417Q\_TOL\_b, 2417J\_TOL\_a, and 333B\_TOL\_a from Toluquilla and F9\_ST\_a as the surrogate  
300 for Aridoamerica. The three individuals can in fact be modeled in the same clade separated from  
301 F9\_ST\_a, with no contribution to this (Fig. 4C).

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

305 individuals can be modeled in the same clade without the need of an admixture event (p-value  
306  $>0.12$ ). In contrast, an additional genetic source is needed to reflect genetic variation in F9\_ST\_a,  
307 Pericúes, Akimel O' odham, and Maya (Table S18).

308

### 309 **Admixture model for Central Mexico populations**

310 Given the complex relationships observed for ancient individuals in central Mexico, we tested  
311 additional demographic models using qpGraph. The base model that fitted the pre-Hispanic  
312 populations includes Anzick (47) individual as the ancestral SNA (16), present-day Athabascan  
313 (16) as the representative for NNA (16), and present-day Indigenous populations in Mexico (37)  
314 representing the northern (Konkaak), central (Nahua from Puebla), and southern (Triqui) regions.  
315 Furthermore, this model involves a split of the Southern Native American into S2A and S2B  
316 sources. The pre-Hispanic populations from Sierra Gorda and Cañada de la Virgen show different  
317 levels of genetic ancestry from Southern Native American and Northern Native American. We  
318 found that the population from Sierra Gorda had a higher percentage of the Southern Native  
319 American branch from S2B (64%) than the population from Cañada de la Virgen (37%) who had  
320 62% genetic ancestry shared with the Northern Native American branch, indicating a different  
321 demographic history for each of these populations despite its relative geographic proximity (Fig.  
322 S26). This model supports previous studies that report several admixture events between the two  
323 branches that have given rise to Central and South American populations (15). For the Michoacán  
324 population, we obtained terminal branches with no support (zeros); thus, we did not make  
325 conclusions with this model for the Michoacán pre-Hispanic population.

326

### 327 **Ghost population contribution to pre-Hispanic Mexico**

328 Earlier studies report the contribution of a “ghost” genetic ancestry from an unsampled group,  
329 designated UpopA, among the present-day Mixe from Mexico (48). We tested the presence of this  
330 UpopA in the pre-Hispanic individuals, using combinations of admixture graph models.  
331 Interestingly, Sierra Tarahumara (represented by F9\_ST\_a) and Cañada de la Virgen (represented  
332 by E8\_CdV\_b) show a genetic ancestry contribution from a ‘ghost’ population, at 28% and 17%,  
333 respectively (Fig. S27A-B), consistent with the “ghost” UpopA genetic ancestry previously  
334 reported in Mixe (Fig. S27C), while models including individuals from the other archaeological  
335 sites were rejected. When we modeled the pre-Hispanic individuals and Mixe together we found  
336 that Sierra Tarahumara and Mixe share the same ghost genetic ancestry UPopA (with a better Z-  
337 score when Sierra Tarahumara receives NA1 genetic ancestry) (Fig. S28), while Cañada de la  
338 Virgen requires the contribution of a second “ghost” genetic ancestry from that found in Mixe  
339 (models without this second “ghost” genetic ancestry are rejected,  $|Z| > 3$ ), which we name  
340 “UPopA2” (Fig. S29). Finally, we included Sierra Tarahumara, Cañada de la Virgen and Mixe and  
341 the two “ghost” populations in the same model and confirmed that F9\_ST\_a and Mixe share the  
342 contribution from UpopA1, while Cañada de la Virgen requires the contribution an additional  
343 unsampled group UpopA2 (Fig. 5) (18).

344

### 345 **Discussion**

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

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

353 Our findings exhibit a geographical structure in pre-Hispanic individuals that differentiates  
354 northern and central populations from Mexico, which is consistent with the northwest-southeast  
355 cline observed in the genetic structure of present-day Indigenous populations (37). The pre-  
356 Hispanic individuals cluster in proximity to present-day Indigenous populations from the same  
357 geographical area in the mdPCA and exhibit similar ancestral components based on ADMIXTURE  
358 analysis, except for individual MOM6\_ST\_a (Fig. 3 and S9-S10). This reflects an overall  
359 conservation of the genetic structure of the populations inhabiting the Mexican territory (37) for  
360 at least 1,400 years (which is the date of the most ancient individual in the dataset). This is  
361 consistent with demographic models based on present-day Indigenous populations (44),(49)  
362 proposing a northern/southern population split between 4,000 and 10,000 years ago followed by  
363 multiple waves of admixture events between them (50). This geographical structure is also  
364 reflected in the maternal lineages (Table S5). The spatial distribution of the haplogroups found,  
365 namely A, B, C and D, closely resembles that of present-day Mexico (32, 51).

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

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

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

397  
398 To directly test the hypothesis of a population replacement at the northern frontier of Mesoamerica  
399 by Aridoamerican hunter-gatherers after the 900-1,300 CE droughts, we studied Sierra Gorda  
400 individuals from the time before and after the climate change episode. Outgroup-f3 analyses  
401 showed that individuals from Sierra Gorda pre- and post-drought shared higher genetic drift  
402 between them than with any other pre-Hispanic individual. Using D statistics, we observed that  
403 pre- and post-drought individuals from Sierra Gorda always form a clade to the exclusion of the  
404 rest of pre-Hispanic or present-day populations from Mexico. Then, qpGraphs showed that pre-  
405 and post-drought individuals from Sierra Gorda can be modeled in the same clade, different from  
406 the one formed by F9\_ST\_a as an Aridoamerican surrogate. Regardless F9\_ST\_a might not belong  
407 to the population that allegedly replaced the population in Sierra Gorda following the drought; we  
408 used him safely as a surrogate for northern genetic ancestry since previous studies have estimated  
409 that the split time between the northern and southern ancestries occurred ca. 7,200ya (44).  
410 Therefore, even though F9\_ST\_a likely did not belong to the population that allegedly replaced  
411 Sierra Gorda inhabitants following the drought, we can confidently use this genome as a surrogate  
412 for northern genetic ancestry since, by the time of the climate change episode, both ancestries were  
413 already well differentiated. Furthermore, with qpWave, we confirmed that pre- and post-drought  
414 individuals from Sierra Gorda genetic make-up can be explained within the same genetic history  
415 without an additional source of genetic variation. Even if this replacement had happened by an  
416 unsampled northern population, we would still expect to see signals that significantly differentiated  
417 the post- drought individuals with an input from Aridoamerica, in the qualitative (mdPCA and  
418 ADMIXTURE) and the quantitative (f3, Dstats, qpGrap and, qpWave) analyses, which is not the  
419 case. Instead, the evidence points to population continuity in Sierra Gorda after the climate change  
420 episode.

421  
422 A possible explanation for a population continuity despite the droughts is that the favorable  
423 climatic conditions at the northern Sierra Gorda maintained higher humidity than in other arid sites  
424 of the northern frontier of Mesoamerica, such as Cañada de la Virgen. Because agriculture was the  
425 main subsistence strategy in Cañada de la Virgen, the droughts forced its inhabitants to migrate to  
426 other regions during the heavy droughts, which resulted in an abandonment of the site between  
427 1,000-1,100 CE (58). In contrast, the main subsistence strategy in Toluquilla and Ranas was the  
428 mining and trade of cinnabar, a valuable mineral of sacred value in pre-Hispanic cultures (57, 61,  
429 62). We hypothesize that the cinnabar trade and the landscape of the Sierra Gorda allowed the  
430 peoples of Toluquilla and Ranas to subsist despite low rainfall conditions during the drought.  
431 Notably, we found shorter ROH segments in the pre-drought pre-Hispanic individual than in the  
432 post-drought individuals. This could indicate a reduction in the population size after the climate  
433 change episode and suggest a possible demographic impact in the same population. Additional  
434 assessment of this hypothesis at other sites at the northern frontier of Mesoamerica will shed light  
435 into the population migrations and dynamics in a wider geographical range.

436  
437 We were also interested in obtaining insights into the demography of central Mexico populations  
438 since previous attempts of demographic modeling have been hampered by the high genetic  
439 heterogeneity observed in these populations (44, 52). Our outgroup-f3 and D statistics results show

440 that, although pre-Hispanic individuals from Mesoamerica, in central Mexico share higher genetic  
441 drift with present-day populations from that region, none of these relationships were statistically  
442 significant. Using qpGraph we found that the pre-Hispanic populations from central Mexico, all  
443 have different genetic ancestry sharing with the NNA and SNA branches. This was expected given  
444 previous studies on ancient genomes from Central and South America reporting multiple  
445 admixture events between these two branches after their split ca. 15,000 years ago (15, 50).  
446 Together, these observations point to a scenario in which populations from central Mexico have  
447 not completely diverged from one another, possibly due to extensive gene flow as expected based  
448 on the active commercial exchange between different Mesoamerican populations for centuries  
449 (63). This interaction was mainly through trade routes and alliances between different nations (63),  
450 as revealed in present-day Indigenous populations who share identity-by-descent (IBD) segments  
451 (37, 50) and through the analysis of STR loci (56). The study of IBD segments in present-day  
452 Indigenous populations from Mexico has also evidenced gene flow between Mesoamerican and  
453 Aridoamerican populations in pre-Hispanic times (50). However, this gene flow occurred less  
454 frequently than within Mesoamerica. The individual MOM6\_ST\_a from Aridoamerica may have  
455 been the result of such admixture between Mesoamerican and Aridoamerican ancestors.

456 Furthermore, the qpGraph admixture models explored for the pre-Hispanic populations showed  
457 F9\_ST\_a and the ancient individuals from Cañada de la Virgen have genetic ancestry from an  
458 “ghost” or unsampled population. A contribution from an unsampled population, named ‘UpopA’  
459 was previously identified in present-day Mixe (48) as well as in present-day northern and central  
460 Indigenous populations from Mexico (50). UpopA was estimated to have diverged ~24,700 years  
461 ago from Native Americans (48). Interestingly, while Sierra Tarahumara shares the contribution  
462 of UpopA with Mixe, a second “ghost” population (‘UpopA2’) is needed to model Cañada de la  
463 Virgen in the admixture model together with Sierra Tarahumara and Mixe (Fig. 5). This  
464 observation reveals a complex population history in the Americas during the late Pleistocene that  
465 needs to be further characterized. Additional aDNA studies from the Americas could contribute to  
466 the identification of the source of both “ghost” ancestries, seemingly contributing to many present-  
467 day Indigenous populations from Mexico.

468  
469 In summary, we find that the pre-Hispanic population structure from over a thousand years ago  
470 can still be observed today. Our work, together with previous studies (15, 37, 44, 50, 52), shows  
471 that the demographic events that gave rise to Aridoamerican and Mesoamerican populations are  
472 more complex than previously thought. Commercial trade routes may have contributed to  
473 increased mobility, facilitating gene flow between different populations within and between  
474 various cultural areas. Furthermore, we found genetic continuity in the Sierra Gorda region at the  
475 northern frontier of Mesoamerica which suggest the local population stayed in their homeland  
476 despite the longstanding droughts that forced other populations to abandon their cities. The  
477 identification of a second “ghost” genetic ancestry contribution to some pre-Hispanic central  
478 Mexico population reveals a complex past that needs to be characterized through the ethical study  
479 of ancient genomes from Mexico. Our study opens the door for further research to address the  
480 questions of the unknown genetic past and population dynamics of Mexican pre-Hispanic  
481 populations, whose genetic legacy is retained today among Indigenous and admixed populations.

## 482 **Materials and Methods**

483  
484



## 485 Laboratory procedures

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

496 DNA extraction from individual P\_CCM\_b was performed taking 150 mg of mummified skin.  
497 The sample was washed in deionized sterile water and dried. Then, the sample was UV irradiated  
498 (256 nm) for 1.5 minutes using a UVP CL-1000 crosslinker. Epithelium was cut in small pieces  
499 and incubated at 50°C for 24 hours in lysis buffer (10mM Tris-HCl, 10mM NaCl, 5mM CaCl<sub>2</sub>, 2.5  
500 mM EDTA, 1% SDS, 10 mg/ml proteinase K, 10 mg/mL DTT). Followed by centrifugation at  
501 16,100 x g for 5 minutes and recovery of the supernatant. Then, DNA purification was performed  
502 according to the protocol in (65).

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

519

## 520 Sequence data processing

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

525

526 The retained reads were mapped to the human reference genome b37 (hg19), with the  
527 mitochondrial sequence replaced by the revised Cambridge reference sequence (rCRS,  
528 NC\_012920). Mapping was done with bwa 0.7.13, aln algorithm, seed disabled (-l 500), and  
529 keeping reads with mapping quality >25. Clonal duplicates and reads mapping to more than one

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

539

#### 540 mtDNA analysis

541 The program Schmutzi (73) was used to estimate mtDNA contamination and to obtain consensus  
542 sequences. The reads aligned to the Cambridge reference sequence (rCRS, NC\_012920) were  
543 analyzed with the script contDeam.pl with parameters --library double and the default --uselength,  
544 and schmutzi.pl was then run with 8 threads. Haplogrep 2 (74) was used to assign a mitochondrial  
545 haplogroup and quality of assignment for each library using the fasta file with the consensus  
546 mitochondrial genome generated by Schmutzi (Table S1).

547

548 For the estimation of the mitochondrial genetic diversity, we aligned the 27 mitochondrial  
549 genomes reconstructed in this study along with the eight public mitochondrial genomes from  
550 Mexico (Table S4), using the BioEdit sequence alignment editor v. 7.0.5.3 (75). The haplotype list  
551 was obtained with DNA Sequence Polymorphism software (DNAsp) v. 6.12.03x64 (76). The sites  
552 with gaps were not considered. Then, we estimated the nucleotide diversity  $P_i$  through the genetic  
553 distance method of Tajima and Nei, with the software Arlequin v. 3.5.1.3 using default parameters  
554 (77).

555

556 Consensus fasta sequences of the ancient mtDNA genomes from this and previous studies were  
557 aligned together with previously published modern sequences (26, 33, 78–91) (Table S4 and S6)  
558 using MEGA X (92). Multiple sequence alignments of mitochondrial genomes of the same  
559 haplogroup (A, B, C or D) were further visually inspected and edited (i.e. removing common  
560 indels, as well as to confirm variants and making sure sequences remained at the same length), to  
561 assure an accurate alignment. Common hypervariable sites that are phylogenetically uninformative  
562 were excluded from the analysis (309, 315, 515-522 AC indels, 3107, 16182-16183, 16193, and  
563 16519) (33). mtDNA sequences were excluded from the alignments based on two criteria: 1) >  
564 200 "N"s" as observed in fasta files or 2) Visual evaluation of an excess of C > T mutations that  
565 cannot be explained by random mutation. Four of the mtDNA sequences (333B\_TOL\_a,  
566 333O\_TOL\_b, E8\_CdV\_b and, BC30) were excluded from EBSP analysis but not from haplotype  
567 network analysis (Table S4), as their high number of N's, would have likely resulted in  
568 overestimation of the mutation rate and biased inferences in ESBP analyses.

569

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

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

577  
578 Past female effective population sizes were reconstructed using a Bayesian skyline approach (94)  
579 in BEAST 2 (95). The ancient mtDNA sequences, mentioned above were merged with present-  
580 day mtDNA sequences collected from published public databases (Table S4 and S6). Alignments  
581 were partitioned into five concatenated regions in the following order (Dloop, Coding, rRNA, and  
582 tRNA) as proposed by (81). HKY+G was found to be the best substitution model for this  
583 arrangement of the data using PartitionFinder 2 software (96). Substitution rates were estimated  
584 following a strict molecular clock model starting from point estimations as in (97). Additionally,  
585 we used tip calibrations for ancient samples using dating estimates (Table S1). We ran Markov  
586 Chain Monte Carlo (MCMC) chains of 100 million steps for each haplogroup alignment with a  
587 sampling of parameters every 10,000 generations, discarding the first 10 million steps as burn-in.  
588 Extended Bayesian skyline analysis was plotted using SkyViz reported in (56). Two independent  
589 runs of EBSP were run, showing a similar behavior (see Table S7 for statistics of each run). Figures  
590 S6A-N correspond to the first of the two EBSP runs.

#### 591 Sex assignment

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

#### 597 Contamination estimation based on the X chromosome

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

#### 605 Y-chromosome haplogroup inference

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

#### 613 Autosomal reference panels

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

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

627  
628 For the analysis within Mexico, we used the reference panel including only the Indigenous  
629 populations from Mexico (without the other three continental populations: YRI, CEU, and CHB)  
630 and keeping the individuals with >90% of Native American genetic ancestry. In total, this  
631 reference panel included genetic information of 561,327 SNVs from 268 individuals (Table S5).

632 We also integrated our generated ancient genome-wide data with previously reported low-depth  
633 genomes (0.09 – 0.3 x) from three ancient Pericúes from Aridoamerica reported in (78, 79),  
634 resulting in a total of 15 pre-Hispanic individuals from Mexico. Our analysis also included ancient  
635 genomic data from previous studies for 21 individuals from across the Americas (Table S3)(46,  
636 78, 79, 105–109).

#### 637 638 Pseudo-haploid calls

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

#### 648 649 Relatedness inference

650 Relatedness between ancient individuals from the same site was assessed using READ (110). A  
651 normalization step with a panel of non-related individuals is required previous to estimating the  
652 relationship in the ancient individuals, this step was performed with all Nahuas from the reference  
653 panel as they are the biggest group in the reference panel. Then, the normalization value was used  
654 to run READ with the ancient individuals. Two individuals from Toluquilla (333C\_TOL\_a and  
655 333B\_TOL\_a) were identified as first-degree relatives. Thus, the one with the highest coverage  
656 (333B\_TOL\_a) was used for downstream analyses. Furthermore, individuals E4\_CdV\_b and  
657 E8\_CdV\_b were identified as second-degree relatives, as well as E7\_CdV\_b and E19\_CdV\_b.

#### 658 659 ADMIXTURE analysis

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

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

674

#### 675 Missing DNA PCA

676 Missing DNA PCA (mdPCA) is a principal component analysis that corrects for genotypes missing  
677 due to genetic ancestry masking or degradation in ancient DNA samples. The correction is  
678 performed by comparing genetic distances between all samples. The genetic distance is computed  
679 as the average number of pairwise differences ( $\pi$  as defined in (112)). The manual of the method  
680 and the description of the parameters can be found in the mdPCA GitHub repository  
681 <https://github.com/AI-sandbox/mdPCA>. The dataset included 15 ancient samples (12 from this  
682 study and 3 Pericúes previously published) and the reference panel including present-day  
683 Indigenous individuals from Mexico. Genotyping data in plink format was converted to an  
684 unphased VCF format with Plink version 1.90 beta (102). The method used this VCF file as an  
685 input with the following parameters: each individual was plotted as an average of both parental  
686 haplotypes (AVERAGE\_PARENTS=True), weights for each individual were used inversely  
687 proportional to the number of samples from the corresponding population  
688 (IS\_WEIGHTED=True), the simple weighted of covariance PCA without any optimization was  
689 used (METHOD=1), no genotype was masked based on local genetic ancestry calls  
690 (IS\_MASKED=False), and only the three first PC's were calculated (NUM\_DIMS=3). Individuals  
691 with  $<0.15x$  whole genome coverage (19,851 SNVs or less) were further down-weighted to  $0.1x$   
692 to avoid having the principal components be selected based on noise variance stemming from these  
693 samples. These included BC30, MOM6\_ST\_a, 11R\_R\_b, E4\_CdV\_b, E7\_CdV\_b, E19\_CdV\_b,  
694 E2\_Mich\_b, and E4\_Mich\_b.

695

#### 696 Principal component analysis

697 We performed principal component analyses (PCA) using smartpca from the software eigensoft  
698 v6.0.1 (35) and projecting the individuals (lsqproject: YES) on the PCs estimated for the modern  
699 populations. The PCA (Fig. S7) with only the ancient individuals as well as the present-day  
700 Indigenous populations from Mexico, only included present-day individuals masked and with  
701  $>90\%$  Native American genetic ancestry.

702

#### 703 Temporal Factor Analysis

704 Temporal Factor Analysis (TFA) (39) were run with the reference panel that includes only the  
705 Indigenous populations from Mexico. The ancient samples included in the TFA consisted  
706 exclusively of individuals from Mexican territory from this study with  $>1x$  genome-wide  
707 coverage: one individual from Sierra Tarahumara (F9\_ST\_b), and three from Toluquilla  
708 (2417Q\_TOL\_b, 2417J\_TOL\_a, and 333B\_TOL\_a). Genotyping data was manipulated with  
709 Plink version 1.90 beta (102). The dataset was converted from Plink format to Eigenstrat's geno  
710 format with the command convertf from Eigensoft version 6.0.1.1 (35). Missing genotypes were  
711 imputed with the snmf function from the LEA package, using the parameters  $K=2$ , entropy=  
712 TRUE, and repetitions=5. The runs with the lowest cross-entropy value were considered for

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

725

### 726 Outgroup f3 and D statistics

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

734

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

746

### 747 Population continuity test between F9\_ST\_a and present-day Rarámuri

748 According to the PCA, TFA, ADMIXTURE, outgroup-f3 and D statistics, individual F9\_ST\_a  
749 from north Mexico is significantly closer to present-day Rarámuri. Thus, we analyzed whether  
750 individual F9 belonged to the present-day Rarámuri ancestral population using the method  
751 published in (40). Derived alleles in individual F9 and present-day Rarámuri were identified using  
752 ancestral alleles' information from  
753 ([ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/pilot\\_data/technical/reference/ancestral\\_alignments](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/pilot_data/technical/reference/ancestral_alignments)).

754 Then, allele counts were performed for each site in the present-day Rarámuri (reference  
755 population), excluding alleles with frequency 0 or 1. We performed allele counts for the pre-  
756 Hispanic individual F9. Parameters alpha and beta necessary for the analysis were estimated using  
757  $\text{min\_samples} = 14$ , which refers to the minimum number of individuals required in the reference  
758 panel that have information for each SNV. Under the null hypothesis, it is assumed that reference

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

761

### 762 qpGraph

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

777

### 778 qpWave

779 We used qpWave version 410 (45) to test whether pre- and post-drought Sierra Gorda (SG)  
780 individuals could be modeled using either SG individuals or other ancient or modern Mexican  
781 indigenous population relative to a set of reference groups (Mbuti, Onge, USR1; Papuan, Anzick,  
782 Han, Mixe, MA1, Zapotec, Karitiana, Piapoco) as in (114). In order to maximize the power for  
783 this test we increased the number of SNVs to 2,224,359k transversion sites ascertained in 1KGP  
784 and having a 5% MAF. Present-day human sequencing data from HGDP genomes (23) was used  
785 to represent Mexican indigenous genetic variation. With these data, we tested whether we could  
786 model the genetic variation of pre (2417Q\_TOL\_b) and post-drought Sierra Gorda (2417J\_TOL\_a  
787 and 333B\_TOL\_a) individually, using as a source either another Sierra Gorda individual or another  
788 Mexican indigenous population. For the latter case we used as references the pre-Hispanic  
789 individual F9\_ST\_a, and Pericúes, along with present-day Akimel O'odham and Mayan. If pre-  
790 and post-drought Sierra Gorda individuals arise from different demographic histories to each other,  
791 then when analyzed together it would be required to add one or more different sources of genetic  
792 variation to explain the data, which would be denoted by a p-value  $< 0.05$ . We also performed  
793 qpWave analysis with the panel of Indigenous populations from Mexico reported in (104) to  
794 increase the number of populations, though at the cost of reducing the number of overlapping sites.  
795 Results are consistent with those obtained using the HGDP Akimel O'odham and Mayan in that  
796 post-drought genetic variation can be explained with pre-drought genetic variation, suggesting  
797 genetic continuity. However, using the Moreno-Estrada reference panel resulted in a low number  
798 of overlapping SNVs for the qpWave analysis and, consequently, some p-values could not be  
799 computed by the program (e.g. when using the pre and post-drought samples as both target (Pop1)  
800 and sources (Pop2) (Table S18) .

801

### 802 Conditional nucleotide diversity

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

805 the individuals per archeological site. This analysis included 16 pre-Hispanic individuals  
806 (Pericúes, Sierra Tarahumara, Sierra Gorda, Cañada de la Virgen and Patagonia) with at least  
807 10,000 SNVs overlapping between them (Table S11 and S12) and present-day populations from  
808 Mexico (19, 23).

809  
810 This consists of counting the differences between the ascertained alleles present in one pair of  
811 individuals from the same population. For this analysis, we used a panel of 1,938,919 transversions  
812 SNVs with a minor allele frequency of 0.01 in the YRI population. In all cases, pseudohaploid  
813 genotype calls were performed before the comparisons. CNV values were calculated for all pairs  
814 of individuals for each present-day Indigenous population and for all pairs of pre-Hispanic  
815 individuals within the same site. For present-day populations we masked the sites with no Native  
816 American genetic ancestry to avoid counting levels of diversity from a different genetic ancestry.  
817 Mean and standard error values were estimated 100 times using a window size of 50 Kb even for  
818 pairs of ancient individuals. Results are reported in a violin plot, where each dot inside the violin  
819 corresponds to the CNV value for a unique pair of individuals (Fig. S19).

820

### 821 Runs of homozygosity

822 We used the tool hapROH (42) to detect runs of homozygosity in pre-Hispanic individuals. The  
823 distribution of ROHs was used as an indirect way of measuring genetic diversity and effective  
824 population size. This method detects the ROH's lengths longer than four centimorgans(cM) and  
825 up to 300 cM. The sum of the ROH detected gives insights into the relatedness of the parents of  
826 each individual and informs about population sizes. Length of ROH and relatedness between  
827 parents are positively correlated. While the higher the number of ROH segments, the smaller the  
828 population size and the smaller the number of heterozygote sites, which means less genetic  
829 diversity. hapROH uses a built-in set of 1240K SNVs from to perform pseudohaploid genotype  
830 calls and a reference panel of 5,008 present-day human haplotypes from (54). We performed a  
831 pseudohaploid variant calling for the 1240K SNVs for each pre-Hispanic genome. hapROH's  
832 manual recommends using samples with a minimum of 400K SNVs intersected with the 1240K  
833 SNVs. However, we also included the individual E8\_CdV\_b, which intersected 369,504 SNVs,  
834 and the Pericú individual B03, which intersected 370,026 SNVs, to have at least one individual of  
835 these sites in the analyses and considering the number of SNVs is not too far from 400K (Table  
836 S13). Detection of runs of homozygosity was made using the set of 5,008 haplotypes mentioned  
837 above.

838

839

### 840 **References:**

- 841 1. A. L. Austin, L. L. Luján, *El pasado indígena* (Fondo de Cultura Económica, 2019).
- 842 2. B. B. Cornejo, *La frontera septentrional de Mesoamérica* (Instituto Nacional de  
843 Antropología e Historia, México, ed. 2, 2000), vol. 1 of *Historia Antigua de México*.
- 844 3. P. Kirchhoff, Mesoamerica: its geographic limits, ethnic composition and cultural  
845 characteristics. *Heritage of conquest*, 17–30 (1952).
- 846 4. M. del C. Solanes, E. Vela, Atlas del México prehispánico. *Arqueología Mexicana*. **5**, 5–  
847 18 (2000).
- 848 5. D. W. Stahle, J. V. Diaz, D. J. Burnette, J. C. Paredes, R. R. Heim, F. K. Fye, *et al.*,  
849 Major Mesoamerican droughts of the past millennium: MESOAMERICAN DROUGHTS.



- 850 *Geophysical Research Letters*. **38** (2011) (available at  
851 <http://doi.wiley.com/10.1029/2010GL046472>).
- 852 6. M. K. Hughes, H. F. Diaz, Was there a ‘medieval warm period’, and if so, where and  
853 when? *Climatic Change*. **26**, 109–142 (1994).
- 854 7. P. Armillas, Condiciones ambientales y movimientos de pueblos en la frontera  
855 septentrional de Mesoamérica. *Homenaje a Fernando Márquez-Miranda*, 62–82 (1964).
- 856 8. K. H. Wogau, H. W. Arz, H. N. Böhnelt, N. R. Nowaczyk, J. Park, High resolution  
857 paleoclimate and paleoenvironmental reconstruction in the Northern Mesoamerican Frontier for  
858 Prehistory to Historical times. *Quaternary Science Reviews*. **226**, 106001 (2019).
- 859 9. D. J. Kennett, S. F. M. Breitenbach, V. V. Aquino, Y. Asmerom, J. Awe, J. U. L. Baldini,  
860 *et al.*, Development and Disintegration of Maya Political Systems in Response to Climate  
861 Change. *Science*. **338**, 788–791 (2012).
- 862 10. J. A. Hoggarth, S. F. M. Breitenbach, B. J. Culleton, C. E. Ebert, M. A. Masson, D. J.  
863 Kennett, The political collapse of Chichén Itzá in climatic and cultural context. *Global and*  
864 *Planetary Change*, 18 (2016).
- 865 11. S. L. O’Hara, S. E. Metcalfe, F. A. Street-Perrott, On the arid margin: The relationship  
866 between climate, humans and the environment. A review of evidence from the highlands of  
867 central Mexico. *Chemosphere*. **29**, 965–981 (1994).
- 868 12. B. Braniff, Oscilación de la frontera norte mesoamericana: un nuevo ensayo.  
869 *Arqueología, segunda época*, 99–114 (1989).
- 870 13. A. Gómez Aí-za, S. Sánchez Vázquez, La Frontera Cultural Meso-Aridoamericana:  
871 Construcción De Imaginarios Nacionalistas En La Historia Mexicana. *Xih*. **9** (2014),  
872 doi:10.37646/xihmai.v9i18.244.
- 873 14. M. Raghavan, M. Steinrucken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, *et*  
874 *al.*, Genomic evidence for the Pleistocene and recent population history of Native Americans.  
875 *Science*. **349**, aab3884–aab3884 (2015).
- 876 15. C. L. Scheib, H. Li, T. Desai, V. Link, C. Kendall, G. Dewar, *et al.*, Ancient human  
877 parallel lineages within North America contributed to a coastal expansion. *Science*. **360**, 1024–  
878 1027 (2018).
- 879 16. M. Raghavan, M. Steinrucken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, *et*  
880 *al.*, Genomic evidence for the Pleistocene and recent population history of Native Americans.  
881 *Science*. **349**, aab3884–aab3884 (2015).
- 882 17. M. C. Ávila-Arcos, C. de la Fuente Castro, M. A. Nieves-Colón, M. Raghavan,  
883 Recommendations for Sustainable Ancient DNA Research in the Global South: Voices From a  
884 New Generation of Paleogenomicists. *Front. Genet*. **13**, 880170 (2022).
- 885 18. Materials and methods are available as supplementary materials on Science Online.
- 886 19. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, *et al.*, The Simons  
887 Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. **538**, 201–206  
888 (2016).
- 889 20. I. De La Cruz, A. González-Oliver, B. M. Kemp, J. A. Román, D. G. Smith, A. Torre-  
890 Blanco, Sex Identification of Children Sacrificed to the Ancient Aztec Rain Gods in Tlatelolco.  
891 *Current Anthropology*. **49**, 519–526 (2008).
- 892 21. D. Perez-Benedico, J. La Salvia, Z. Zeng, G. A. Herrera, R. Garcia-Bertrand, R. J.  
893 Herrera, Mayans: a Y chromosome perspective. *Eur J Hum Genet*. **24**, 1352–1358 (2016).
- 894 22. K. Sandoval, A. Moreno-Estrada, I. Mendizabal, P. A. Underhill, M. Lopez-Valenzuela,  
895 R. Peñaloza-Espinosa, *et al.*, Y-chromosome diversity in Native Mexicans reveals continental

- 896 transition of genetic structure in the Americas. *Am. J. Phys. Anthropol.* **148**, 395–405 (2012).
- 897 23. A. Bergström, S. A. McCarthy, R. Hui, M. A. Almarri, Q. Ayub, P. Danecek, *et al.*,
- 898 Insights into human genetic variation and population history from 929 diverse genomes. *Science*.
- 899 **367**, eaay5012 (2020).
- 900 24. B. Z. González-Sobrino, A. P. Pintado-Cortina, L. Sebastián, F. Morales-Mandujano, A.
- 901 V. Contreras, J. Chávez-Benavides, *et al.*, Genetic Diversity and Differentiation in Urban and
- 902 Indigenous Populations of Mexico: Patterns of Mit, 21.
- 903 25. H.-J. Bandelt, C. Herrnstadt, Y.-G. Yao, Q.-P. Kong, T. Kivisild, C. Rengo, *et al.*,
- 904 Identification of Native American Founder mtDNAs Through the Analysis of Complete mtDNA
- 905 Sequences: Some Caveats. *Ann Human Genet.* **67**, 512–524 (2003).
- 906 26. A. Y. Morales-Arce, M. H. Snow, J. H. Kelley, M. Anne Katzenberg, Ancient
- 907 mitochondrial DNA and ancestry of Paquimé inhabitants, Casas Grandes (A.D. 1200–1450).
- 908 *American Journal of Physical Anthropology.* **163**, 616–626 (2017).
- 909 27. J. Mata-Míguez, L. Overholtzer, E. Rodríguez-Alegría, B. M. Kemp, D. A. Bolnick, The
- 910 genetic impact of aztec imperialism: Ancient mitochondrial DNA evidence from Xaltocan,
- 911 Mexico. *American Journal of Physical Anthropology.* **149**, 504–516 (2012).
- 912 28. A. Y. Morales-Arce, G. McCafferty, J. Hand, N. Schmill, K. McGrath, C. Speller,
- 913 Ancient mitochondrial DNA and population dynamics in postclassic Central Mexico: Tlatelolco
- 914 (ad 1325–1520) and Cholula (ad 900–1350). *Archaeological and Anthropological Sciences*
- 915 (2019), doi:10.1007/s12520-018-00771-7.
- 916 29. A. González-Oliver, L. Márquez-Morfín, J. C. Jiménez, A. Torre-Blanco, Founding
- 917 Amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo: mtDNA
- 918 Lineages in Ancient Maya. *American Journal of Physical Anthropology.* **116**, 230–235 (2001).
- 919 30. A. Y. Morales-Arce, C. A. Hofman, A. T. Duggan, A. K. Benfer, M. A. Katzenberg, G.
- 920 McCafferty, *et al.*, Successful reconstruction of whole mitochondrial genomes from ancient
- 921 Central America and Mexico. *Scientific Reports.* **7**, 18100 (2017).
- 922 31. B. M. Kemp, A. Gonzalez-Oliver, R. S. Malhi, C. Monroe, K. B. Schroeder, J.
- 923 McDonough, *et al.*, Evaluating the Farming/Language Dispersal Hypothesis with genetic
- 924 variation exhibited by populations in the Southwest and Mesoamerica. *Proceedings of the*
- 925 *National Academy of Sciences.* **107**, 6759–6764 (2010).
- 926 32. M. Bodner, U. A. Perego, J. E. Gomez, R. M. Cerda-Flores, N. Rambaldi Migliore, S. R.
- 927 Woodward, *et al.*, The Mitochondrial DNA Landscape of Modern Mexico. *Genes.* **12**, 1453
- 928 (2021).
- 929 33. M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human
- 930 mitochondrial DNA variation. *Human Mutation.* **30**, E386–E394 (2009).
- 931 34. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in
- 932 unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- 933 35. N. Patterson, A. L. Price, D. Reich, Population Structure and Eigenanalysis. *PLoS*
- 934 *Genetics.* **2**, e190 (2006).
- 935 36. D. Reich, A. L. Price, N. Patterson, Principal component analysis of genetic data. *Nature*
- 936 *Genetics.* **40**, 491–492 (2008).
- 937 37. A. Moreno-Estrada, C. R. Gignoux, J. C. Fernandez-Lopez, F. Zakharia, M. Sikora, A. V.
- 938 Contreras, *et al.*, The genetics of Mexico recapitulates Native American substructure and affects
- 939 biomedical traits. *Science.* **344**, 1280–1285 (2014).
- 940 38. The 1000 Genomes Project Consortium, A global reference for human genetic variation.
- 941 *Nature.* **526**, 68–74 (2015).

- 942 39. O. François, F. Jay, Factor analysis of ancient population genomic samples. *Nat*  
943 *Commun.* **11**, 4661 (2020).
- 944 40. J. G. Schraiber, Assessing the Relationship of Ancient and Modern Populations.  
945 *Genetics.* **208**, 383–398 (2018).
- 946 41. P. Skoglund, H. Malmstrom, A. Omrak, M. Raghavan, C. Valdiosera, T. Gunther, *et al.*,  
947 Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers.  
948 *Science.* **344**, 747–750 (2014).
- 949 42. H. Ringbauer, J. Novembre, M. Steinrücken, Parental relatedness through time revealed  
950 by runs of homozygosity in ancient DNA. *Nat Commun.* **12**, 5425 (2021).
- 951 43. D. I. Cruz Dávalos, Y. O. Arizmendi Cárdenas, M. J. Bravo-Lopez, S. Neuenschwander,  
952 S. Reis, M. Q. R. Bastos, *et al.*, “Indigenous peoples in eastern Brazil: insights from 19<sup>th</sup> century  
953 genomes and metagenomes” (preprint, Genomics, 2022), , doi:10.1101/2022.01.27.477466.
- 954 44. M. C. Ávila-Arcos, K. F. McManus, K. Sandoval, J. E. Rodríguez-Rodríguez, V. Villa-  
955 Islas, A. R. Martin, *et al.*, Population History and Gene Divergence in Native Mexicans Inferred  
956 from 76 Human Exomes. *Molecular Biology and Evolution.* **37**, 994–1006 (2020).
- 957 45. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, *et al.*, Ancient  
958 Admixture in Human History. *Genetics.* **192**, 1065–1093 (2012).
- 959 46. J. V. Moreno-Mayar, B. A. Potter, L. Vinner, M. Steinrücken, S. Rasmussen, J. Terhorst,  
960 *et al.*, Terminal Pleistocene Alaskan genome reveals first founding population of Native  
961 Americans. *Nature.* **553**, 203–207 (2018).
- 962 47. M. Rasmussen, S. L. Anzick, M. R. Waters, P. Skoglund, M. DeGiorgio, T. W. Stafford,  
963 *et al.*, The genome of a Late Pleistocene human from a Clovis burial site in western Montana.  
964 *Nature.* **506**, 225–229 (2014).
- 965 48. J. V. Moreno-Mayar, L. Vinner, P. de Barros Damgaard, C. de la Fuente, J. Chan, J. P.  
966 Spence, *et al.*, Early human dispersals within the Americas. *Science.* **362**, eaav2621 (2018).
- 967 49. H. García-Ortiz, *et al.*, The genomic landscape of Mexican Indigenous populations brings  
968 insights into the peopling of the Americas. *Nat Commun.* **12**, 5942 (2021).
- 969 50. H. García-Ortiz, The genomic landscape of Mexican Indigenous populations brings  
970 insights into the peopling of the Americas, 12.
- 971 51. A. Gorostiza, V. Acunha-Alonzo, L. Regalado-Liu, S. Tirado, J. Granados, D. Sámano, *et*  
972 *al.*, Reconstructing the History of Mesoamerican Populations through the Study of the  
973 Mitochondrial DNA Control Region. *PLoS ONE.* **7**, e44666 (2012).
- 974 52. S. Romero-Hidalgo, Demographic history and biologically relevant genetic variation of  
975 Native Mexicans inferred from whole-genome sequencing. *NATURE COMMUNICATIONS*, 8.
- 976 53. J. Ramos-Madriral, B. D. Smith, J. V. Moreno-Mayar, S. Gopalakrishnan, J. Ross-Ibarra,  
977 M. T. P. Gilbert, *et al.*, Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into  
978 the Early Stages of Maize Domestication. *Current Biology.* **26**, 3195–3201 (2016).
- 979 54. M. A. Nieves-Colón, W. J. Pestle, A. W. Reynolds, B. Llamas, C. de la Fuente, K.  
980 Fowler, *et al.*, Ancient DNA Reconstructs the Genetic Legacies of Precontact Puerto Rico  
981 Communities. *Molecular Biology and Evolution.* **37**, 611–626 (2020).
- 982 55. B. Llamas, L. Fehren-Schmitz, G. Valverde, J. Soubrier, S. Mallick, N. Rohland, *et al.*,  
983 Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas.  
984 *Science Advances.* **2**, e1501385 (2016).
- 985 56. B. D. O’Fallon, L. Fehren-Schmitz, Native Americans experienced a strong population  
986 bottleneck coincident with European contact. *Proceedings of the National Academy of Sciences.*  
987 **108**, 20444–20448 (2011).

- 988 57. A. J. Herrera Muñoz, E. Mejía Campos, Un minero en la Sierra Gorda: caso de  
989 contaminación ocupacional multielemental de metales pesados a finales del periodo Clásico.  
990 *Arqueología (México, DF)*, 127–135 (2017).
- 991 58. E. Nalda, La arqueología de Guanajuato: trabajos recientes. *Arqueología mexicana*. **16**,  
992 36–43 (2008).
- 993 59. M. E. Smith, Mesoamerica’s First World City: Teotihuacan in Comparative Context.  
994 *Teotihuacan: The World Beyond the City* (2020).
- 995 60. B. A. Álvarez-Sandoval, L. R. Manzanilla, M. González-Ruiz, A. Malgosa, R. Montiel,  
996 Genetic Evidence Supports the Multiethnic Character of Teopancazco, a Neighborhood Center of  
997 Teotihuacan, Mexico (AD 200-600). *PLoS ONE*. **10**, e0132371 (2015).
- 998 61. E. Campos, J. Campos, A. Muñoz, Mercurialism Determination in Fetuses Bone Remains  
999 from Toluquilla, Queretaro, Mexico. *JSRR*. **8**, 1–10 (2015).
- 1000 62. E. M. Pérez Campos, A. J. Herrera Muñoz, MINAS Y MINEROS: PRESENCIA DE  
1001 METALES EN SEDIMENTOS Y RESTOS HUMANOS AL SUR DE LA SIERRA GORDA  
1002 DE QUERÉTARO EN MÉXICO. *Chungará (Arica)*. **45**, 161–176 (2013).
- 1003 63. D. A. Corona, Expansión Territorial Comercial en Mesoamérica y Mesoamérica  
1004 Septentrional por medio del Intercambio. **19**, 10 (2015).
- 1005 64. N. Rohland, M. Hofreiter, Ancient DNA extraction from bones and teeth. *Nat Protoc*. **2**,  
1006 1756–1762 (2007).
- 1007 65. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, *et al.*,  
1008 Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from  
1009 ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*. **110**, 15758–15763  
1010 (2013).
- 1011 66. M. Meyer, M. Kircher, Illumina Sequencing Library Preparation for Highly Multiplexed  
1012 Target Capture and Sequencing. *Cold Spring Harbor Protocols*. **2010** (2010) (available at  
1013 <http://www.cshprotocols.org/cgi/doi/10.1101/pdb.prot5448>).
- 1014 67. T. Daley, A. D. Smith, Predicting the molecular complexity of sequencing libraries.  
1015 *Nature Methods*. **10**, 325–327 (2013).
- 1016 68. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming,  
1017 identification, and read merging. *BMC Research Notes*. **9**, 88 (2016).
- 1018 69. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, *et al.*, The Sequence  
1019 Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
- 1020 70. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, *et al.*,  
1021 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA  
1022 sequencing data. *Genome Research*. **20**, 1297–1303 (2010).
- 1023 71. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast  
1024 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. **29**, 1682–  
1025 1684 (2013).
- 1026 72. V. Link, A. Kousathanas, K. Veeramah, C. Sell, A. Scheu, D. Wegmann, “ATLAS:  
1027 Analysis Tools for Low-depth and Ancient Samples” (preprint, Bioinformatics, 2017), ,  
1028 doi:10.1101/105346.
- 1029 73. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: estimation of contamination and  
1030 endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology*. **16**, 224 (2015).
- 1031 74. H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt,  
1032 *et al.*, HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput  
1033 sequencing. *Nucleic Acids Res*. **44**, W58–W63 (2016).

- 1034 75. T. A. Hall, in *Nucleic acids symposium series* ([London]: Information Retrieval Ltd.,  
1035 c1979-c2000., 1999), vol. 41, pp. 95–98.
- 1036 76. J. Rozas, A. Ferrer-Mata, J. C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado, S. E.  
1037 Ramos-Onsins, *et al.*, DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets.  
1038 *Molecular Biology and Evolution*. **34**, 3299–3302 (2017).
- 1039 77. L. Excoffier, H. E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform  
1040 population genetics analyses under Linux and Windows. *Molecular ecology resources*. **10**, 564–  
1041 567 (2010).
- 1042 78. M. Raghavan, *et. al*, Genomic evidence for the Pleistocene and recent population history  
1043 of Native Americans. **349**, aab3884–aab3884 (2015).
- 1044 79. C. L. Scheib, *et al.*, Ancient human parallel lineages within North America contributed to  
1045 a coastal expansion. *Science*. **360**, 1024–1027 (2018).
- 1046 80. S. Mallick, *et. al*, The Simons Genome Diversity Project: 300 genomes from 142 diverse  
1047 populations. *Nature*. **538**, 201–206 (2016).
- 1048 81. B. Llamas, *et al.*, Ancient mitochondrial DNA provides high-resolution time scale of the  
1049 peopling of the Americas. *Science Advances*. **2**, e1501385 (2016).
- 1050 82. M. Van Oven, PhyloTree Build 17: Growing the human mitochondrial DNA tree.  
1051 *Forensic Science International: Genetics Supplement Series*. **5**, e392–e394 (2015).
- 1052 83. S. Kumar, C. Bellis, M. Zlojutro, P. E. Melton, J. Blangero, J. E. Curran, Large scale  
1053 mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American  
1054 origins. *BMC evolutionary biology*. **11**, 1–17 (2011).
- 1055 84. U. A. Perego, A. Achilli, N. Angerhofer, M. Accetturo, M. Pala, A. Olivieri, *et al.*,  
1056 Distinctive Paleo-Indian Migration Routes from Beringia Marked by Two Rare mtDNA  
1057 Haplogroups. *Current Biology*. **19**, 1–8 (2009).
- 1058 85. A. Achilli, U. A. Perego, H. Lancioni, A. Olivieri, F. Gandini, B. H. Kashani, *et al.*,  
1059 Reconciling migration models to the Americas with the variation of North American native  
1060 mitogenomes. *Proceedings of the National Academy of Sciences*. **110**, 14308–14313 (2013).
- 1061 86. A. Achilli, A. Olivieri, M. Pala, B. Hooshiar Kashani, V. Carossa, U. A. Perego, *et al.*,  
1062 Mitochondrial DNA Backgrounds Might Modulate Diabetes Complications Rather than T2DM  
1063 as a Whole. *PLoS ONE*. **6**, e21029 (2011).
- 1064 87. U. A. Perego, N. Angerhofer, M. Pala, A. Olivieri, H. Lancioni, B. H. Kashani, *et al.*, The  
1065 initial peopling of the Americas: A growing number of founding mitochondrial genomes from  
1066 Beringia. *Genome Research*. **20**, 1174–1179 (2010).
- 1067 88. D. M. Behar, M. van Oven, S. Rosset, M. Metspalu, E.-L. Loogväli, N. M. Silva, *et al.*, A  
1068 “Copernican” Reassessment of the Human Mitochondrial DNA Tree from its Root. *The*  
1069 *American Journal of Human Genetics*. **90**, 675–684 (2012).
- 1070 89. A. Hartmann, M. Thieme, L. K. Nanduri, T. Stempfl, C. Moehle, T. Kivisild, *et al.*,  
1071 Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum.*  
1072 *Mutat.* **30**, 115–122 (2009).
- 1073 90. R. S. Just, T. M. Diegoli, J. L. Saunier, J. A. Irwin, T. J. Parsons, Complete mitochondrial  
1074 genome sequences for 265 African American and U.S. “Hispanic” individuals. *Forensic Science*  
1075 *International: Genetics*. **2**, e45–e48 (2008).
- 1076 91. M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, Mitochondrial genome variation and  
1077 the origin of modern humans. *Nature*. **408**, 708–713 (2000).
- 1078 92. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary  
1079 genetics analysis across computing platforms. *Molecular biology and evolution*. **35**, 1547 (2018).

- 1080 93. J. W. Leigh, D. Bryant, POPART : full-feature software for haplotype network  
1081 construction. *Methods Ecol Evol.* **6**, 1110–1116 (2015).
- 1082 94. J. Heled, A. J. Drummond, Bayesian inference of population size history from multiple  
1083 loci. *BMC Evol Biol.* **8**, 289 (2008).
- 1084 95. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, *et al.*, BEAST 2: A  
1085 Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput Biol.* **10**, e1003537  
1086 (2014).
- 1087 96. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2: New  
1088 Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological  
1089 Phylogenetic Analyses. *Mol Biol Evol*, msw260 (2016).
- 1090 97. P. Soares, L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Röhl, *et al.*, Correcting for  
1091 Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *The American*  
1092 *Journal of Human Genetics.* **84**, 740–759 (2009).
- 1093 98. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of  
1094 ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science.* **40**,  
1095 4477–4482 (2013).
- 1096 99. Y. Huang, H. Ringbauer, hapCon: estimating contamination of ancient genomes by  
1097 copying from reference haplotypes. *Bioinformatics.* **38**, 3768–3777 (2022).
- 1098 100. The 1000 Genomes Project Consortium, D. G. Poznik, *et al.*, Punctuated bursts in human  
1099 male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics.*  
1100 **48**, 593–599 (2016).
- 1101 101. The 1000 Genomes Project Consortium, An integrated map of genetic variation from  
1102 1,092 human genomes. *Nature.* **491**, 56–65 (2012).
- 1103 102. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, *et al.*,  
1104 PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.  
1105 *The American Journal of Human Genetics.* **81**, 559–575 (2007).
- 1106 103. A. Moreno-Estrada, *et al.*, The genetics of Mexico recapitulates Native American  
1107 substructure and affects biomedical traits. *Science.* **344**, 1280–1285 (2014).
- 1108 104. B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: A Discriminative  
1109 Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of*  
1110 *Human Genetics.* **93**, 278–288 (2013).
- 1111 105. M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, *et al.*,  
1112 Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature.* **463**, 757–762 (2010).
- 1113 106. P. Flegontov, N. E. Altınışık, P. Changmai, N. Rohland, S. Mallick, N. Adamski, *et al.*,  
1114 Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature.* **570**,  
1115 236–240 (2019).
- 1116 107. M. Rasmussen, *et al.*, The genome of a Late Pleistocene human from a Clovis burial site  
1117 in western Montana. *Nature.* **506**, 225–229 (2014).
- 1118 108. C. Posth, N. Nakatsuka, I. Lazaridis, P. Skoglund, S. Mallick, T. C. Lamnidis, *et al.*,  
1119 Reconstructing the Deep Population History of Central and South America. *Cell.* **175**, 1185-  
1120 1197.e22 (2018).
- 1121 109. C. de la Fuente, M. C. Ávila-Arcos, J. Galimany, M. L. Carpenter, J. R. Homburger, A.  
1122 Blanco, *et al.*, Genomic insights into the origin and diversification of late maritime hunter-  
1123 gatherers from the Chilean Patagonia. *Proceedings of the National Academy of Sciences.* **115**,  
1124 E4006–E4012 (2018).
- 1125 110. J. M. Monroy Kuhn, M. Jakobsson, T. Günther, Estimating genetic kin relationships in

- 1126 prehistoric populations. *PLoS ONE*. **13**, e0195491 (2018).
- 1127 111. A. A. Behr, K. Z. Liu, G. Liu-Fang, P. Nakka, S. Ramachandran, pong: fast analysis and  
1128 visualization of latent clusters in population genetic data. *Bioinformatics*. **32**, 2817–2823 (2016).
- 1129 112. Nei, Masatoshi, *Molecular Evolutionary Genetics* (Columbia University Press, New  
1130 York).
- 1131 113. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, *et al.*, The Simons  
1132 Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*. **538**, 201–206  
1133 (2016).
- 1134 114. K. Nägele, C. Posth, M. Iraeta Orbeago, Y. Chinique de Armas, S. T. Hernández  
1135 Godoy, U. M. González Herrera, *et al.*, Genomic insights into the early peopling of the  
1136 Caribbean. *Science*. **369**, 456–460 (2020).
- 1137 115. The 1000 Genomes Project Consortium, R. A. Gibbs, E. Boerwinkle, H. Doddapaneni, Y.  
1138 Han, V. Korchina, *et al.*, A global reference for human genetic variation. *Nature*. **526**, 68–74  
1139 (2015).

1140  
1141

## 1142 Acknowledgments

1143

1144 **Acknowledgments:** We greatly acknowledge the IT support of Luis Alberto Aguilar  
1145 Bautista, Alejandro de León Cuevas, Carlos Sair Flores Bautista and Jair Garcia Sotelo of  
1146 the Laboratorio Nacional de Visualización Científica Avanzada (LAVIS UNAM). We also  
1147 thank Alejandra Castillo Carbajal and Carina Uribe Díaz from LIIGH-UNAM for wetlab  
1148 technical support during the project. We thank Alexandra Sockell for technical support at  
1149 the Bustamante Lab. We thank Leonardo Correa-Mendoza for technical support with  
1150 qpWave. We thank Diego Ortega-Del Vecchyo and Angélica González Oliver for  
1151 enriching comments throughout the project on earlier versions of the manuscripts. We  
1152 thank Yazmín Jiménez Islas for her help in the Summary figure. We also thank Rafael  
1153 Montiel Duarte for his early contribution in the Cañada de la Virgen project.

1154

1155 **Funding:** Viridiana Villa-Islas is a doctoral student from the Programa de Doctorado en Ciencias  
1156 Biomédicas, Universidad Nacional Autónoma de México (UNAM) and has received  
1157 CONACyT fellowship (Número de becaria: 282188). M.C.A.-A. received support from  
1158 UNAM-PAPIIT (Project numbers: IA201219 and IA203821) and the Wellcome Trust Seed  
1159 Award in Science (grant number: 208934/Z/17/Z), and the International Centre for Genetic  
1160 Engineering and Biotechnology (project number: CRP/MEX10-03). M.C.A.-A., E.H.-S.,  
1161 and F.J. receive support from the Human Frontiers Science Program (grant number:  
1162 RGY0075/201). R.F. was supported by a grant (PALEUCOL; PGC2018-094101-A-I00)  
1163 from FEDER/Ministerio de Ciencia e Innovación—Agencia Estatal de Investigación and  
1164 from CajaCanarias and La Caixa Banking foundations (GENPAC; 2018PATRI16). Data  
1165 generation for the Sierra Tarahumara Individuals was funded with a Trainee Research  
1166 Grant to M.C.A.-A. by Stanford Center for Computational, Evolutionary and Human  
1167 Genomics. E.H.-S. received support from NIH R35GM128946 and the Alfred P. Sloan  
1168 Foundation. Data generation for the Cañada de la Virgen individuals was supported by The  
1169 Mexican Biobank Project (Grants FONCICYT/50/2016 and MR/N028937/1 from  
1170 CONACyT and Newton Fund, respectively) awarded to A.M.-E. Also, this work has been  
1171 partially supported by the government of Guanajuato (Grant No. 093/2016 CINVESTAV)

1172 awarded to K.S. And, data generation for the Aridoamerican individuals was supported in  
1173 part by a Leakey Foundation granted to F.A.V.

1174  
1175 **Author contributions:** M.C.A.-A. and V.V.-I. conceived and designed the study with input from  
1176 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.  
1177 V.V.-I. processed all sequence data. V.V.-I. carried out most analyses with help from, A.I.-  
1178 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.-  
1179 M., M.V.-M, R.A.-H., A.G.-O., C.V. contributed samples, provided the archaeological  
1180 context and discussed results. F.A.V., A.M.-E., F.J., A.I., E.H.-S.V.M.-M., and F.S.-Q.  
1181 suggested analyses, discussed results and provided feedback throughout the project. V.V.-  
1182 I. and M.C.A.-A. wrote and edited the manuscript with input from all authors.

1183  
1184 **Competing interests:** The authors declare no competing interests.

1185  
1186 **Data and materials availability:** Data are available from the European Nucleotide  
1187 Archive under the project accession PRJEB51440.

1188  
1189  
1190



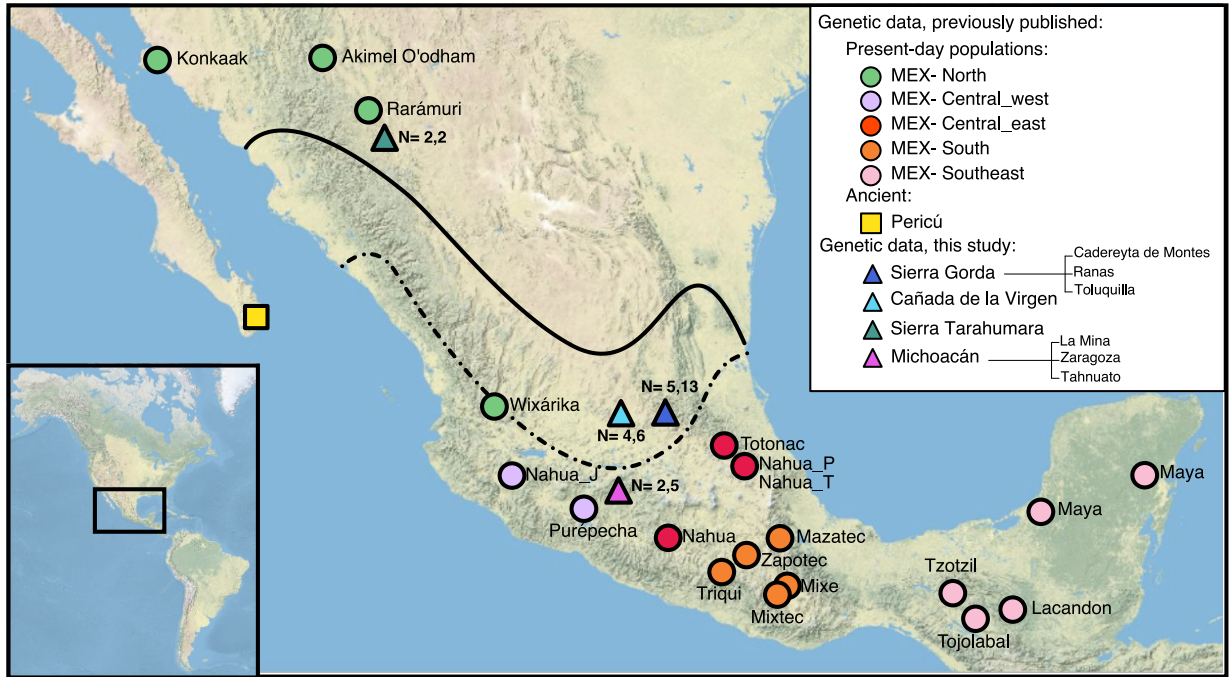
1191 **References Supplementary Material:**

- 1192
- 1193 116. J. Bracho, Narrativa e identidad. El mestizaje y su representación historiográfica. *LREL*,
- 1194 55 (2009).
- 1195 117. P. Wade, in *The Companion to Latin American Studies* (Routledge, 2014), pp. 197–211.
- 1196 118. E. Mejía Pérez Campos, X. Chávez Balderas, R. Chávez Sánchez, Pepita. La momia de la
- 1197 Sierra Gorda de Querétaro. *Arqueología mexicana*, 70–75 (2009).
- 1198 119. E. Mejía Pérez Campos, thesis, Escuela Nacional de Antropología e Historia, Ciudad de
- 1199 México, Méx (2010).
- 1200 120. G. Zepeda García Moreno, Cañada de la Virgen, refugio de los muertos y los ancestros.
- 1201 *Guanajuato, Ediciones La Rana* (2010).
- 1202 121. I. D. Lara Barajas, G. Islas Estrada, G. Zepeda García Moreno, “Informe Final del
- 1203 Análisis Antropofísico del Entierro 19,” *Proyecto Arqueológico Cañada de la Virgen* (Centro
- 1204 INAH-Querétaro, 2011).
- 1205 122. A. Goguitchaichvili, V. Ortega, J. Archer, J. Morales, A. T. Guerrero, Absolute
- 1206 geomagnetic intensity record from pre-Columbian pottery dates elite Tlailotlacan Woman in
- 1207 ancient Teotihuacan. *Journal of Archaeological Science: Reports*. **14**, 146–151 (2017).
- 1208 123. A. Brito-Mayor, A dos y a cuatro patas: el Occidente de Mesoamérica y la relación entre
- 1209 el perro, el humano y viceversa. Una aproximación a través de estudios de caso. *A*. **28**, 169
- 1210 (2019).
- 1211 124. T. Maricic, M. Whitten, S. Pääbo, Multiplexed DNA Sequence Capture of Mitochondrial
- 1212 Genomes Using PCR Products. *PLoS ONE*. **5**, e14004 (2010).
- 1213 125. M. L. Carpenter, J. D. Buenrostro, C. Valdiosera, H. Schroeder, M. E. Allentoft, M.
- 1214 Sikora, *et al.*, Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of
- 1215 Ancient DNA Sequencing Libraries. *The American Journal of Human Genetics*. **93**, 852–864
- 1216 (2013).
- 1217 126. A. García, R. Nores, J. M. B. Motti, M. Pauro, P. Luisi, C. M. Bravi, *et al.*, Ancient and
- 1218 modern mitogenomes from Central Argentina: new insights into population continuity, temporal
- 1219 depth and migration in South America. *Human Molecular Genetics*. **30**, 1200–1217 (2021).
- 1220 127. A. González-Martín, A. Gorostiza, L. Regalado-Liu, S. Arroyo-Peña, S. Tirado, I. Nuño-
- 1221 Arana, *et al.*, Demographic History of Indigenous Populations in Mesoamerica Based on mtDNA
- 1222 Sequence Data. *PLOS ONE*. **10**, e0131791 (2015).
- 1223 128. B. Z. González-Sobrino, *et al.*, Genetic Diversity and Differentiation in Urban and
- 1224 Indigenous Populations of Mexico: Patterns of Mitochondrial DNA and Y-Chromosome
- 1225 Lineages. *Biodemography and Social Biology*. **62**, 53–72 (2016).
- 1226 129. S. Romero-Hidalgo, *et al.*, Demographic history and biologically relevant genetic
- 1227 variation of Native Mexicans inferred from whole-genome sequencing. *Nature Communications*.
- 1228 **8**, 1005 (2017).
- 1229 130. J. Ramos-Madrigo, B. D. Smith, J. V. Moreno-Mayar, S. Gopalakrishnan, J. Ross-Ibarra,
- 1230 M. T. P. Gilbert, *et al.*, Genome Sequence of a 5,310-Year-Old Maize Cob Provides Insights into
- 1231 the Early Stages of Maize Domestication. *Current Biology*. **26**, 3195–3201 (2016).
- 1232 131. R. Acuna-Soto, Megadrought and Megadeath in 16th Century Mexico. *Emerging*
- 1233 *Infectious Diseases*. **8**, 360–362 (2002).
- 1234 132. G. Kubler, *The Population of Central Mexico in the Sixteenth Century*. (JSTOR, 1948).
- 1235 133. P. Skoglund, B. H. Northoff, M. V. Shunkov, A. P. Derevianko, S. Pääbo, J. Krause, *et*
- 1236 *al.*, Separating endogenous ancient DNA from modern day contamination in a Siberian

- 1237 Neandertal. *Proceedings of the National Academy of Sciences*. **111**, 2229–2234 (2014).
- 1238 134. R. Fregel, F. L. Méndez, Y. Bokbot, D. Martín-Socas, M. D. Camalich-Massieu, J.
- 1239 Santana, *et al.*, Ancient genomes from North Africa evidence prehistoric migrations to the
- 1240 Maghreb from both the Levant and Europe. *Proc Natl Acad Sci USA*. **115**, 6774–6779 (2018).
- 1241 135. S. J. Macfarlan, C. N. Henrickson, Inferring Relationships Between Indigenous Baja
- 1242 California Sur and Seri/Comcáac Populations Through Cultural Traits. *Journal of California and*
- 1243 *Great Basin Anthropology*. **30**, 18 (2021).
- 1244 136. J. V. Moreno-Mayar, *et al.*, Early human dispersals within the Americas. *Science*. **362**,
- 1245 eaav2621 (2018).
- 1246 137. E. M. P. Campos, INTERPRETACIÓN PRELIMINAR RESPECTO A LA
- 1247 TEMPORALIDAD DE TOLUQUILLA, QUERÉTARO, 24.
- 1248 138. H. Rangel-Villalobos, V. M. Martínez-Sevilla, G. Martínez-Cortés, J. A. Aguilar-
- 1249 Velázquez, M. Sosa-Macías, R. Rubi-Castellanos, *et al.*, Importance of the geographic barriers to
- 1250 promote gene drift and avoid pre- and post-Columbian gene flow in Mexican native groups:
- 1251 Evidence from forensic STR Loci: GEOGRAPHIC BARRIERS, DIFFERENTIATION AND
- 1252 ADMIXTURE IN AMERINDIANS. *Am. J. Phys. Anthropol.* **160**, 298–316 (2016).
- 1253
- 1254

1255  
1256  
1257

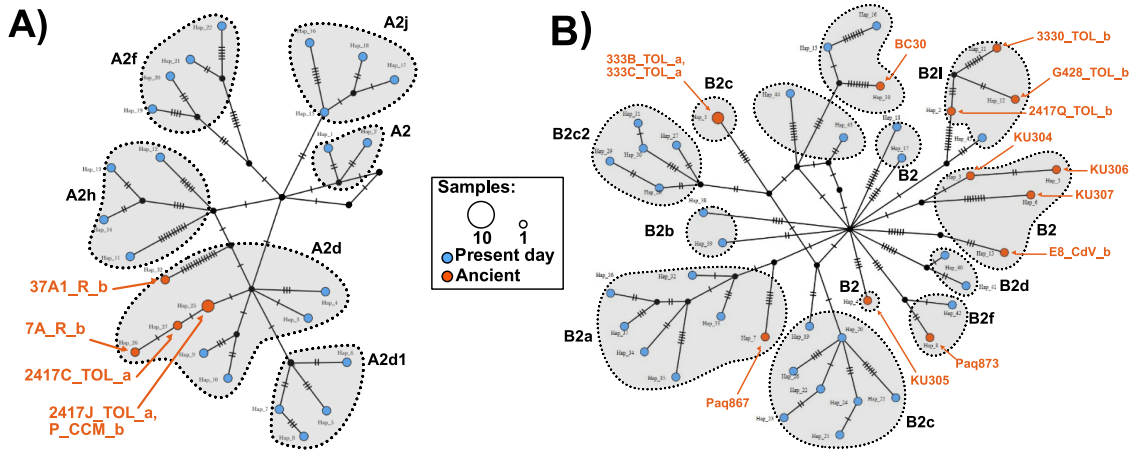
## Figures and Tables



1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270

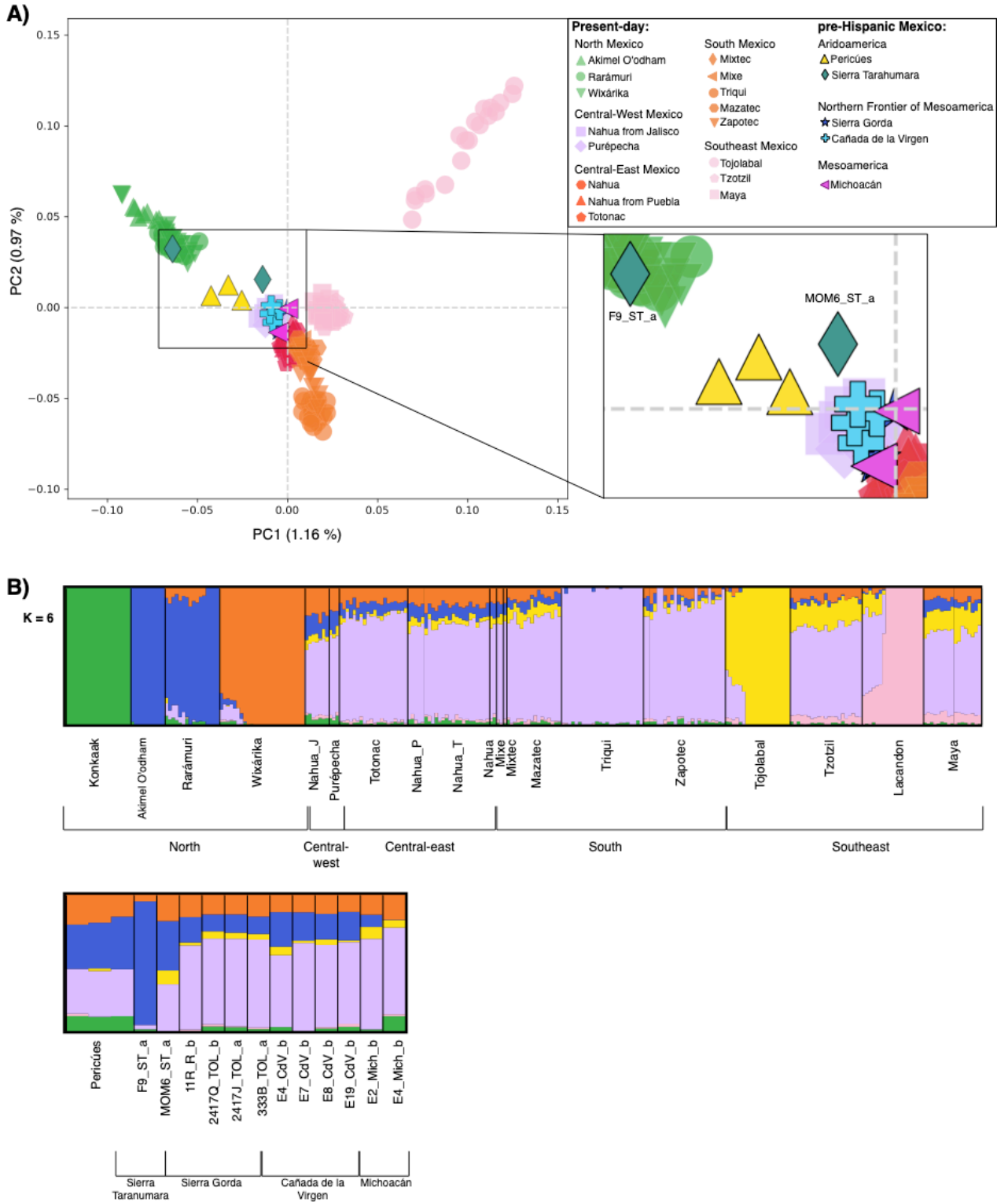
**Fig.1. Map of archeological sites and modern populations.** Map of Mexico indicating the location of the pre-Hispanic sites analyzed here (triangles) and in previous studies (squares) (15, 16) and the approximate locations of present-day Indigenous populations sampled in previous studies (circles) (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).

1271  
1272



1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282

**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.

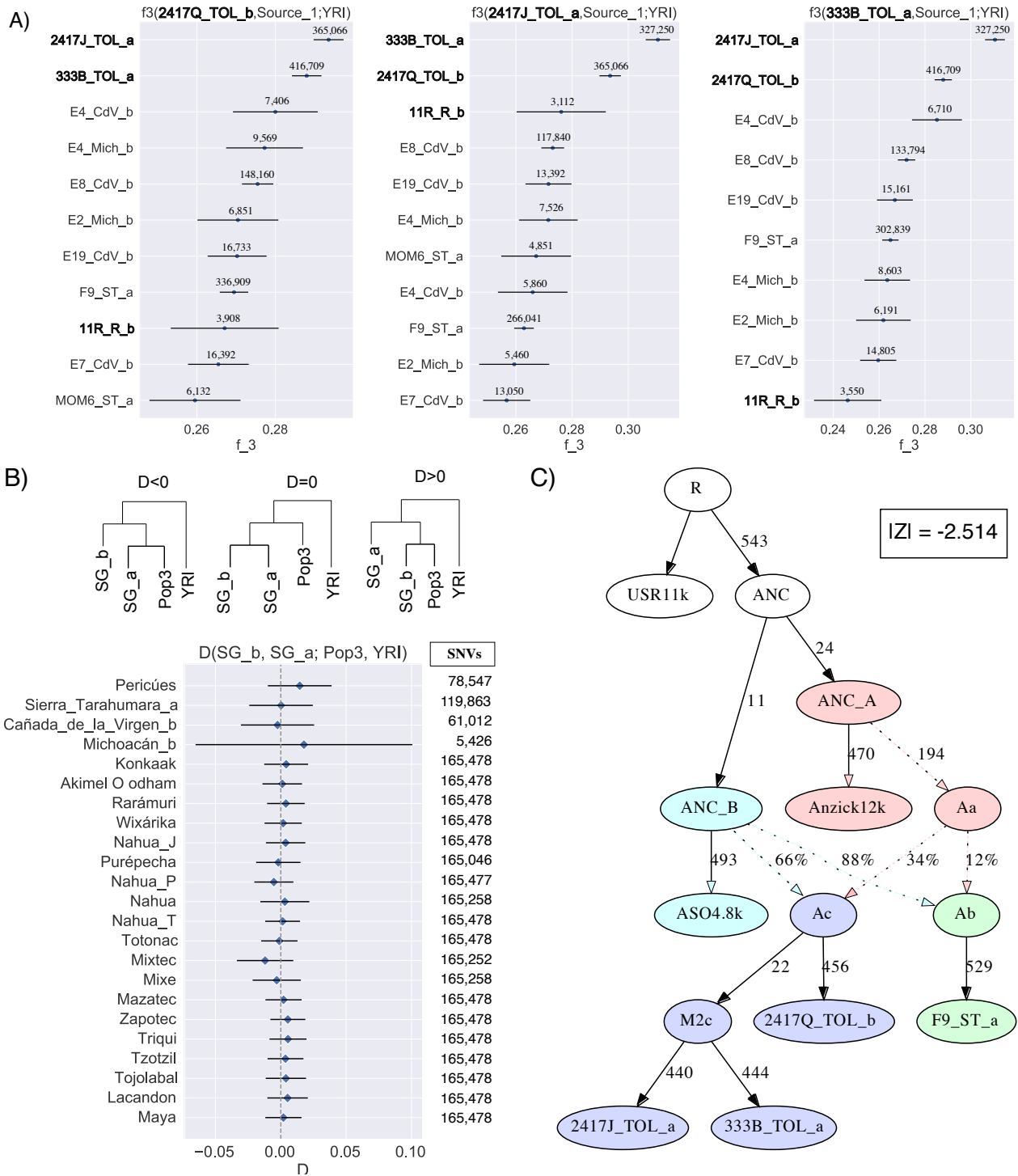


1283  
1284  
1285  
1286  
1287  
1288

**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

1289  
1290  
1291  
1292

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

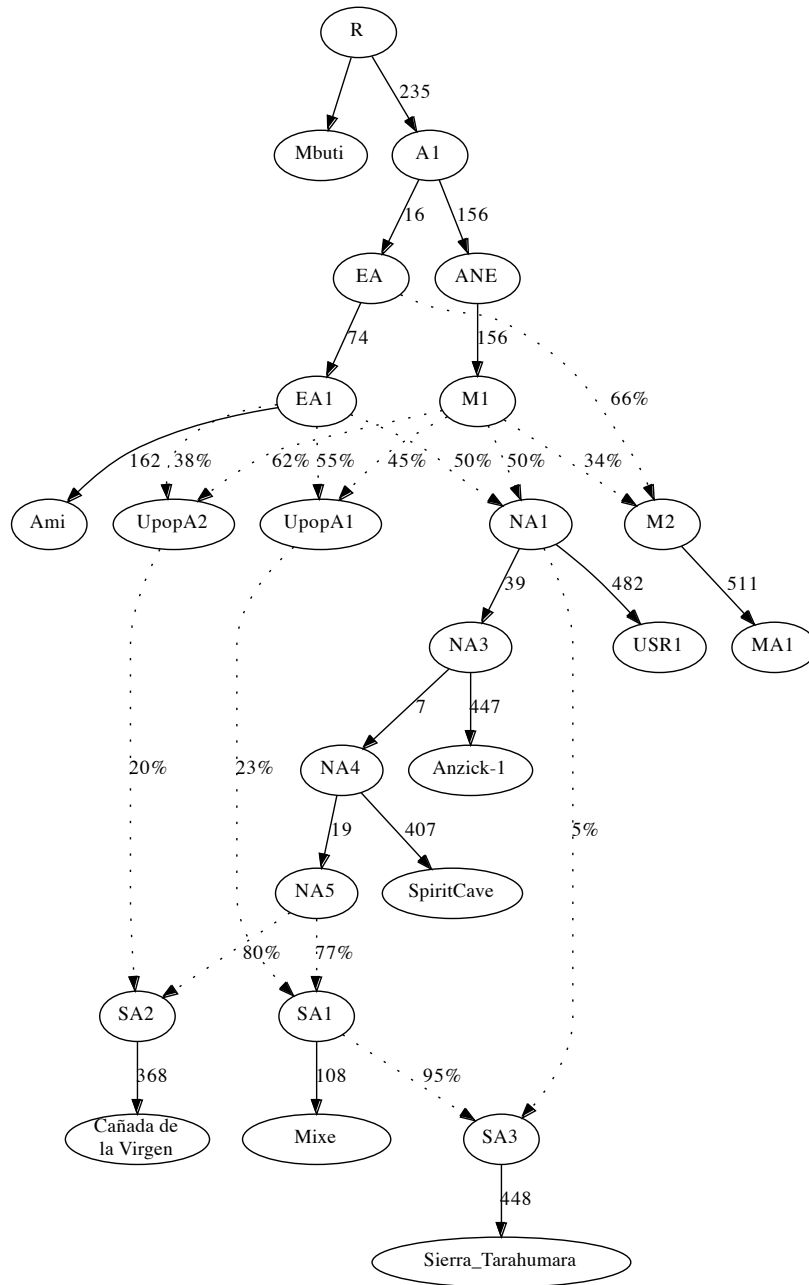


**Fig.4. Population continuity in Sierra Gorda.** (A) Shared genetic drift and genetic affinities of pre-Hispanic individuals from Toluquilla, 2417Q\_TOL\_b, 2417J\_TOL\_a, and 333B\_TOL\_a compared with pre-Hispanic individuals from Mexico. Higher values of  $f_3$  indicate higher shared genetic drift. 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

1294  
1295  
1296  
1297  
1298  
1299  
1300

1301 11R\_R\_b, and SG\_a 2417J\_TOL\_b and 333B\_TOL\_a. Pop3 is any of the other pre-Hispanic or  
1302 present-day populations from Mexico and is shown in the Y axis. YRI as an outgroup corresponds  
1303 to Yoruba from Africa from the 1000 Genomes Project (38, 101). Expected tree topologies  
1304 according to different D values are drawn on the top of the plot. Blue dots indicate no significant  
1305 deviations from D=0 ( $|Z|>3$ ). The number of SNV's used in each test is shown in the right. (C)  
1306 qpGraph model of individuals from Sierra Gorda from before and after the drought are shown in  
1307 the same clade, without contribution of F9\_ST\_a to the post-drought individuals.  
1308  
1309





1311  
 1312  
 1313  
 1314  
 1315  
 1316  
 1317  
 1318  
 1319

**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.

1  
2  
3  
4 **Supplementary Materials for**

5  
6 **Demographic history and genetic structure in Pre-Hispanic Central Mexico**

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

15  
16 \*Corresponding author. Email: mavila@liigh.unam.mx  
17  
18  
19  
20  
21

22 **This PDF file includes:**

23 Supplementary Text  
24 Figs. S1 to S33  
25 References (116-138)  
26

27 **Other Supplementary Materials for this manuscript include the following:**

28 Supplementary\_Table\_S1\_Sample\_summary\_information.xlsx  
29 Supplementary\_Table\_S2\_Sample\_information\_per\_sequencing.xlsx  
30 Supplementary\_Table\_S3\_Reference\_public\_ancient\_nuclear\_DNA.xlsx  
31 Supplementary\_Table\_S4\_mtDNA\_Mexico.xlsx  
32 Supplementary\_Table\_S5\_mtDNA\_new\_variants.xlsx  
33 Supplementary\_Table\_S6\_mtDNA\_for\_Networks\_and\_EBSP.xlsx  
34 Supplementary\_Table\_S7ab\_EBSP\_statistics.xlsx  
35 Supplementary\_Table\_S8\_Reference\_panel\_present-day.xlsx  
36 Supplementary\_Table\_S9\_F3outgroup\_prehispanic-present-day.xlsx  
37 Supplementary\_Table\_S10\_Dstats\_prehispanic-present-day.xlsx  
38 Supplementary\_Table\_S11\_individuals\_sites\_for\_CND\_analysis.xlsx  
39 Supplementary\_Table\_S12\_CND\_pairwise\_values.xlsx  
40 Supplementary\_Table\_S13\_hapROH.xlsx  
41 Supplementary\_Table\_S14\_F3outgroup\_prehispanic-prehispanic.xlsx  
42 Supplementary\_Table\_S15\_Dstats\_SG\_and\_prehispanic\_each\_individual.xlsx  
43 Supplementary\_Table\_S16\_Dstats\_SG\_and\_prehispanic\_per\_site\_Mexico.xlsx

---

44 Supplementary\_Table\_S17\_Dstats\_SG\_and\_prehispanic\_per\_site\_Mexico\_Transversion  
45 s.xlsx  
46 Supplementary\_Table\_S18\_qpWave.xlsx  
47

## 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)

93 Sierra Gorda, Querétaro (Mummy-P)

94  
95  
96 Cadereyta cave, Sierra Gorda, Querétaro (P\_CCM\_b)

97  
98 In 2002, the discovery of the mummified remains of individual P\_CCM\_b was reported in the  
99 vicinity of the village of Altamira, in the municipality of Cadereyta, Querétaro. The remains were  
100 removed and taken to the public prosecutor's offices in the municipal capital. Once the remains  
101 were handed over to INAH and archaeologist Elizabeth Mejia Perez Campos, the site was visited  
102 and cleaned. Fragments of textiles, organic matter such as bird feathers and wood formed a  
103 bracelet, and human hair braids were found. The piece was taken to the Templo Mayor Museum  
104 in Mexico City, where computerized axial tomography, necropsy with an endoscope to see internal  
105 organs and samples were taken. Likewise, mtDNA studies were performed by PCR amplification  
106 of the HVR at the Centro de Investigación y Estudios Avanzados of the Instituto Politécnico  
107 Nacional (CINVESTAV - IPN) in Mexico City and the textile and skin were dated by C13 and  
108 C14 at Beta Analytical in Florida, USA. The results showed a twenty-months-old infant who lived  
109 in 340 BC, shrouded with a cotton blanket with human hair interwoven and with an offering of  
110 maguey quills, stork feathers, pine, oak, and sotol leaves (118).

111  
112  
113 Toluquilla, Sierra Gorda, Querétaro

114  
115 Toluquilla archaeological site is located in the Sierra Gorda, in the State of Querétaro. The  
116 occupation of Toluquilla spans ca. 2,000 years, from 400 BCE to 1,550 CE.

117  
118 Eight individuals, 333A\_TOL\_a, 333B\_TOL\_a, 333C\_TOL\_a, 333O\_TOL\_b, 333Q\_TOL\_b,  
119 2417C\_TOL\_a, 2417Q\_TOL\_b and, 2417J\_TOL\_a, from the archaeological site of Toluquilla  
120 were sampled for this study. We analyzed samples from individuals from two ritual burials at the  
121 site. The burials were analyzed by Dr. Elizabeth Mejia Perez Campos as part of her doctoral thesis  
122 (6) and were named based on the relevance of the offerings. These burials correspond to the  
123 'Manatee context', and the 'Mirror context'.

124  
125 Manatee Context (Individuals 333x)

126  
127 The manatee context has a great quantity, quality, and variety of offerings. It includes shells and  
128 snails from both coasts (Gulf of Mexico and the Pacific Ocean), obsidian from the Sierra de las  
129 Navajas (Hidalgo, Mexico), local animals such as river fish lynx, armadillo plates, and foreign  
130 aquatic species from the deep sea.

131  
132 Individuals 333A\_TOL\_a, 333B\_TOL\_a and 333C\_TOL\_a 333O\_TOL\_b, 333Q\_TOL\_b from  
133 this context were sampled for this study. This context is located in building 33 for which dates are  
134 available, either by C14 dating, obsidian hydration dating, or the analysis of ceramic materials.

135  
136 In this context, the more recent burial is dated between 1,300 and 1,400 CE. In this burial  
137 individuals 333A\_TOL\_a, 333B\_TOL\_a, and 333C\_TOL\_a were found. Charcoal was deposited  
138 among the sediment that covered the bodies and has been C14 dated to 1,350 ± 50 CE.

139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184

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  $700 \pm 50$  CE. Individual 333Q\_TOL\_b was found in an older burial C14 dated to  $520 \pm 40$  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

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

188  
189 Five individuals were found in the third level of the northeast of the assemblage. The most relevant  
190 is individual 2417J\_TOL\_a, who presents pathologies such as spondylitis, evident in the sacrum,  
191 iliac and lumbar vertebrae. In addition to poor dental health, he also performed exercises that  
192 developed his muscles and left a mark on his arms and legs.

193  
194 In the fourth and last level, the remains of four more individuals were located, all adults, two  
195 secondary, and two remnant primary portions. Individual 2417Q\_TOL\_b corresponds to a remnant  
196 primary portion that was placed in the ventral decubitus position (face down) of a male between  
197 forty-five and fifty-five years old. Under individual 2417Q\_TOL\_b, there were two isolated bones,  
198 a coxal, and a vertebrae. Also under this individual, a pyrite mosaic mounted on a ceramic base  
199 was found, thought to serve as a mirror. The mirror is a significant offering; it is related to the  
200 priests and the god Tezcatlipoca, the deity of the underworld and the night, characterized because  
201 it carries a mirror, which is also used in the arts of divination, to obtain omens and to provoke evils  
202 (*119*).

203  
204 Individual 6428A\_TOL\_b

205  
206 The body of individual 6428A\_TOL\_b was found in building 64, being part of burial 28. It  
207 corresponds to an adult male in a flexed dorsal decubitus position with a significant offering  
208 including a flint bifacial, two jade beads, and ten amulets carved in bone (*119*).

209  
210 Ranas, Sierra Gorda, Querétaro

211  
212 Ranas archaeological site is located in the Sierra Gorda, in the State of Querétaro, and the  
213 occupation of the site is recorded from 400 BCE to 1,300 CE.

214  
215 Individual 11R\_R\_b

216  
217 Individual 11R\_R\_b was found in a burial that contained three individuals, one primary  
218 (11R\_R\_b) and two secondary arranged on the sides of the primary. 11R\_R\_b corresponds to an  
219 adult male over 35 years old. He was found deposited in dorsal decubitus, with the trunk slightly  
220 flexed upwards. The arms were placed in front of the body and the wrists were touching each other.  
221 His hands were on the pelvic region and in the right hand he was holding a ceramic pipe. Two  
222 bone circles forming a truncated cone were found in the area of the pubic symphysis. He presented  
223 osteological characteristics of repetitive work and high stress to the body. The enteropathies in the  
224 long bones of the legs, arm, vertebrae and calcaneus may be directly related to mining. Thus,  
225 individual 11R\_R\_b is inferred to have worked in mining for several years. The measurement of  
226 high values of heavy metals in a combination of Pb, Hg, As and Sb indicate chronic exposure to a  
227 wide variety of heavy metals present in geological contexts, which were systematically deposited  
228 in the bones of the individual (*57*).

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

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

233  
234 Individuals 7A\_R\_b and 3A7I\_R\_b were found in structure 3A, in burial 7 at the archaeological  
235 site of Ranas. Burial 7 is a complex burial, where initially five areas of intrusion are defined, and  
236 inside each of them, there is evidence of skeletal remains of one or more individuals.

237  
238 Individual 7A\_R\_b

239  
240 This individual is the most important since it is located just below the altar. It is a direct primary  
241 burial with a very flexed seated disposition. The individual was deposited as a mortuary bundle  
242 strongly tied (at the height of the shoulders and waist), so when the body decomposed in its  
243 taphonomic process, the head fell between the legs. The care taken in the deposition and the fact  
244 that limestone slabs covered the individual from its lowest level helped to keep most of the skeleton  
245 in its primary position, where the only element of post-depositional alteration was a rodent burrow  
246 that altered the metatarsal bones and phalanges of the left foot and altered the distal phalanges of  
247 the right foot.

248  
249 Individual 3A7I\_R\_b

250  
251 The case of individual 3A7I\_R\_b is another burial of an important dignitary since it is the only  
252 one deposited in an extended form with the head to the south, its face looking to the east, with the  
253 right arm flexed on its chest. The hand was deposited on its left shoulder, while its left arm rested  
254 on the belly, and its hand is extended between the right ribs and the right humerus, with the torso  
255 slightly leaning forward. The pelvic region was horizontal on the ground while the legs were  
256 separated with the right knee to the east, the left knee to the west, and the feet under the coxae.  
257 Therefore, this is a primary burial, in dorsal decubitus slightly raised from the back and head, with  
258 legs flexed on themselves. As it is a primary individual, the different body portions could be  
259 recognized and recovered, although in many cases, they were found in regular and poor  
260 preservation.

261  
262 Cañada de la Virgen, Guanajuato

263  
264 Individual E2\_CdV\_b

265  
266 Individual E2\_CdV\_b is a direct secondary burial, located in layer III of Table V-4 (although it  
267 partially intrudes up to layer IV). The burial presents only proximal fragments of incomplete  
268 femurs corresponding to an adult individual of undetermined age. Although it is possible that its  
269 partial intrusion into floor 1 and layer IV is accidental, the only difference with respect to the rest  
270 of the skeletons from Room 3 is the fact that it does not have contact with floor 2, but is found  
271 right at the level of floor 1. It seems more likely that this burial, due to its position and secondary  
272 character, functioned as a ritual seal for floor 2, in a termination ritual aimed at "killing the  
273 funerary space between the two floors (120).

274  
275 Individual E4\_CdV\_b

276



277 Located in layer IV of Table X-4. It is an indirect primary burial deposited in a cist in extended  
278 dorsal decubitus facing east. The estimated age is 5 years +/- 24 months and indeterminate sex. It  
279 presents 60% of the skeleton with absence of bones of the feet, patella, sacrum and coccyx (120).  
280

281 Individual E6\_CdV\_b

282  
283 Located in layer IV of Table V-3. Indirect secondary burial in cist, corresponding to a male  
284 individual of 40-45 years of age. It presents only fragmented iliacs, complete right radius, proximal  
285 part of both femurs at 50% and part of the proximal portion of the right tibia. Position and  
286 orientation not discernible (120).  
287

288 Individual E7\_CdV\_b

289  
290 Located in layer IV of Table X-3. Indirect primary burial deposited in cist in dorsal decubitus  
291 flexed oriented to the east. Corresponds to a female of 25 to 30 years of age. Feet and hands are  
292 missing (120).  
293

294 Individual E8\_CdV\_b

295  
296 Located in layer IV of table U-3. Direct primary burial in flexed dorsal decubitus oriented to the  
297 east. It corresponds to a male individual between 35 and 40 years old. He presents 90% of the  
298 skeleton with absence of feet (120).  
299

300 Individual E9\_CdV\_b

301  
302 Located in layer IV of Table T-4. It is an indirect primary burial deposited in flexed dorsal  
303 decubitus with orientation to the east. It corresponds to a male subject of 20 to 25 years of age. It  
304 presents 80% of the skeleton with absence of feet (120).  
305

306 Individual E10\_CdV\_b

307  
308 Located in layer IV of box V-3. This is a direct secondary burial with no discernible anatomical  
309 position. It consists only of a left femur diaphysis and a mandible (120).  
310

311 Individual E11\_CdV\_b

312  
313 Located in layer IV of squares U/V-3 and 4. It is a direct secondary burial formed by three  
314 fragments of long bones, fragments of ribs, a clavicle, some phalanges, fragments of pelvis and  
315 skull, as well as some molars (120).  
316

317 Individual E14\_CdV\_b

318  
319 Located in complex C, which is a circular structure, with an association to the wind and related to  
320 the North Star and Ursa Minor, which is linked to Tezcatlipoca. This burial was found at the foot  
321 of the access ramp to the second structure (120).  
322

323 Individual E15\_CdV\_b

324  
325 This individual was called the "girl of the rain". This subadult individual, through osteological  
326 method was estimated to be approximately 7 years old and possible a female, but we take this with  
327 caution since the individual is a subadult. She was found in complex B, and linked to the pluvial  
328 drainage, which is large and well-constructed with superimposed slabs. She was 1.10 m tall and is  
329 thought to be part of the ancestral veneration practiced in the place, since she was accompanied by  
330 a coyote and placed in such a way that it was necessary to decompose her body in order to  
331 accommodate her seated and compressed. Possible postmortem treatment in the temazcal that is at  
332 the top of the pyramidal base. The "girl of the rain", carried in her neck a small necklace of beads  
333 that includes a bead in the form of butterflies in the central part of the necklace. The offering is  
334 completed with a ceramic spindle and a fragment of marine shell. Next to the girl and as a  
335 companion and vestige of the practices of ancestral veneration, a small coyote was placed -a  
336 complete specimen of *Canis latrans*, an animal of totemic importance in the region. Forensic  
337 studies performed on the girl of the rain provided her height, and some pathologies associated with  
338 infectious processes such as periostitis or infection of the periosteum, as well as cranial lesions  
339 probably diagnostic of tuberculosis (120).

340  
341 Individual E18\_CdV\_b

342  
343 "The decapitated one". Part of his burial offering consisted of an elongated handle smoker. He was  
344 found associated with the enclosed patio located on the elevated platform on one side of the  
345 pyramidal base of Complex B. The individual is male, aged between 20 and 30 years old and with  
346 an unusual height of 1.73 centimeters - the left femur itself measured 45 centimeters. Forensic  
347 analysis indicates evidence of disease, specifically ankylosing arthritis, as well as apparent  
348 postmortem decapitation (120).

349  
350 Individual E19\_CdV\_b

351  
352 According to the skeleton, the burial of E19\_CdV\_b corresponds to a female individual, 20-25  
353 years old. It was identified as an indirect primary burial in a dorsal decubitus position (slightly to  
354 the right) flexed. This burial presented a series of alterations regarding the original position, and  
355 the hands were absent. Because there are no marks of violence or dismemberment, archaeologists  
356 suggest that the most likely hypothesis is that these elements were removed during some looting  
357 and were subsequently lost by agricultural work or by the action of rodents. The remains of this  
358 burial were found under a layer of killed pottery and associated remains of a canid (*Canis*  
359 *familiaris*) (121).

360  
361 Michoacán

362  
363 Eight individuals were retrieved from the state of Michoacán. Individuals E1A\_Mich\_b,  
364 E2\_Mich\_b and E4\_Mich\_b from the site La Mina; individual M1\_Mich from the site Tanhuato  
365 and Individuals M2\_Mich\_b, M3\_Mich\_b, M4\_Mich\_b and M5\_Mich\_b from the site Zaragoza.

366  
367 La Mina, Michoacán

369 At the end of December 2014, the INAH Michoacán Center was notified of the discovery of a  
370 series of skeletons located in the construction of a perimeter wall in the Telesencundaria #133  
371 school, in the town of La Mina, municipality of Álvaro Obregón, Michoacán.  
372

373 Upon arriving at Telesecundaria School #133, the archaeologists detected that the funerary context  
374 had been seriously altered due to the construction of a trench for the foundation of the perimeter  
375 wall. From the initial reports obtained from the masons, it was possible to determine that the bodies  
376 were concentrated in the same sector. During the first visit, towards the end of December 2014, a  
377 collection of those osteological elements extracted due to the construction work was carried out.  
378

379 So far, the extension of the site of La Mina is unknown due to the lack of monumental architecture  
380 and the growth of the current settlement, which occupies much of the area with vestiges.  
381 Excavations have shown that La Mina appears to be on what was once the shores of a village  
382 founded on the southwestern shore of Lake Cuitzeo.  
383

384 The absence of architectural elements on the surface, in the intervened area, and the surroundings,  
385 suggests an architecture based on perishable materials. The bioarchaeological remains were  
386 located in the northern sector of an embankment covering an area of approximately 1200 m<sup>2</sup>. The  
387 ceramic materials on the surface indicate that the settlement was occupied between 200 B.C. and  
388 900 A.D. Although this data indicates an extensive temporal extension, the stratigraphic unit  
389 indicates a single moment of occupation. The excavations carried out at this site were only aimed  
390 at recovering the bioarchaeological materials affected by the construction work, so the  
391 archaeologists do not rule out the existence of more burials at the site.  
392

#### 393 E1A\_Mich\_b

394  
395 Individual E1A\_Mich\_b was in a poor state of preservation. Only part of the lower limbs, two-  
396 thirds of the mandible, and the upper dental arcade were located. This individual corresponded to  
397 a male young adult placed in a right lateral decubitus position and was oriented in a west-east  
398 direction.  
399

400 The grave goods consisted of a vessel of the Paso Ancho/graffito type placed at the level of the  
401 lumbar vertebrae and showed clear traces of use. Two reused sherds deposited in front of the face  
402 completed the offering. The intentional placement of this ceramic fragment leads us to suppose  
403 that it is an object used in the development of the funerary ritual and later placed as part of the  
404 trousseau. These two ceramic fragments were dated together with another sample from another  
405 burial using the archaeomagnetic technique. The results indicated that the burial corresponded to  
406 647-825 CE (122) during the so-called Epiclassic period. A third element that made up the offering  
407 consisted of a dog in its juvenile development phase. Despite its poor state of preservation, it was  
408 possible to identify that the animal was not more than eight months old, considering the process  
409 of ossification of the epiphyses (123).  
410

#### 411 E2\_Mich\_b

412  
413 Individual E2\_Mich\_b corresponds to an early infant who was approximately six months old. This  
414 infant did not present any artifact or other element that could be identified as part of an offering.

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

417  
418 E4\_Mich\_b

419  
420 The individual E4\_Mich\_b corresponds to a skull collected at the site without context because it  
421 was removed by personnel building the school's perimeter fence. This skull was placed next to the  
422 burial of individual E1A\_Mich\_b.

423  
424 Tanhuato, Michoacán

425  
426 M1\_Mich corresponds to the burial excavated in a rescue excavation in the town of Tanhuato,  
427 Michoacán, during the construction of a canal in a street inside the town. Tanhuato is located in  
428 the municipality of the same name in the state of Michoacán. The burial was found at 1,543 meters  
429 above sea level without a greater context, since it is located in an urbanized area. The surrounding  
430 area is almost flat and there is a large amount of cultivated land. The burial was mixed, and the  
431 inhumations were made in an enclosure dug in the soil matrix, whose bottom is one meter deep  
432 with respect to the level of the street; it covers an area of about 5 m<sup>2</sup>. Primary and secondary burials  
433 were found in the pit and two incomplete individuals, two ossuaries and a funerary pot could be  
434 identified. Unfortunately, both bones and materials had already been removed by construction  
435 workers, so the context was very altered. Absolute dates for this individual are not yet available.

436  
437 Zaragoza, Michoacán

438  
439 The Zaragoza site is located in the municipality of La Piedad, on a flat-topped, isolated elevation  
440 called Mesa Acuitzio. This site presents a large rocky front that the settlers of Zaragoza took  
441 advantage of as a support for their settlement. The monumental area of the site was built on a  
442 relatively flat part located more or less halfway up the mesa, around 1,753 meters above sea level,  
443 while the living and cultivation area is distributed downwards, both to the east and north, along  
444 the slopes through artificial terraces.

445  
446 Samples M2\_Mich\_b and M3\_Mich\_b come from a mixed burial, excavated in 2004 by Anyul  
447 Cuellar and Diana Bustos, called Burial 1. The space was used in successive events, that is, not all  
448 the bodies were deposited there simultaneously. It is located outside, on the east side, of an  
449 architectural unit that we have identified as a possible temascal (low heat sweat lodge), located at  
450 the southern end of one of the main group of structures. Among the buildings that make up this  
451 complex, a ball court stands out, almost 57 meters long (including the headers) and 14 meters  
452 wide.

453  
454 A total of 14 individuals were located in the funerary context, identified from the number of skulls,  
455 i.e., not necessarily complete bodies were found. The remains were in a very poor state of  
456 preservation, but at least 5 primary deposits were identified. M2\_Mich\_b comes from individual  
457 7, located at the western end of the pit, next to the wall that delimits the east side of the temascal.  
458 M3\_Mich\_b corresponds to a second molar of skull 13. Both skulls were found without anatomical  
459 connection, although they were part of the group of bodies deposited in the burial space.

460

461 M4\_Mich\_b and M5\_Mich\_b, also from the Zaragoza site, but from Burial 2, were also mixed and  
462 excavated in 2012 by Alfredo Salas. This burial is located to the south of structure 1, which is part  
463 of the same complex as the court, to the east of it. The pit, of 24 m<sup>2</sup>, presents a total of 8 or 9  
464 individuals, which are not specified yet. The samples correspond to a first molar of M4\_Mich\_b  
465 and a third molar of M5\_Mich\_b, respectively. Individual M4\_Mich\_b belongs to an individual  
466 placed in extended dorsal decubitus with a south-north axis, slightly deviated from 6 to 8 degrees  
467 to the east; it is important to note that this is the only individual with this orientation inside the pit.  
468 This individual had an associated offering composed of a cajete (ceramic bowl), a pot and a  
469 "Capiral" lid. Individual M5\_Mich\_b was the most complete skeleton, although in a poor state of  
470 preservation; M5\_Mich\_b was found in an extended dorsal decubitus position, with arms flexed  
471 and with an east-west orientation with a 287° deviation with respect to north.  
472

## 473 **Extended Results**

### 474 **Sample selection and sequencing strategy**

475  
476 We screened the DNA content of archeological samples from 39 pre-Hispanic individuals (Table  
477 S1). The central Mexico sites include three in the Sierra Gorda (SG) in Querétaro state, namely  
478 Toluquilla (TOL, n=9), Ranas (R, n=3) and a cave in Cadereyta de Montes (CCM, n=1); Cañada  
479 de la Virgen (CdV) in Guanajuato state (n=16); and the Zaragoza (n=4), Tanhuato (n=1) and La  
480 Mina (n=3) sites in Michoacán state (Mich). Furthermore, we produced additional genome-wide  
481 data for two mummies from the Sierra Tarahumara in northern Mexico (Table S1), which were  
482 sequenced at lower depths in a previous study (78) (Fig. 1).

483 For each sample, we evaluated the extent of endogenous ancient DNA (aDNA) via shotgun  
484 sequencing. Out of the 39 samples, only three had endogenous content above 10% (13 – 39%)  
485 (Table S2). We then generated additional sequence data for these three samples using shotgun  
486 sequencing, while for the remaining libraries, we performed whole-genome or mitochondrial DNA  
487 (mtDNA) capture (124, 125) (Table S2). For all of these, we verified the presence of the  
488 characteristic misincorporation patterns of aDNA using mapDamage (71) (Fig. S1-S4). We  
489 obtained informative amounts of genome-wide data for 12 individuals (0.01 – 4.7 x). Ten were  
490 from Mesoamerica: Sierra Gorda (n=4), Cañada de la Virgen (n=4), and Michoacán (n=2); and  
491 two from the Sierra Tarahumara in Aridoamerica (Table S1). We integrated these data with  
492 previously reported low-depth genomes (0.09 – 0.3 x) from three ancient Pericúes from  
493 Aridoamerica reported in (78, 79), resulting in a total of 15 pre-Hispanic individuals from Mexico  
494 for which we had genome-wide data. Our analysis also included ancient genomic data from  
495 previous studies for 21 individuals from across the Americas (Table S3)(46, 78, 79, 105–109).  
496 Additionally, we reconstructed the mitochondrial genomes for 27 individuals across Mexico (3.2  
497 – 1284.8x), spanning 1,600 years (320 BCE to 1,351 CE). Of these, 25 were from Mesoamerica:  
498 SG (n=13), CdV (n=7), and Michoacán (n=5); and two individuals were from Aridoamerica: Sierra  
499 Tarahumara (n=2) (78) (Table S1 and S2). These mitochondrial genomes were analyzed along  
500 with previously published mitochondrial genomes from Mexico and South America (26, 78, 80,  
501 81, 83–85, 87–91, 108, 126) (30, 78, 79) (Table S4 and S6).  
502

### 503 **Mitochondrial haplogroup structure and diversity**

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

506 previously observed in ancient and contemporary populations from Mexico (26–32, 80).  
507 Specifically for central Mexico, we found the presence of all four autochthonous haplogroups, A,  
508 B, C, and D, similar to previous observations on ancient and present-day populations inhabiting  
509 this region (20, 27, 28, 51, 127, 128). This highlights the greater genetic diversity and  
510 heterogeneity in this particular area (44, 129), and is in agreement with larger Pre-Hispanic  
511 agriculture-based civilizations and greater population mobility compared to Aridoamerica.  
512 Furthermore, we detected unreported variants in the mtDNA PhyloTree (33) for seven individuals  
513 assigned to the sub-haplogroups A2d (n=4), B2c (n=2), and B2l (n=1). The absence of these  
514 variants could reflect a poor sampling of Native American mitochondrial genomes in the  
515 PhyloTree database, or a loss of these haplotypes because of drift. Accordingly, we propose adding  
516 these new unreported variants to the PhyloTree database (Table S5).

517 In Sierra Tarahumara, we found two sub-lineages of the mtDNA haplogroup C (C 100%), as  
518 previously reported. In contrast, Central Mexico shows a higher diversity of mtDNA haplogroups.  
519 Cañada de la Virgen was the site with the highest number of sub-haplogroups, distributed in B  
520 42.3%, A 28.6%, C 14.3%, and D 14.3%. Michoacán presented four different sub-haplogroups (A  
521 40%, D 40%, and C 20%). Sierra Gorda also presented four different sub-haplogroups but  
522 corresponding to major lineages A 46.15%, B 46.15%, and D 7.7%, respectively.  
523

524 In Sierra Gorda, out of six A2d individuals four belong to the pre-drought period according to  
525 dating based either on C14 or the archaeological context, and two to the drought period. The  
526 haplotype median-joining network for the haplogroup A2 shows a clustering of five Sierra Gorda  
527 individuals carrying the sub-haplogroup A2d (the sixth A2d individual, 333Q\_TOLb, was not  
528 included in the network as it did not pass the quality filters for inclusion) despite a time transect  
529 of 1,480 years (Fig. 2A). One of the five individuals, 37A1\_R\_b, harbors several polymorphisms  
530 with respect to the other four. Unfortunately, this individual barely passed our filtering criteria  
531 (high presence of ‘N’s’ and excess C > T mutations). Because haplotype networks can be biased  
532 by sequence uncertainty, most of this individual's polymorphisms are likely from sequencing  
533 errors and/or ancient damage. This makes this individual difficult to place and compare with  
534 confidence. However, the other four individuals (7A\_R\_b, 2417C\_TOL\_a, 2417J\_TOL\_a and  
535 P\_CCM\_b) form a unique clade with each other showing genetic continuity. The clade that they  
536 form is striking given their high sequence similarity, location and time span. This observation  
537 supports a continuity of this sub-haplogroup, despite the severe droughts between 900 – 1,300 CE  
538 in this region of the Northern frontier of Mesoamerica. In contrast, the haplotype network for  
539 haplogroup B, shows three pre-drought individuals sharing sub-haplogroup B2l, while the two  
540 post-drought are found in the same node representing sub-haplogroup B2c. No individuals from  
541 Sierra Gorda carried C haplogroups and only one individual (11R\_R\_b) had a D sub-haplogroup  
542 (D1) (Fig. S5).  
543

544 We estimated the diversity of the mitochondrial genomes per pre-Hispanic site and period using  
545 the 27 mtDNA genomes, along with eight public mitochondrial genomes from Mexico (Table S4).  
546 We calculated nucleotide diversity  $\pi$  through the genetic distance method of Tajima and Nei, with  
547 the software Arlequin v. 3.5.1.3 (77). As expected, we find that the pre-Hispanic populations from  
548 Mesoamerica have higher  $\pi$  values (Sierra Gorda pre-drought: 0.003022, Sierra Gorda post-  
549 drought: 0.003219, Cañada de la Virgen: 0.005609, Michoacán: 0.004394) than those from  
550 Aridoamerica (Sierra Tarahumara: 0.001895, Cueva Candelaria: 0.000413, Paquimé: 0.002118).  
551



552

### **Past female effective population size inference**

553  
554  
555  
556

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).

557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574

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.

575

### **Sex determination and Y chromosome haplogroups**

576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595

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 ( $R_y$ ). According to the method,  $R_y > 0.075$  corresponds to XY, while  $R_y < 0.016$  corresponds to XX. For another four individuals, although the  $R_y$  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  $R_y$  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).

596 After applying the filters, a total of 20 individuals got a molecular sex assignment; nine agreed  
597 with the previous morphological assignment, one was not in agreement (individual  
598 2417J\_TOL\_a), and ten corresponded to individuals for which there was not a morphological  
599 assignment due to poor preservation of the skull or youngness of the individual (Table S1).  
600

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

#### 616 **Autosomal genetic structure**

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

623 We applied a novel approach named “missing DNA” PCA (mdPCA) on the 15 pre-Hispanic  
624 individuals from Mexico, 12 from this study (Table S1) and the 3 Pericúes previously published  
625 (Table S3). Ancient individuals clustered closely with present-day populations from the same  
626 geographical region. Individual F9\_ST\_a from Aridoamerica clusters closely with present-day  
627 Rarámuri populations from north Mexico. Notably, MOM6\_ST\_a, also from Aridoamerica,  
628 clustered closely with present-day Indigenous populations from central-west Mexico (Purépecha  
629 and Nahuas from Jalisco). Pericúes from Baja California (78, 79) cluster closely between  
630 Wixárika, and central west Mexico Nahuas from Jalisco and Purépechas. Regarding the ancient  
631 individuals from central Mexico (Sierra Gorda, Cañada de la Virgen, and Michoacán), these cluster  
632 closely with present-day Indigenous populations from this same region (Nahua from Jalisco,  
633 Purépecha, Totonac, Nahua from Puebla, and Nahua) (Fig. 3A). Both pre-, and post-drought Sierra  
634 Gorda individuals cluster together with central Mexico present-day populations, contrary to what  
635 would be expected if a population replacement from Aridoamerica in the Sierra Gorda would have  
636 taken place (Fig. 3A). A similar structure was observed with standard PCA, except that here  
637 F9\_ST\_a clusters closer with Wixárika than with Rarámuri (Fig S9).

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



640 dates of the pre-Hispanic individuals. For this, we included only four pre-Hispanic individuals  
641 with >1x genome-wide coverage (F9\_ST\_a, 2417Q\_TOL\_b, 2417J\_TOL\_a, and 333B\_TOL\_a).  
642 This analysis confirmed the closer genetic affinity of individual F9\_ST\_b with populations from  
643 Aridoamerica, while Sierra Gorda individuals from Toluquilla were slightly shifted towards the  
644 central-east populations in contrast with their position in the mdPCA between central-west and  
645 central-east populations. Most notably, despite the different periods of the individuals from  
646 Toluquilla, they cluster together with all three dimension reduction approaches.

647 In agreement, ADMIXTURE analysis (K=6) reveals that pre-Hispanic individuals from Sierra  
648 Tarahumara and Pericúes have a bigger proportion of genetic ancestry shared with present-day  
649 northern populations Akimel O’odham and Rarámuri (blue component in Fig. 3B). Pre-Hispanic  
650 individuals from central Mexico also display a similar genetic composition to present-day central-  
651 west and central-east Mexico populations (Nahua from Jalisco, Purépecha, Totonac and, Nahua  
652 from Puebla). Interestingly, individuals from Sierra Gorda show a homogeneous genetic  
653 composition independently of the period in which they lived. The genetic composition of  
654 individuals from Cañada de la Virgen and Sierra Gorda is similar, but Cañada de la Virgen  
655 individuals have a slightly higher proportion of the blue component shared with northern  
656 populations. In addition, the two individuals from Michoacán have a different proportion of the  
657 components shared with northern populations, E2\_Mich\_b presents the orange and blue  
658 components while E4\_Mich\_b presents orange and green. However, due to the low-coverage of  
659 these individuals (0.03x and 0.04x, respectively), we take this observation with caution.

### 660 **Outgroup-f3 and D-statistics**

661 We computed outgroup-f3 statistics of the form  $f_3(\text{Test}, \text{Source1}; \text{YRI})$ , where the Test refers to  
662 the pre-Hispanic individual under analysis, and Source1 being a present-day Indigenous  
663 population. D-statistics were of the form  $D(\text{Pop1}, \text{Pop2}; \text{Test}, \text{YRI})$ , where Test refers to the pre-  
664 Hispanic individual under analysis, while Pop1 and Pop2 were all possible combinations of  
665 present-day Indigenous populations. For both analysis we considered only the combinations that  
666 have >1,000 overlapping Single Nucleotide Variants (SNVs), being 1,181 the lowest amount of  
667 SNVs plotted.

668 We found F9\_ST\_a to share higher genetic drift with present-day northern Mexico populations:  
669 Rarámuri, Akimel O’odham and Konkaak (0.291, 0.282, 0.277), than with any other present-day  
670 population (Fig. S11) (Table S9). Consistently, D-statistics showed F9\_ST\_a to be significantly  
671 more related to the present-day Rarámuri than to any other population (Pop1) when tested in the  
672 form  $D(\text{Pop1}, \text{Rarámuri}; \text{F9\_ST\_a}, \text{YRI})$  ( $D = -0.0763$  to  $-0.0210$ ,  $|Z| = 6.913$  to  $20.281$ ) (Fig. S12)  
673 (Table S10).

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

680

### 681 **Conditional Nucleotide Diversity (CND) and Runs of Homozygosity (ROH)**

682           CND, a relative measure of genetic diversity, was estimated by genetic pairwise comparison of the  
683           genomic positions that are variable in African individuals (41). For these comparisons we grouped  
684           the individuals per archeological site and calculated CND by pairs of individuals. This analysis  
685           included 16 pre-Hispanic individuals (Pericúes, Sierra Tarahumara, Sierra Gorda, Cañada de la  
686           Virgen and Patagonia) with at least 10,000 SNVs overlapping between each pair of individuals  
687           (Table S11 and S12). We found similar values of CND in pairs of pre-Hispanic individuals to those  
688           observed in present-day Indigenous populations from Mexico, except for Pericúes from Baja  
689           California, who had the lowest CND values (Fig. S19) (Table S12). This is in line with Pericúes  
690           being hunter gatherers, likely with small population sizes and living in isolation in Baja California  
691           (135).

692           We calculated ROH in six pre-Hispanic individuals with the highest coverage; namely two from  
693           Aridoamerica (the Pericú B03 and F9\_ST\_a from Sierra Tarahumara) and four from central  
694           Mexico (three from Toluquilla and one from Cañada de la Virgen) (Table S13). Of the three  
695           Toluquilla individuals included in this analysis, one was dated pre-drought (2417Q\_TOL\_b), and  
696           two post-drought (2417J\_TOL\_a and 333B\_TOL\_a). Pre-Hispanic individuals display fewer  
697           segments in ROH than some present-day Indigenous individuals, especially present-day  
698           individuals from northern Mexico. The segment size distribution of ROHs suggests that the pre-  
699           Hispanic individuals studied here belonged to populations with small effective population ( $N_e$ )  
700           sizes ( $2N_e= 1,600-6,400$ ) (Fig. S20), which is similar to  $N_e$  previously calculated for present-day  
701           northern (Rarámuri,  $N_e= 2,419$ ; Wixárika,  $N_e= 2,193$ ) and southern (Triqui,  $N_e= 2,407$ ; Maya,  
702            $N_e= 2,750$ ) populations (44). We contrasted the CND values and ROH estimations of Pericúes  
703           from Mexico with another isolated pre-Hispanic population from Patagonian hunter gatherers  
704           (1,275 – 895 BP) (109), expecting similarly low levels of genetic diversity. We found that two  
705           pairs of Pericúes had similar CND levels as Patagonia individuals ( $\sim 0.16-0.175$ ), and the third pair  
706           of Pericúes presented a lower CND value ( $\sim 0.14$ ). However, in ROH analysis, we found that the  
707           four individuals from Patagonia have a higher sum of segments in ROH than the B03 Pericú  
708           individual, suggesting that the Pericú population had a bigger population size than that of South  
709           American Patagonian individuals (Fig. S20). In summary, the CND and ROH information showed  
710           that pre-Hispanic and present-day individuals from northern Mexico (Pericúes and Akimel  
711           O’odham, respectively) present the lowest CND values and the highest sum of inferred ROH  $>4$   
712           cM compared to other populations of their respective period. Individuals from the Sierra  
713           Tarahumara, Toluquilla and Cañada de la Virgen show similar values of genetic diversity and sum  
714           of inferred ROH segments. Furthermore, in the case of Toluquilla, where we have three individuals  
715           from different dates, we found that the CND value increases with the date difference between pairs  
716           of individuals, probably reflective of the accumulation of new mutations or genetic drift through  
717           time during the 489-680 years of difference between 2417Q\_TOL\_b and the individuals of a later  
718           period (2417J\_TOL\_a and 333B\_TOL\_a).

719

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

721           We decided to evaluate the possible scenario of population continuity in the Sierra Tarahumara  
722           from northern Mexico, considering that the individual F9\_ST\_a was significantly more related to  
723           present-day Rarámuri than to any other indigenous population from Mexico in  $f_3$  and D statistics.  
724           We used the population continuity test described in (40), with the null hypothesis of population  
725           continuity tested by using read counts in the ancient individual and allele frequencies in the

726 reference population (Rarámuri). Using this method, we found no population continuity between  
727 F9\_ST\_a and present-day Rarámuri and rejected the null hypothesis with a p-value of  $10^{-499.3}$ .

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

735 Interestingly, Sierra Tarahumara mummy MOM6\_ST\_a has a higher shared genetic drift with  
736 F9\_ST\_a, as evidenced by outgroup-f3 tests, but mdPCA and ADMIXTURE analyses show that  
737 he has also genetic ancestry components found in central west present-day populations. Using D  
738 statistics, we found MOM6\_ST\_a to be equally related to all individuals from Aridoamerica and  
739 Mesoamerica, supporting the idea of gene flow between these two areas.  
740

### 741 **Ghost population contribution to pre-Hispanic Mexico**

742 We tested the presence of UpopA in the pre-Hispanic individuals, using an admixture graph model  
743 with the same topological framework but some modifications compared to (136). We used Mbuti  
744 as an outgroup, Ami as the East Asian, MA1 as the ancient north Eurasian, USR1 as the ancient  
745 Beringian, Anzick-1 (46) and Spirit Cave (136) as the SNA, and Athabascan as the NNA.  
746 Interestingly, the Mummy F9\_ST\_a from the north and the individuals from Cañada de la Virgen  
747 from central Mexico show genetic ancestry from a ‘ghost’ population, at 28% and 17%,  
748 respectively ( $|Z\text{-score}|$ : 2.142 and 1.655, respectively) (Fig. S27A-B), consistent with the “ghost”  
749 UpopA genetic ancestry reported in Mixe. Models including individuals from other archaeological  
750 sites had  $|Z\text{-score}|$  values over 3 and branches with inner zeros, thus we rejected them. We repeated  
751 the test with the Mixe population to verify the contribution of the ghost population in our tested  
752 admixture graph structures, and found Mixe to have a genetic ancestry coming from an unsampled  
753 UpopA population at 13% ( $|Z\text{-score}|$ : 2.34) (Fig. S27C), which is similar to previous estimates of  
754 11% (136). To confirm these results, we tested whether the ghost genetic ancestry observed in pre-  
755 Hispanic individuals was the same found in Mixe, by modeling the pre-Hispanic individuals along  
756 with Mixe, applying two different scenarios: 1) pre-Hispanic modeled in a same clade with Mixe,  
757 receiving UpopA genetic ancestry, and 2) pre-Hispanic and Mixe receiving UpopA genetic  
758 ancestry, with an extra contribution of NNA in the pre-Hispanic. Interestingly, Results show that  
759 Sierra Tarahumara and Mixe share the same ghost genetic ancestry UPopA (with F9\_ST\_a better  
760 modeled when having a higher contribution of NNA genetic ancestry) (Fig. S28), while in both  
761 models Cañada de la Virgen does not share the same UPopA contribution ( $|Z| > 3$ ) (Fig. S29).  
762 Then we tested a third model with Mixe and Cañada de la Virgen to see whether the genetic  
763 ancestry in the latter, comes from a secondary unsampled population (UPopA2) and not the  
764 previously observed in Mixe (UPopA1). Results show Cañada de la Virgen can be modeled  
765 receiving genetic ancestry from a different unsampled population “UPopA2” suggesting the  
766 existence of another ghost population not sampled yet (Fig S29). We then modeled the two pre-  
767 Hispanic, Mixe and the two unsampled populations within the same model to confirm and  
768 summarize the results (Fig. 5). We attempted different combinations for the placement of UpopA,  
769 as well as adding and removing USR1 to confirm whether Mixe and Cañada de la Virgen could be

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

772

## 773 **Extended Discussion**

774

### 775 **Population continuity in Sierra Gorda**

776

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

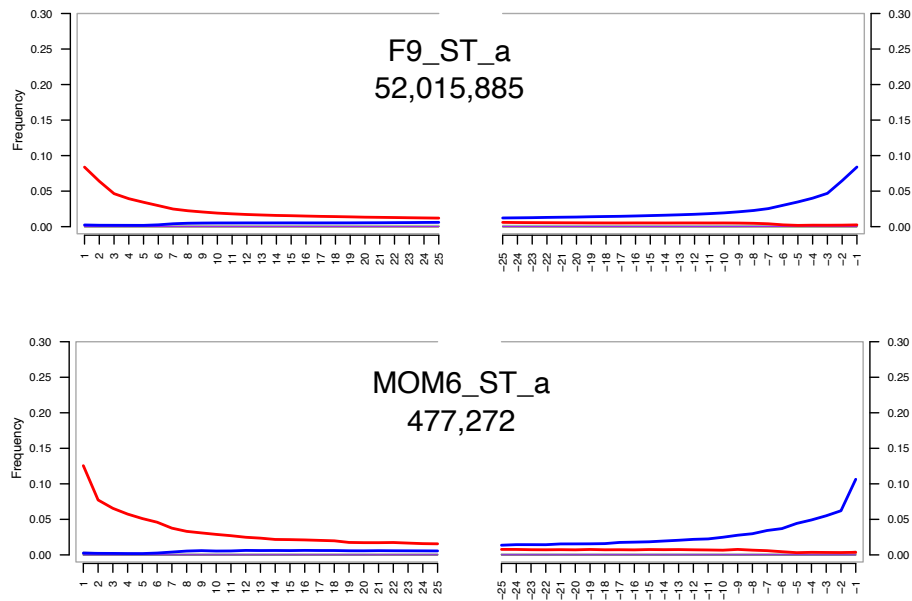
795

### 796 **Demography of central Mexico populations**

797

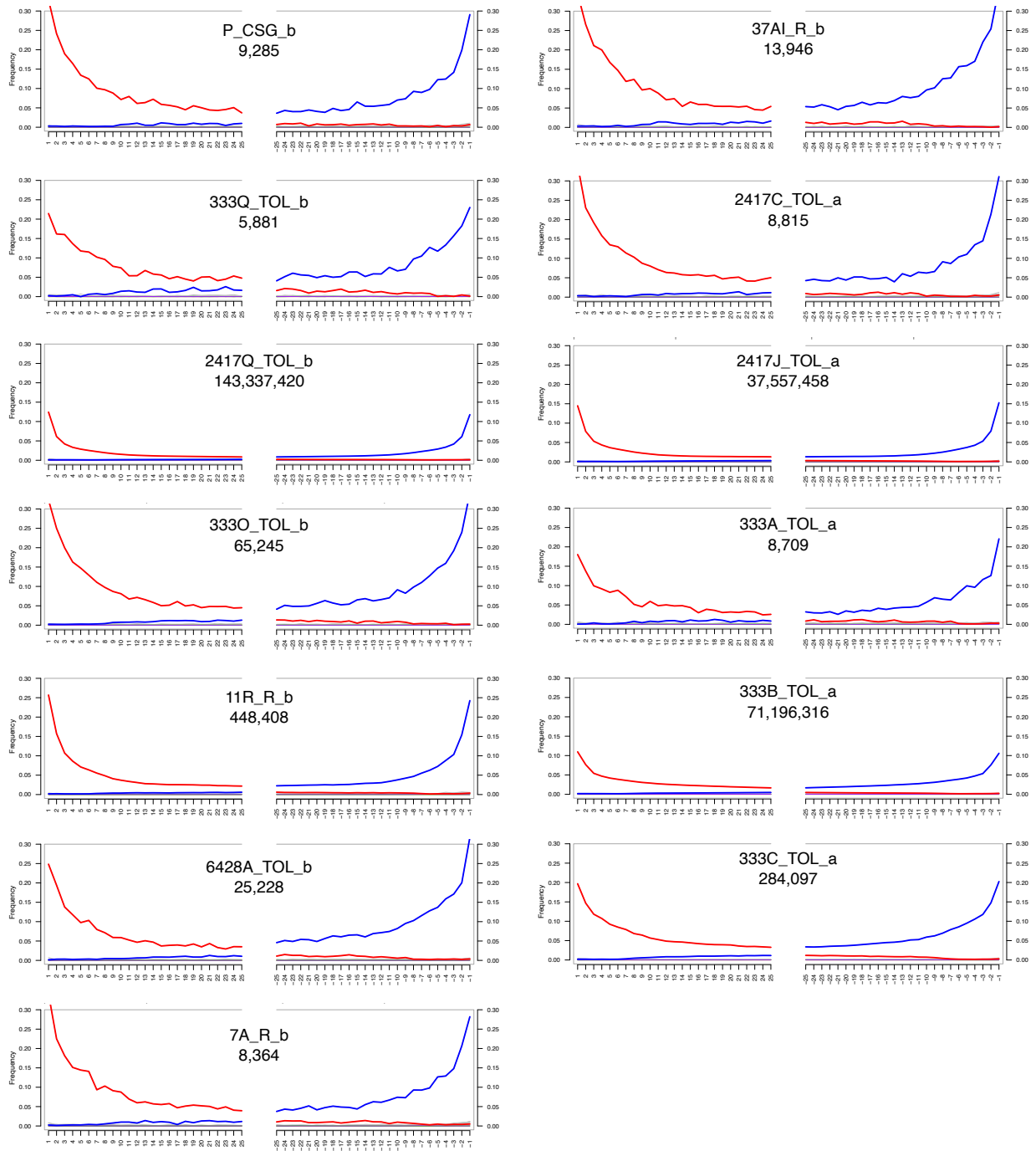
798 We were also interested in obtaining insights into the demography of central Mexico populations  
799 since previous attempts of demographic modeling have been hampered by the high genetic  
800 heterogeneity observed in these populations (44, 52). Our outgroup-f3 and D statistics results show  
801 that, although pre-Hispanic individuals from central Mexico share higher genetic drift with  
802 present-day populations from that region, none of these relationships are statistically significant.  
803 We thus leveraged the ancient genome data to propose an admixture model for central Mexico  
804 populations. Using qpGraph we found that the pre-Hispanic populations from central Mexico, all  
805 have different genetic ancestry sharing with the NNA and SNA branches. This was expected given  
806 previous studies on ancient genomes from Central and South America reporting multiple  
807 admixture events between these two branches after their split ca. 15,000 years ago (49, 79).  
808 Together, these observations point to a scenario in which populations from central Mexico have  
809 not completely diverged from one another, possibly due to extensive gene flow as expected based  
810 on the active commercial exchange between different Mesoamerican populations for centuries  
811 (63). This interaction was mainly through trade routes and alliances between different nations (63),  
812 as observed in present-day Indigenous populations who share identity-by-descent (IBD) segments  
813 (37, 49) and through the analysis of STR loci (138). There was also gene flow between  
814 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

815 flow occurred less frequently than within Mesoamerica. The individual MOM6\_ST\_a from  
816 Aridoamerica may have been the result of such admixture between Mesoamerican and  
817 Aridoamerican ancestors. This individual has a higher shared genetic drift with F9\_ST\_a in  
818 outgroup-f3 tests. However, ADMIXTURE analysis show that MOM6\_ST\_a has genetic ancestry  
819 components found in central west present-day populations. Also, using  $\bar{D}$  statistics, we found  
820 MOM6\_ST\_a to be equally related to all individuals from Aridoamerica and Mesoamerica,  
821 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.

823  
824  
825  
826  
827  
828  
829  
830  
831

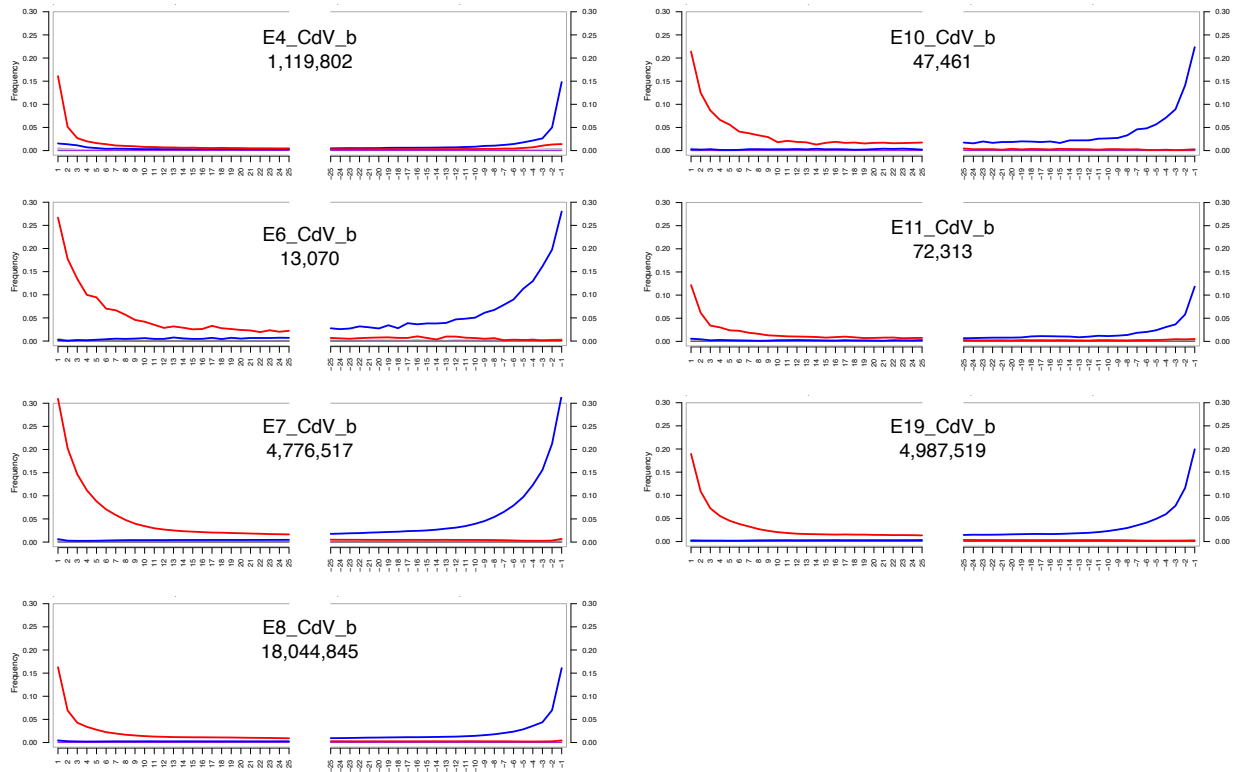


833  
834  
835  
836  
837  
838  
839

**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.



840  
841

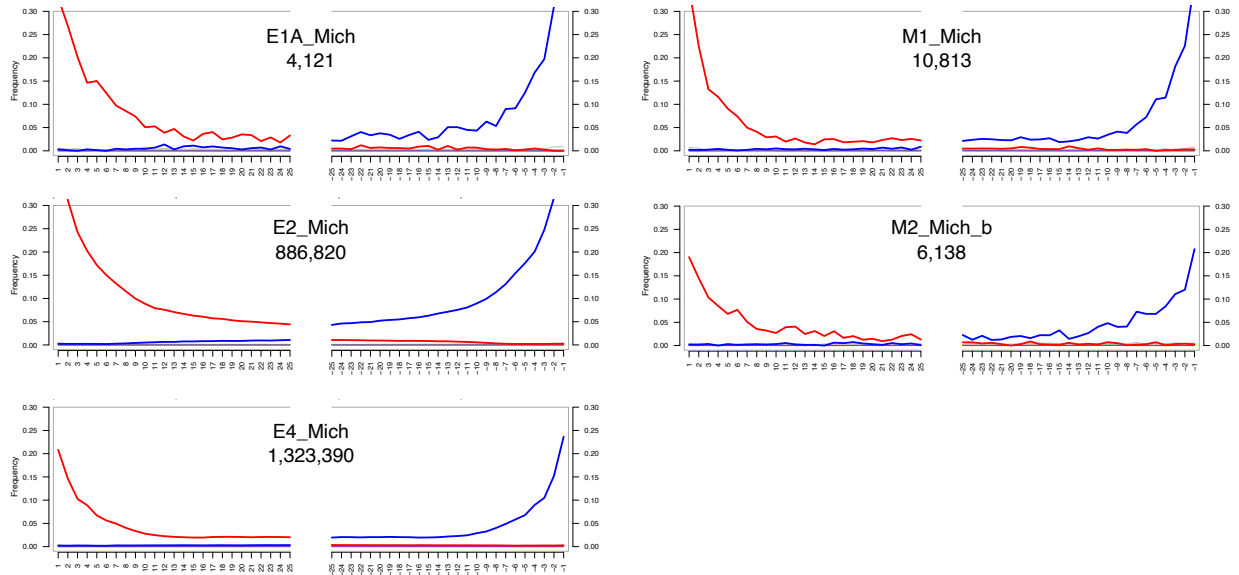


842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853

**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.



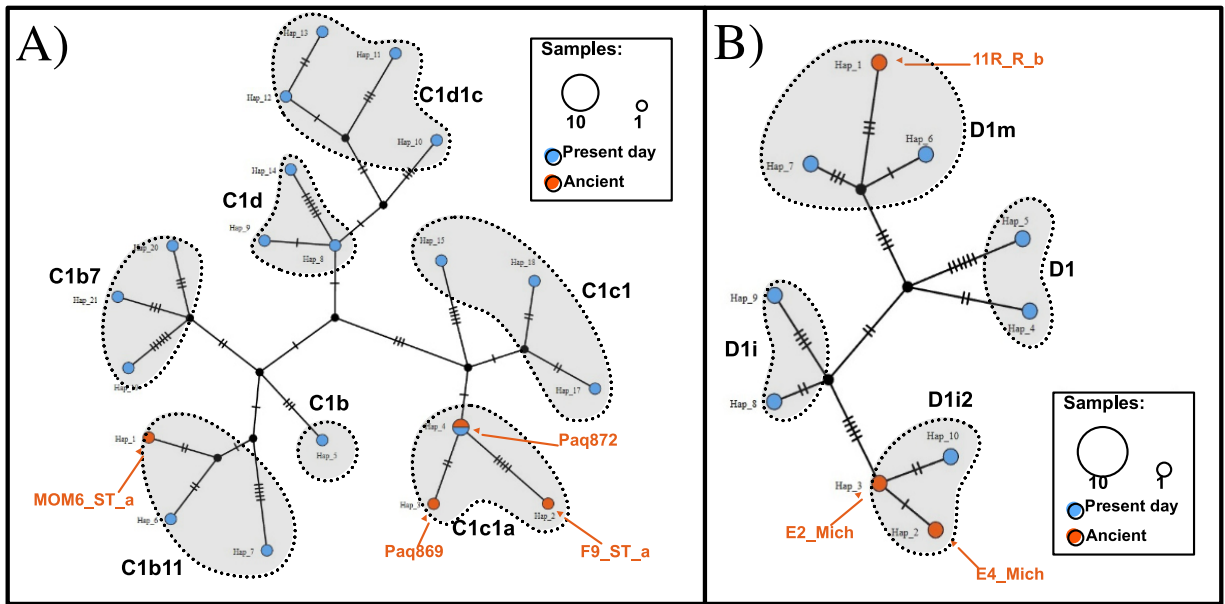
854  
855  
856



857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867

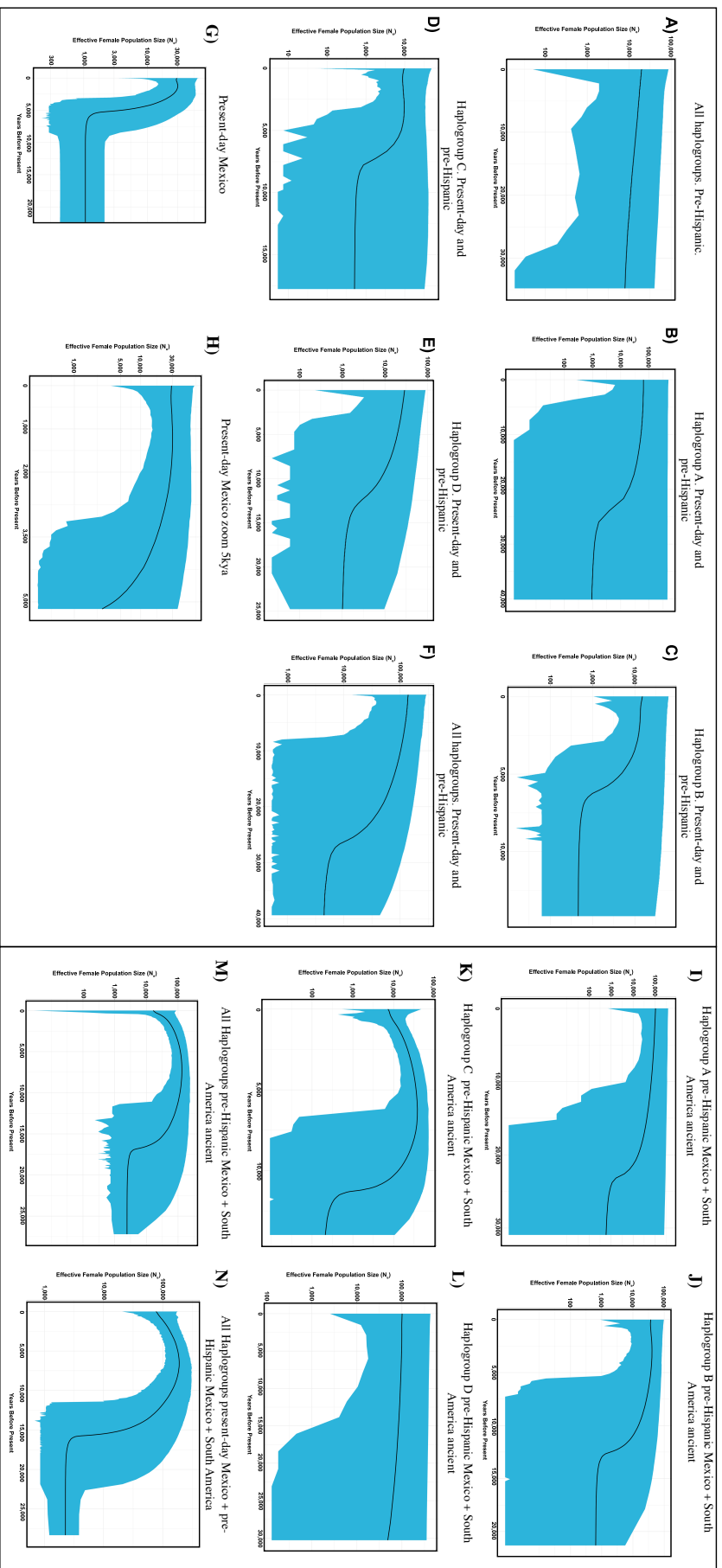
**Fig S4. Damage patterns of the genomic data from Michoacán pre-Hispanic individuals.** The number of reads included in the analysis is indicated below the individual's ID. The red color indicates the frequency of cytosine to thymine changes, while the blue color indicates guanine to adenine changes. 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.

868  
869  
870

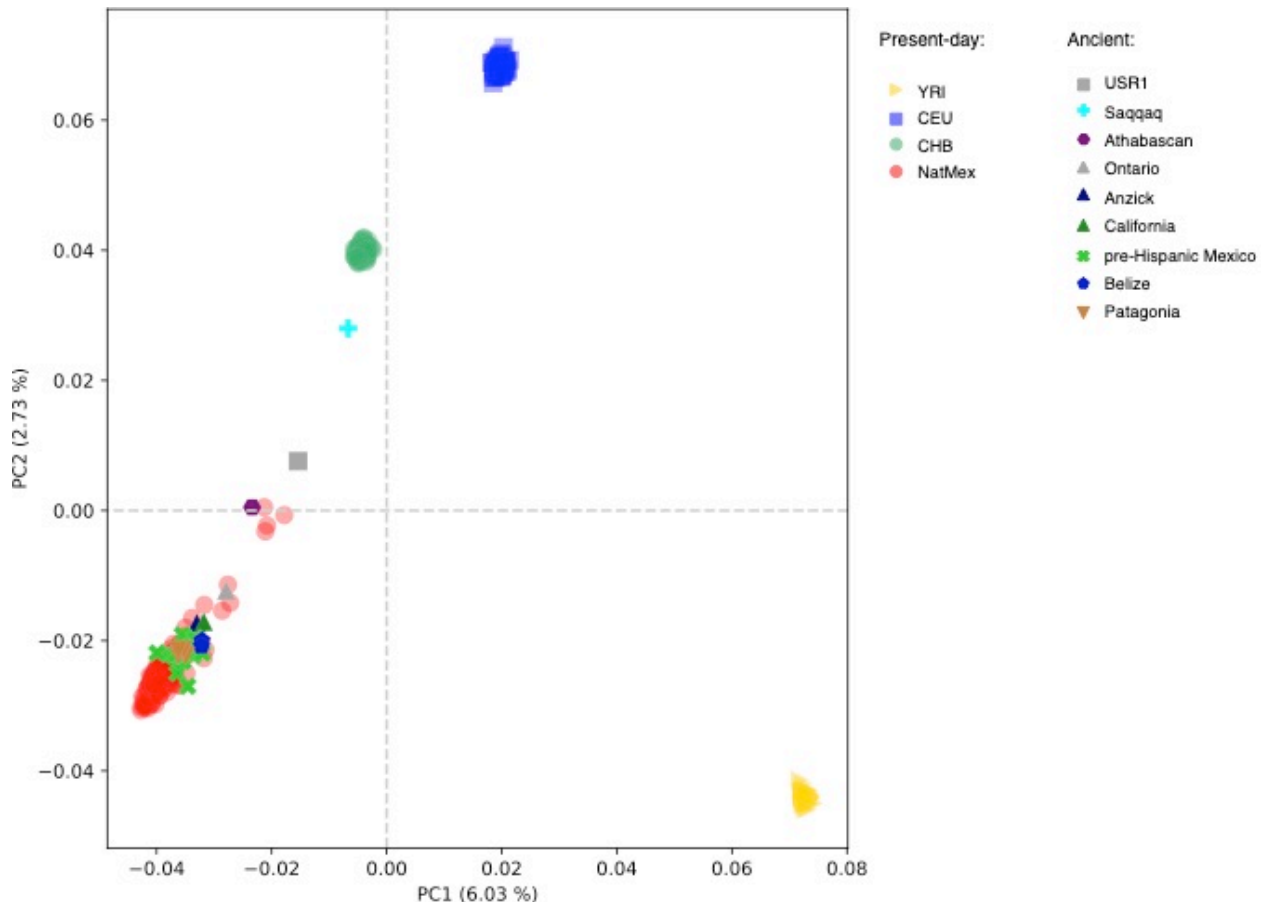


871  
872  
873  
874  
875  
876  
877

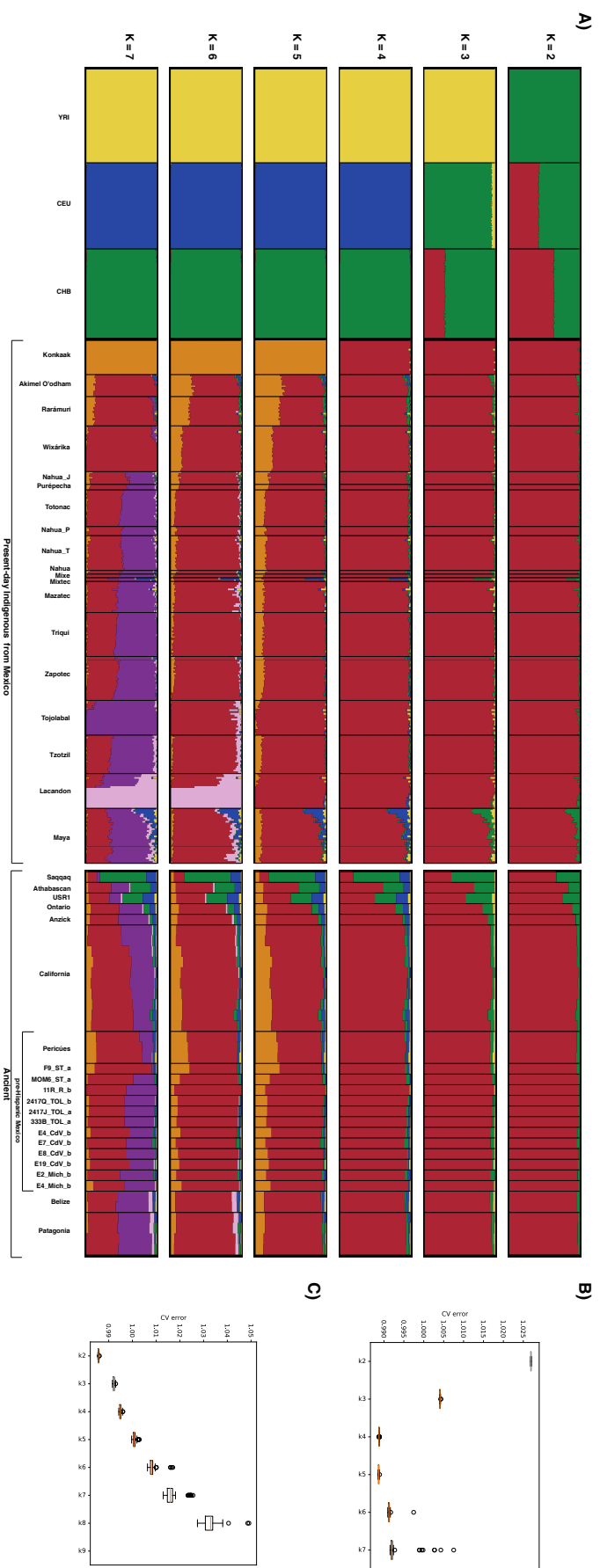
**Fig S5. Haplotype Networks for mitochondrial haplogroups of pre-Hispanic individuals from Mexico.** (A) Haplogroup C, (B) Haplogroup D. Sub-haplogroups are shaded with gray. Orange arrows point to ancient individuals from the present study and public available data used for the analysis (Table S4).



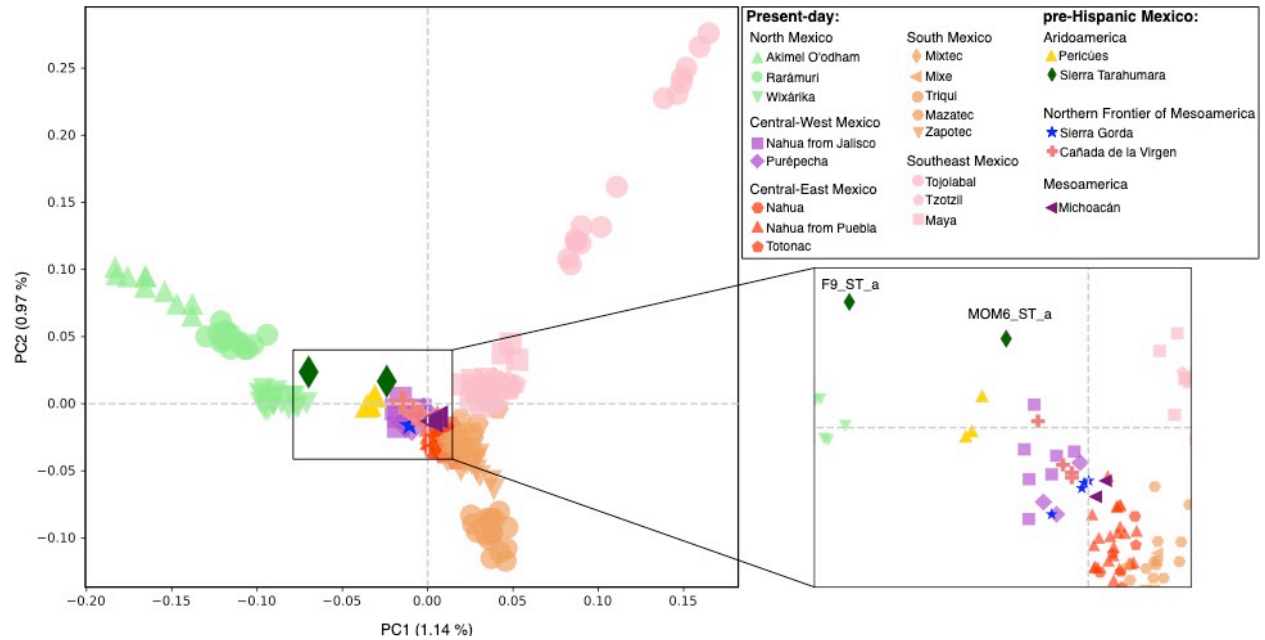
**Fig S6. Inference of past effective female population size through Extended Bayesian skyline plots (EBSP).** The EBSPs were constructed with ancient and present-day mitochondrial DNA. The median is shown as a black line, and the credibility interval of 95% is shown in blue.



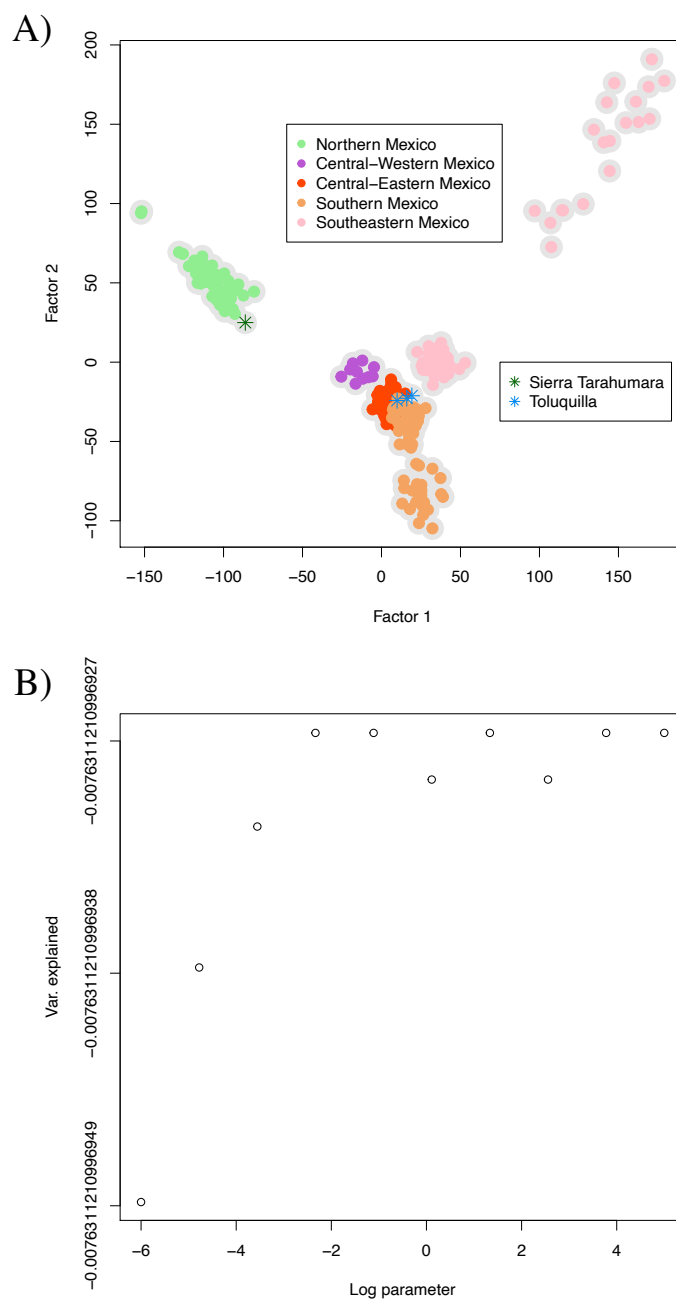
5 **Fig. S7. Global Principal Components Analysis (PCA) including pre-Hispanic individuals from this study.** The PCA includes ancient individuals Mexico from this and other studies, as well as ancient individuals from the Americas (Supplementary Table S1 and S3), Indigenous populations from Mexico (NatMex),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



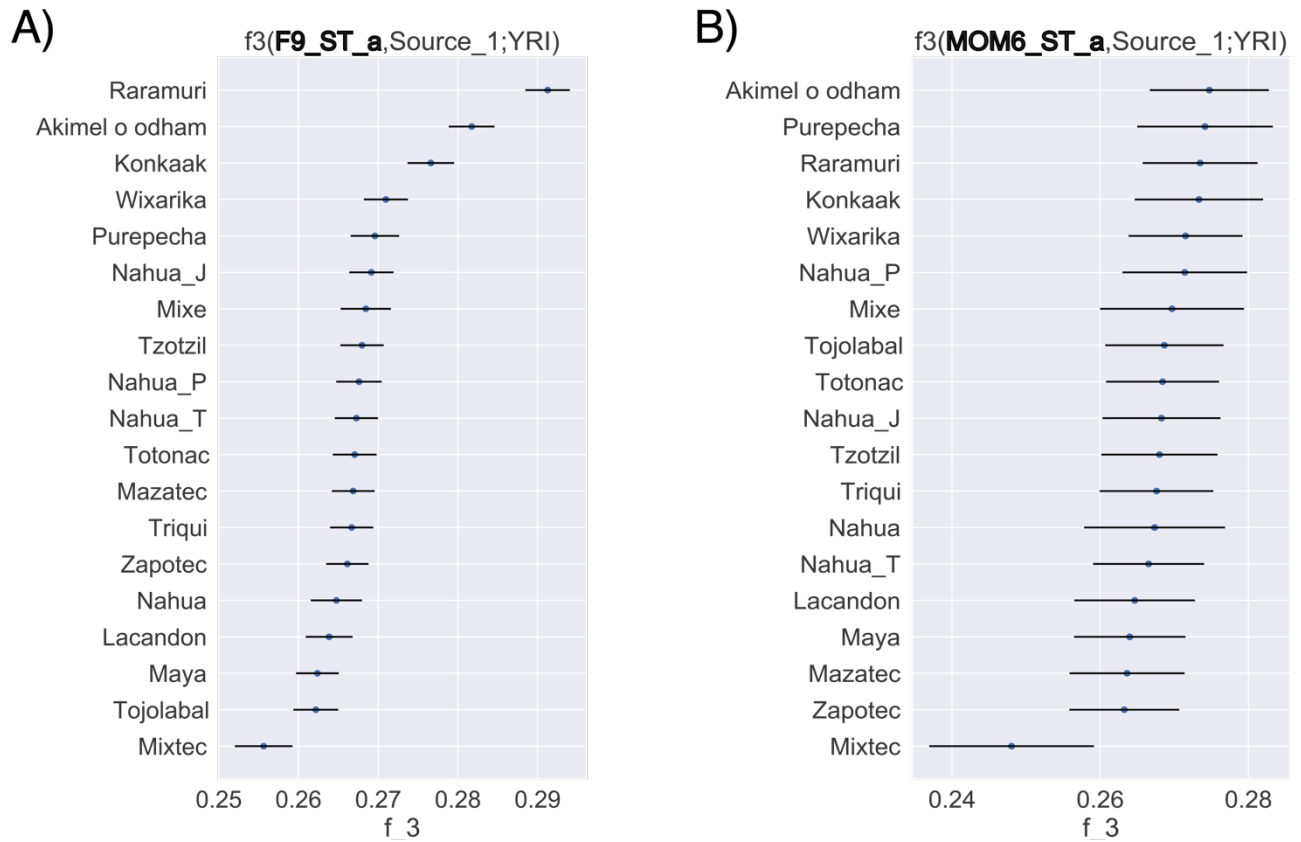
**Fig. S8. ADMIXTURE analysis including pre-Hispanic individuals from Mexico.** (A) Continental ADMIXTURE plot shown for K=2 through K=7. Individuals include continental references from Africa (YRI), Europe (CEU), and Asia (CHB) from 1000 Genomes as well as present-day Indigenous populations from Mexico, 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.

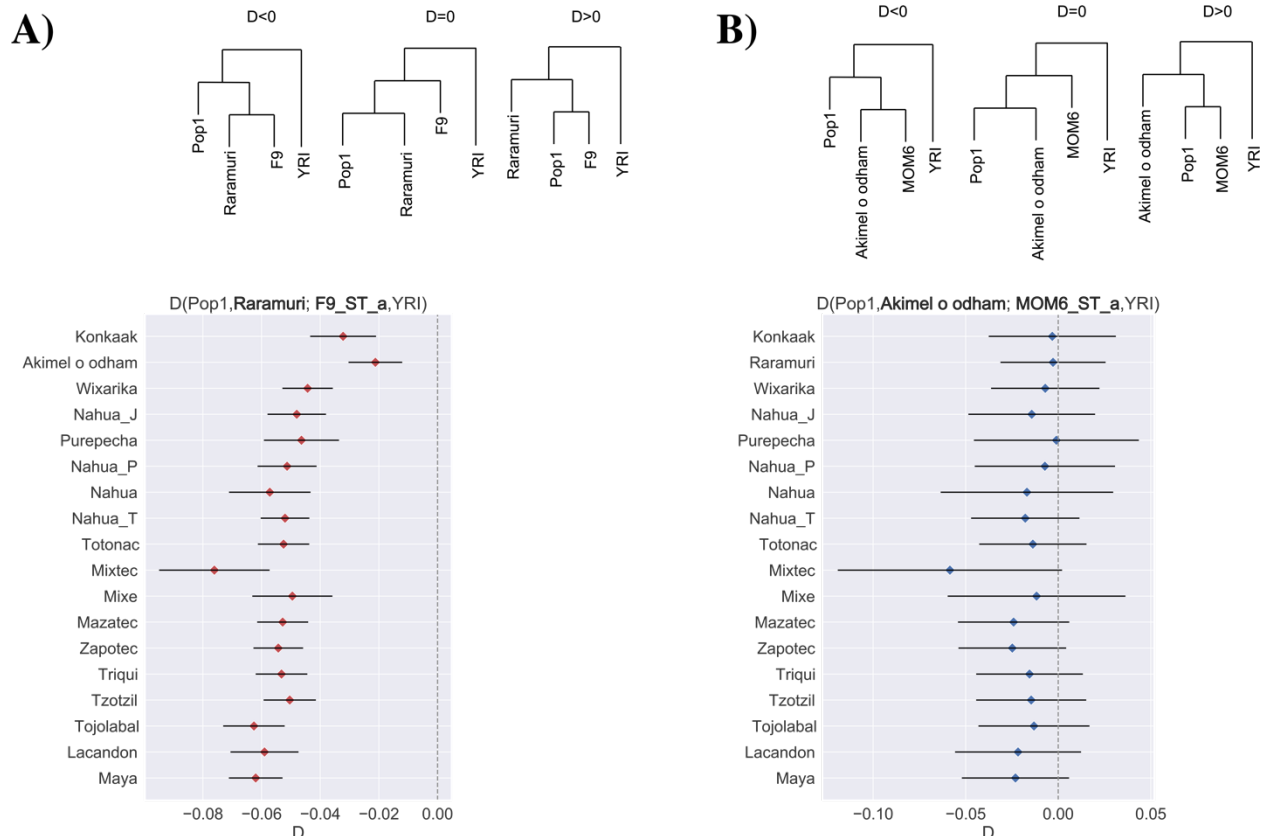


**Fig. S10. Temporal factor analysis (TFA) of ancient and present-day Indigenous individuals from Mexico.** (A) TFA includes the four pre-Hispanic individuals, F9\_ST\_a from Sierra Tarahumara and 2417Q\_TOL\_b, 2417J\_TOL\_a 333B\_TOL\_a from Toluquilla, Sierra Gorda. (B) Percentage of variance explained 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.

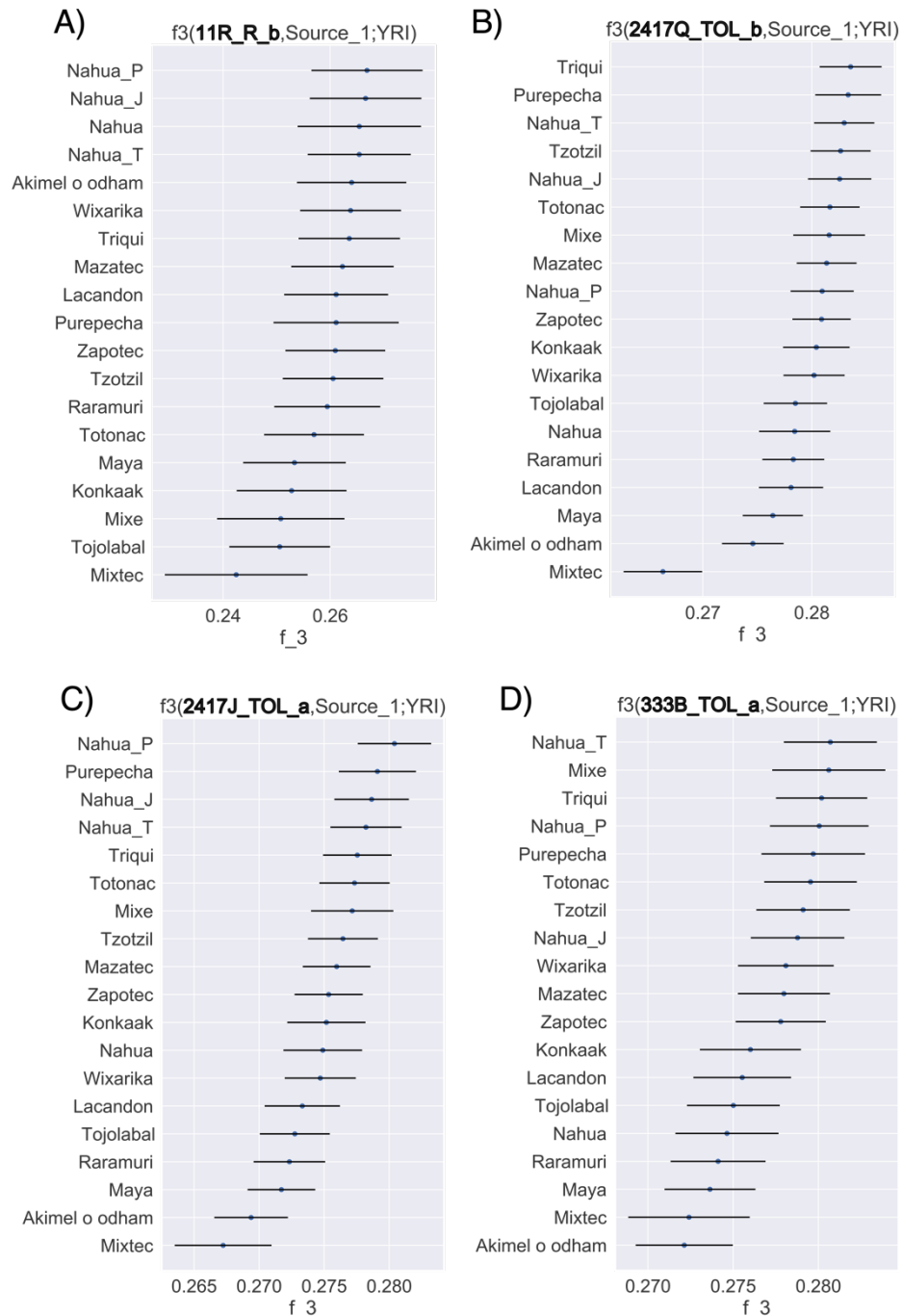


**Fig. S11. Shared generic drift between Sierra Tarahumara mummies and other present-day populations of Mexico.** Outgroups  $f_3$  statistics are shown for (A) Mummy F9\_ST\_a and (B) Mummy MOM6\_ST\_a. Higher values of  $f_3$  indicate higher shared genetic drift. Point estimates and one standard error 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.

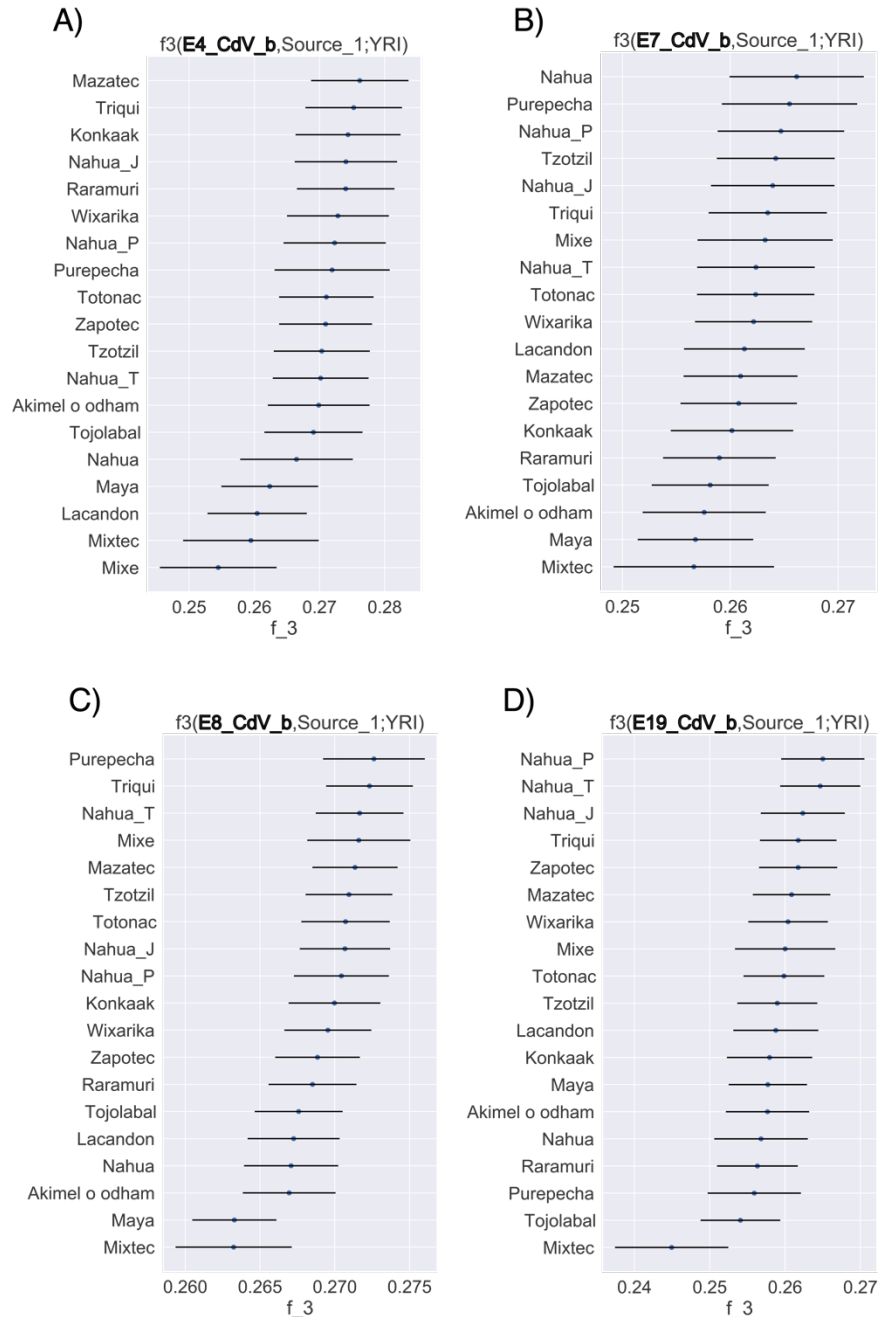




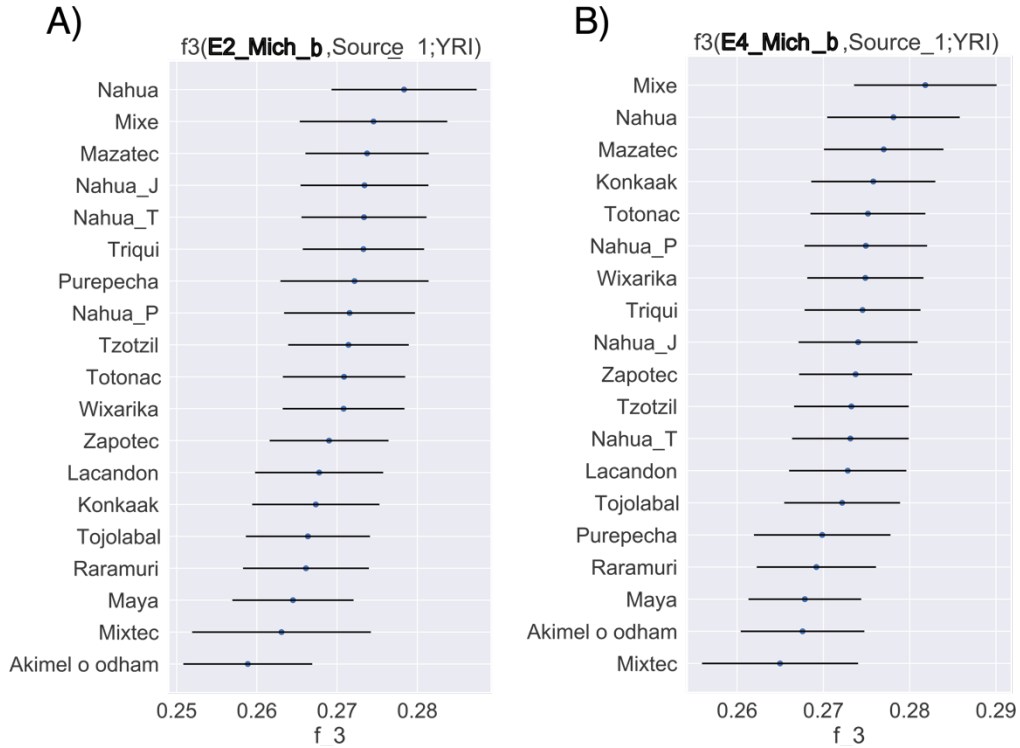
**Fig. S12. D statistics for pre-Hispanic individuals from Sierra Tarahumara and other present-day populations.** (A) D statistics in the form  $D(\text{Pop1}, \text{Raramuri}; \text{F9\_ST\_a}, \text{YRI})$ , (B) D statistics in the form  $D(\text{Pop1}, \text{Akimel o'odham}; \text{MOM6\_ST\_a}, \text{YRI})$ . 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 than with any other present-day population. The individual MOM6\_ST\_a seems to be equally related to all present-day Indigenous populations from Mexico.



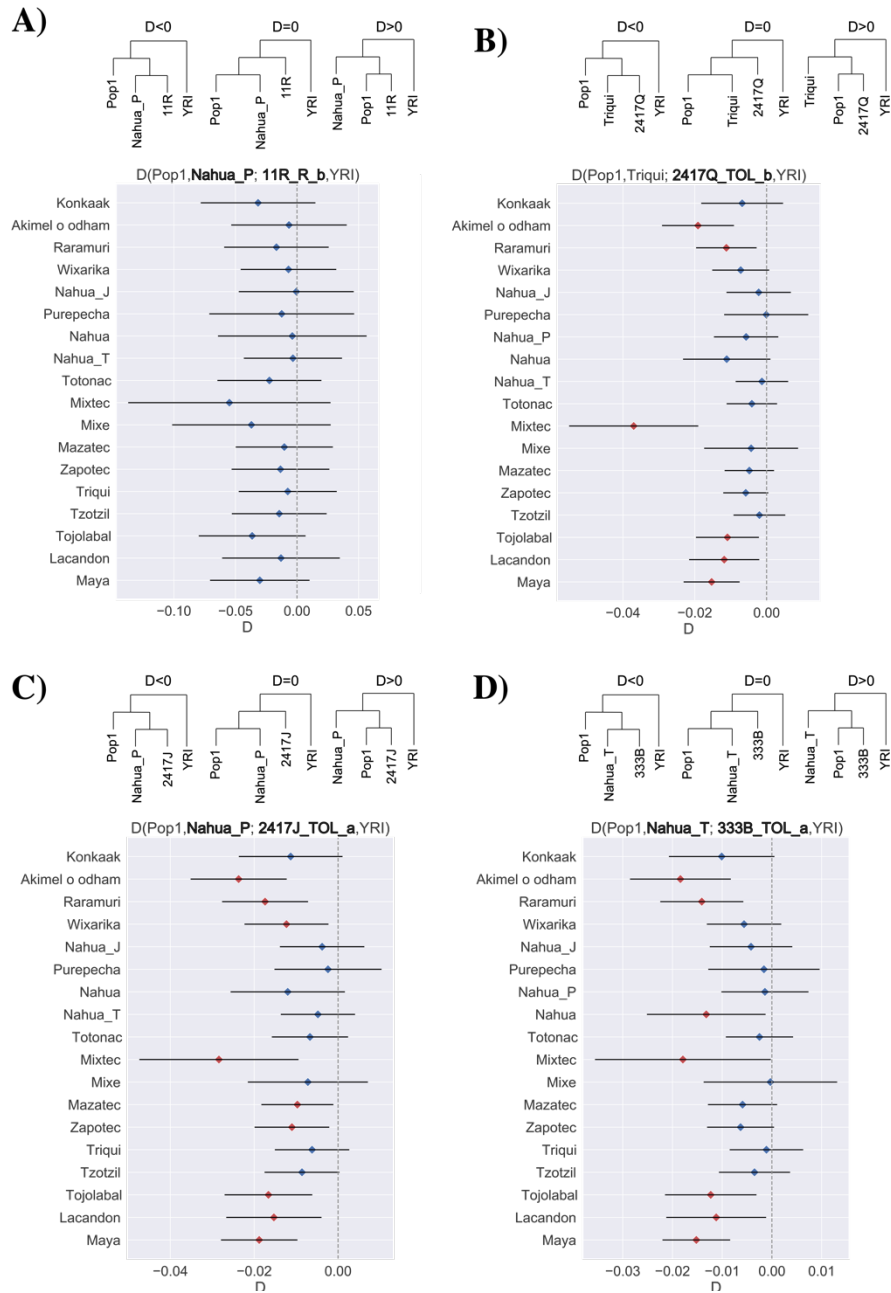
**Fig. S13. Outgroup f<sub>3</sub> statistics for pre-Hispanic individuals from Sierra Gorda and present-day Indigenous populations.** (A) 11R\_R\_b, (B) 2417Q\_TOL\_b, (C) 2417J\_TOL\_a, and (D) 333B\_TOL\_a. Higher values of f<sub>3</sub> indicate higher shared genetic drift. Point estimates and one standard error are shown. All individuals show a similar genetic drift shared with all present-day Indigenous populations from Mexico.



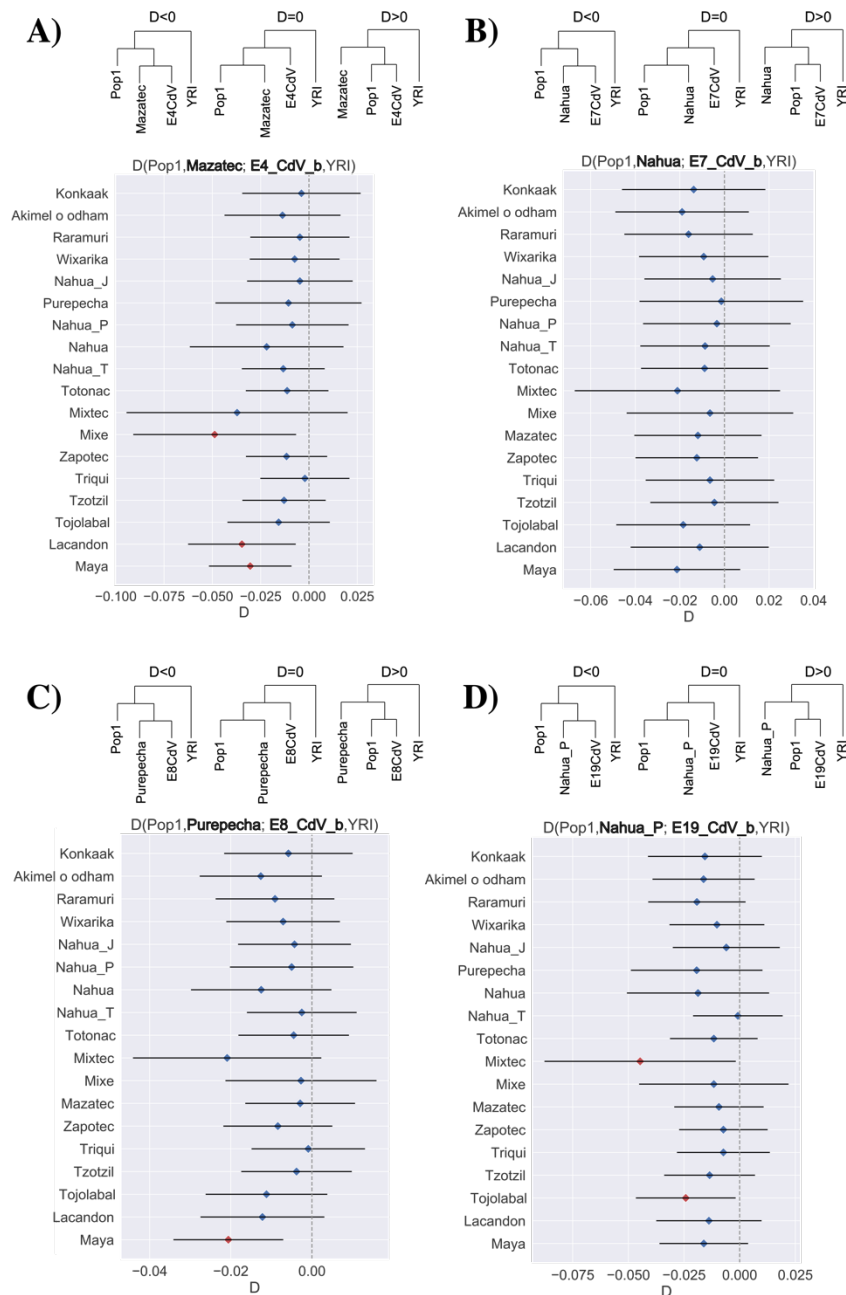
**Fig. S14. Outgroups f3 statistics for pre-Hispanic individuals from Cañada de la Virgen (CdV) and present-day Indigenous populations.** (A) E4\_CdV\_b, (B) E7\_CdV\_b, (C) E8\_CdV\_b, and (D) E19\_CdV\_b. Higher values of f3 indicate higher shared genetic drift. Point estimates and one standard error are shown. All pre-Hispanic CdV individuals show a similar genetic drift shared with all present-day Indigenous populations from Mexico, except with Mixtec.



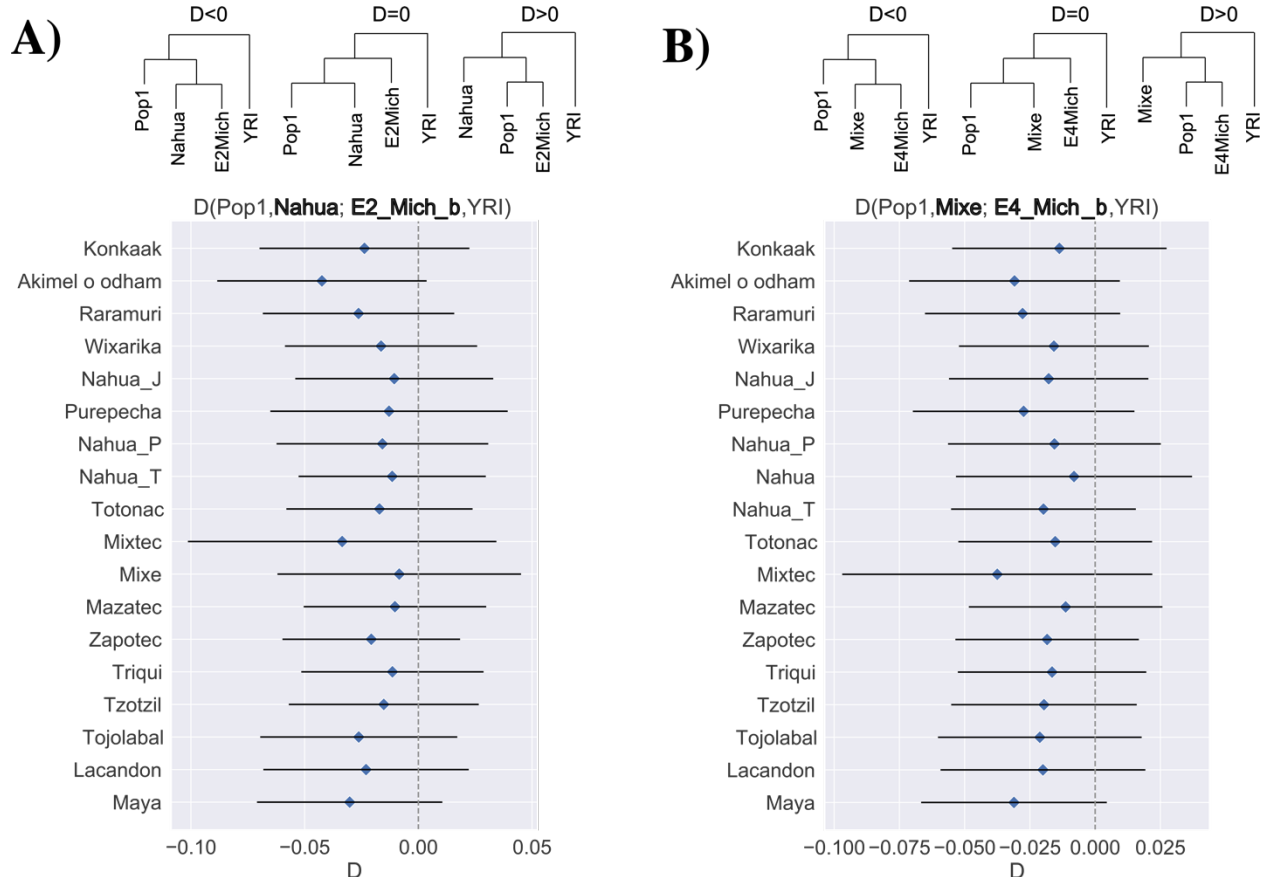
**Fig. S15. Outgroups  $f_3$  statistics for pre-Hispanic individuals from Michoacán and present-day Indigenous populations.** (A) E2\_Mich\_b, and (B) E4\_Mich\_b. Higher values of  $f_3$  indicate higher shared genetic drift. Point estimates and one standard error are shown. Both individuals show a similar genetic drift shared with all present-day Indigenous populations from Mexico.



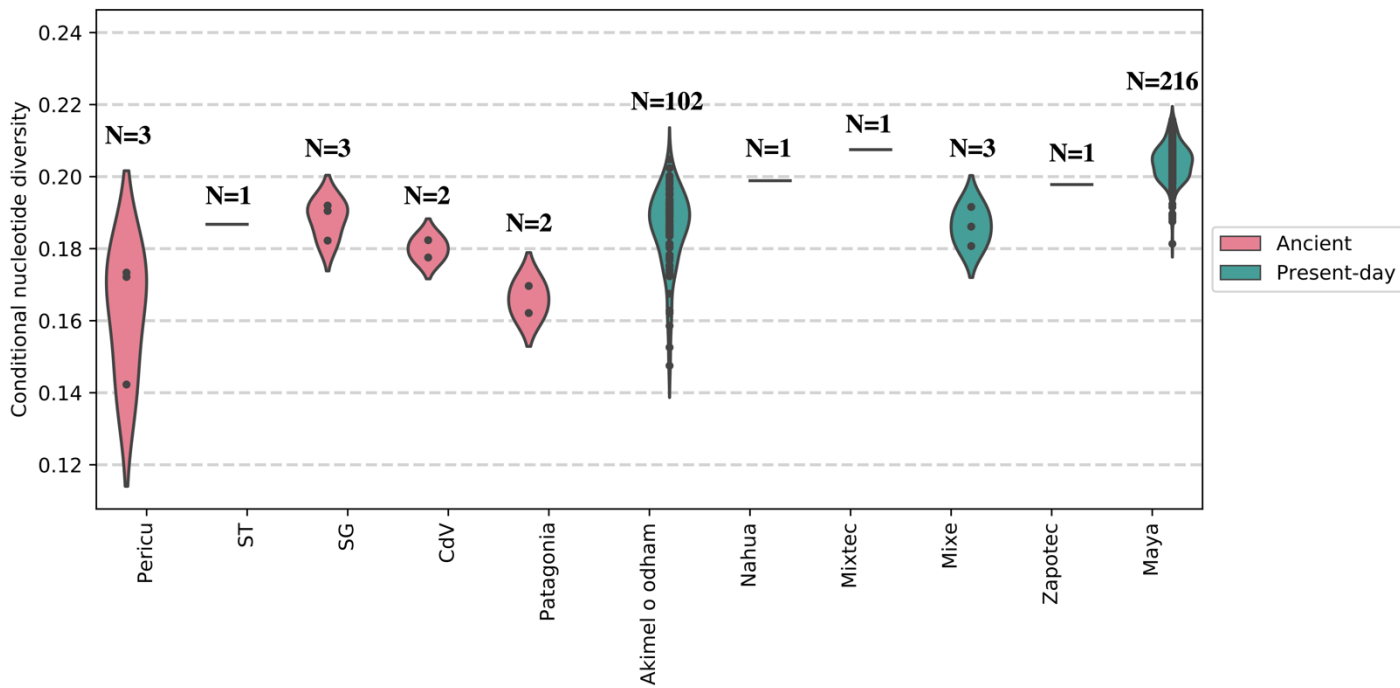
**Fig. S16. D statistics for pre-Hispanic individuals from Sierra Gorda and present-day Indigenous populations.** (A)  $D(\text{Pop1, Nahua\_P; 11R\_R\_b, YRI})$ , (B)  $D(\text{Pop1, Triqui; 2417Q\_TOL\_b, YRI})$ , (C)  $D(\text{Pop1, Nahua\_P; 2417J\_TOL\_a, YRI})$  and (D)  $D(\text{Pop1, Nahua\_T; 333B\_TOL\_a, YRI})$ . 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)  $D(\text{Pop1, Mazatec; E4\_CdV\_b, YRI})$ , (B)  $D(\text{Pop1, Nahua; E7\_CdV\_b, YRI})$ , (C)  $D(\text{Pop1, Purepecha; E8\_CdV\_b, YRI})$  and (D)  $D(\text{Pop1, Nahua\_P; E19\_CdV\_b, YRI})$ . 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 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.

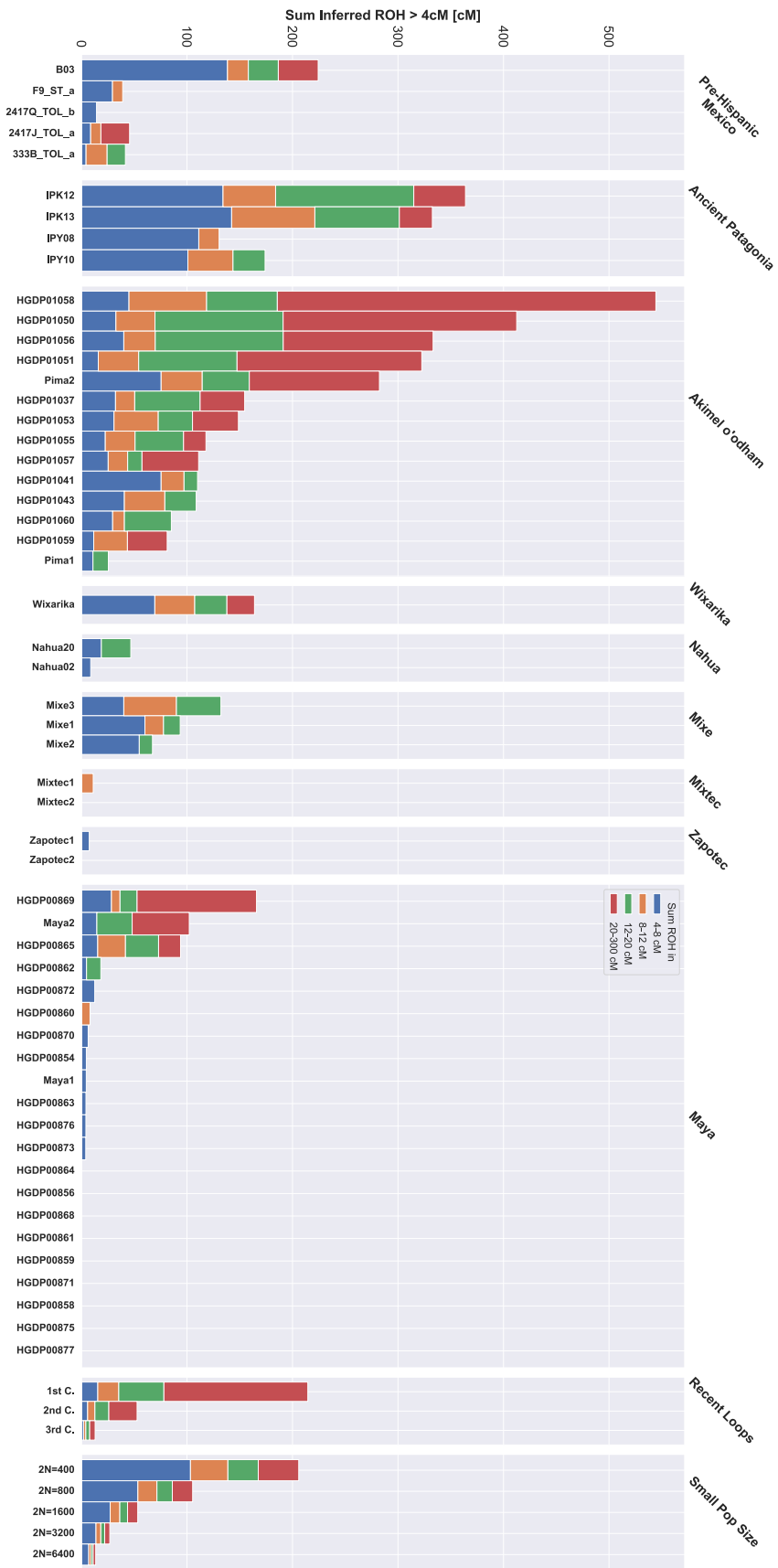


**Fig. S18. D statistics for pre-Hispanic individuals from Michoacán and present-day Indigenous populations.** (A)  $D(\text{Pop1, Nahua; E2\_Mich\_b, YRI})$ , (B)  $D(\text{Pop1, Mixe; E4\_Mich\_b, YRI})$ . Expected tree topologies according to D value are drawn on the top of the plot, individual IDs 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.

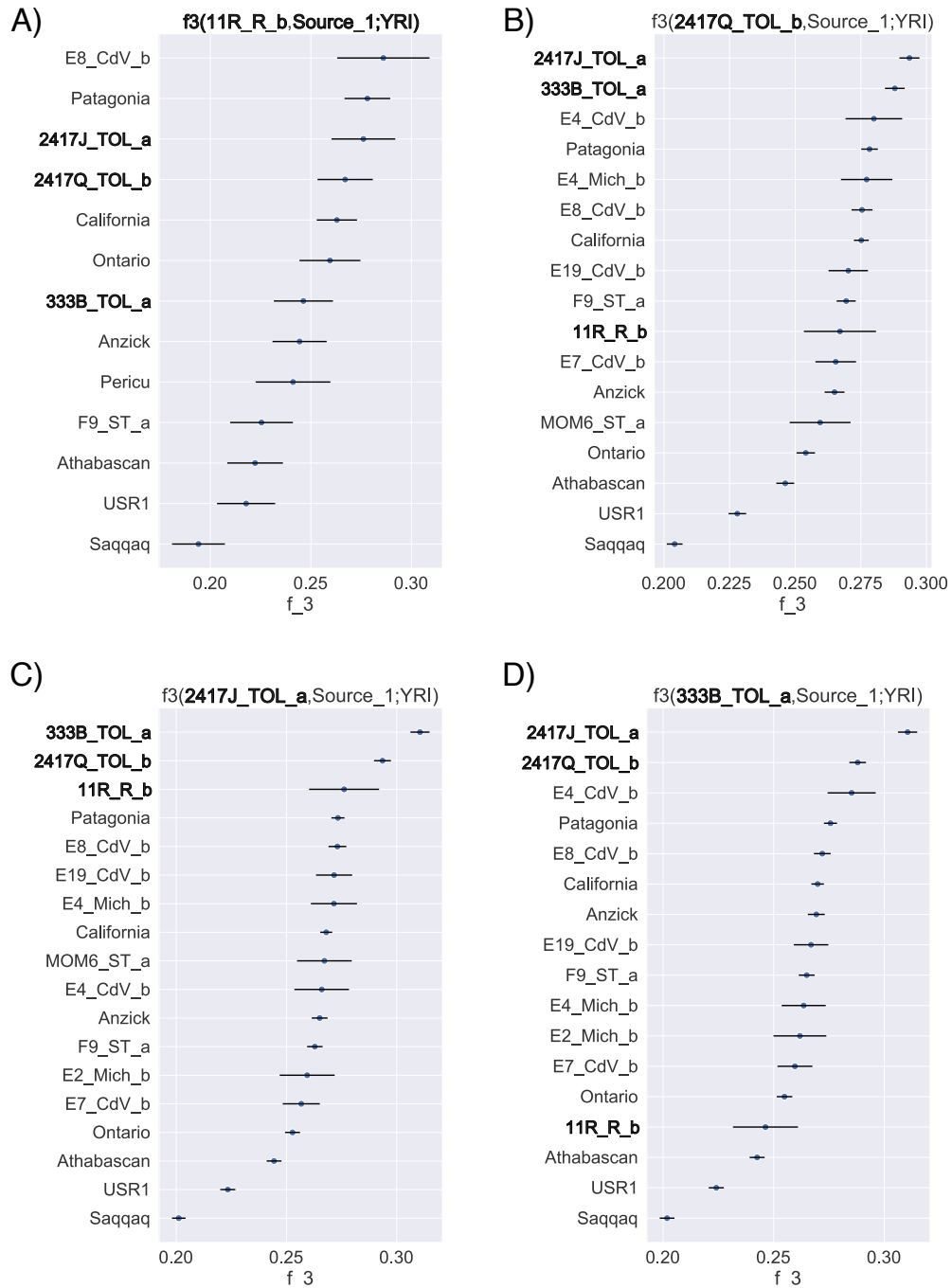


**Fig. S19. Conditional Nucleotide Diversity of pairs of ancient pre-Hispanic individuals and present-day Indigenous populations from Mexico.** Violin plot indicating the CND values estimated by pairs of individuals from the same archaeological site or population. The N value above each violin indicates 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).

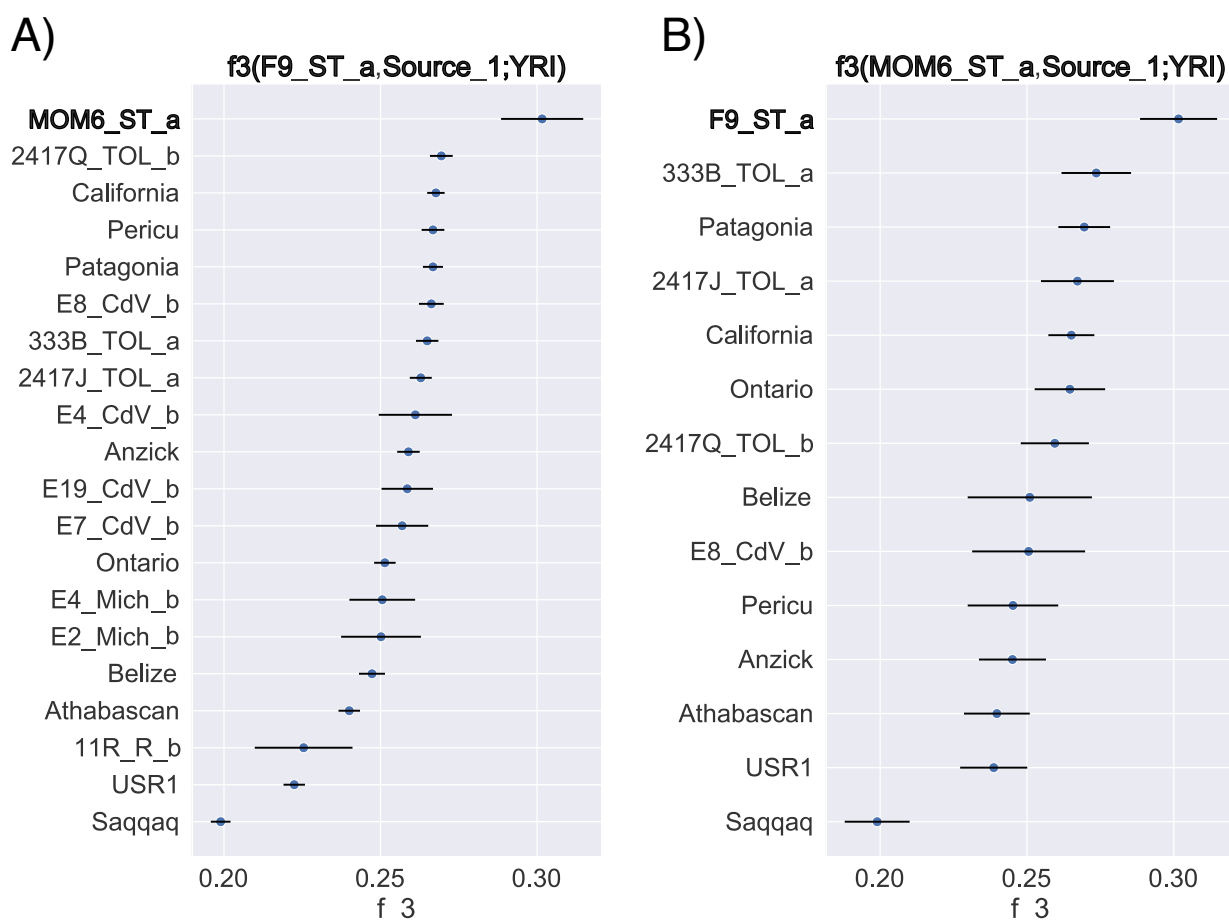




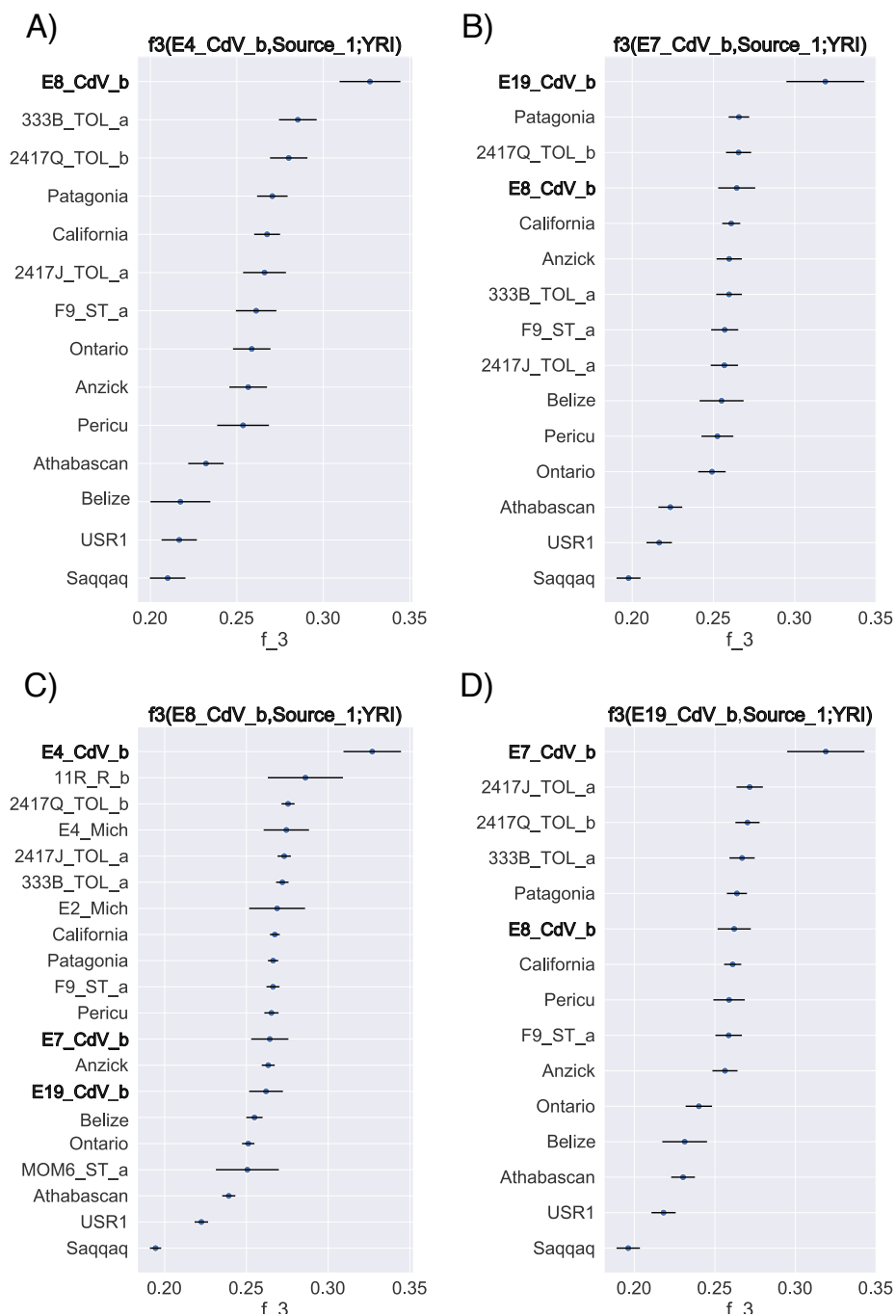
**Fig. S20. Runs of homozygosity in pre-Hispanic and present-day Indigenous individuals from Mexico.** Four pre-Hispanic individuals from Patagonia were analyzed for comparison. Length of ROH are shown in different colors. Pre-Hispanic individual B03 (Pericu) and the present-day Akimel O'odham population 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 1<sup>st</sup> cousins, 2<sup>nd</sup> cousins and 3<sup>rd</sup> cousins. Small Pop Size refers to the expected values in individuals from small populations.



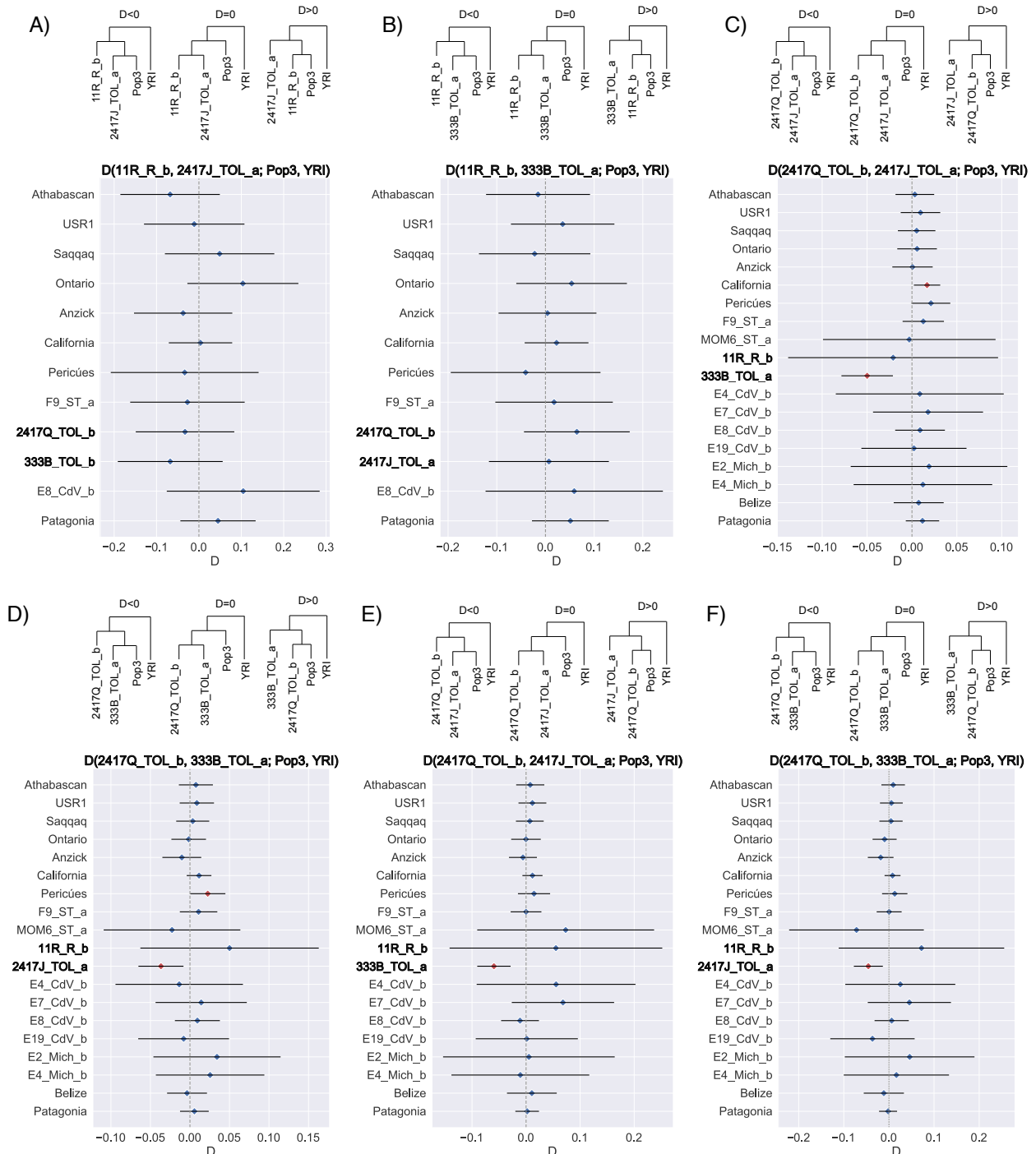
**Fig. S21. Outgroup  $f_3$  statistics for pre-Hispanic individuals from Sierra Gorda.** (A) 11R\_R\_b, (B) 2417Q\_TOL\_b, (C) 2417J\_TOL\_a, and (D) 333B\_TOL\_a compared with ancient individuals from Mexico and the Americas. Sierra Gorda individuals are shown in bold. Higher values of  $f_3$  indicate higher shared genetic drift. Point estimates and one standard error are shown.



**Fig. S22. Outgroup  $f_3$  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  $f_3$  indicate higher shared genetic drift.

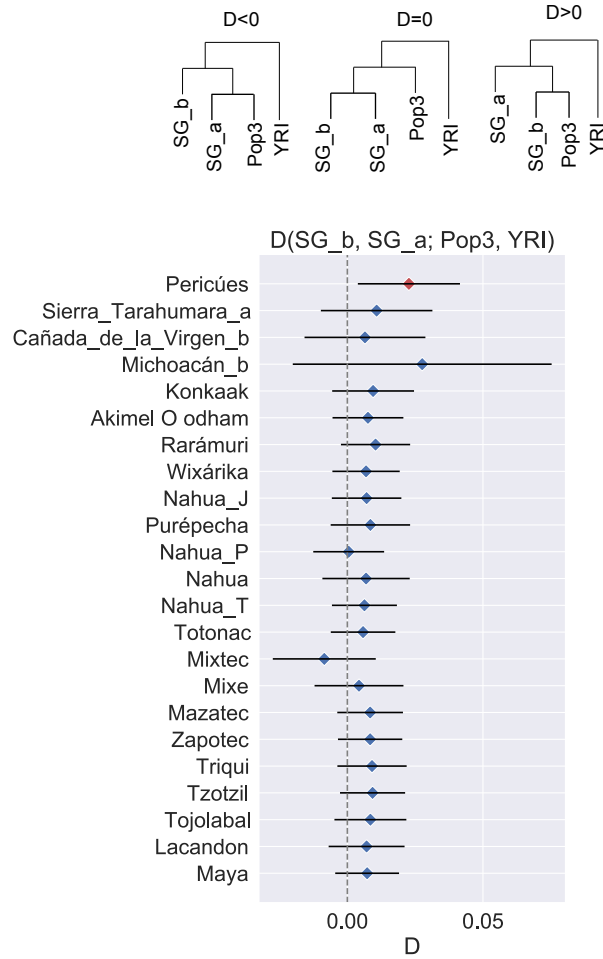


**Fig. S23. Outgroup  $f_3$  statistics for pre-Hispanic individuals from Cañada de la Virgen.** (A) E4\_CdV\_b, (B) E7\_CdV\_b, (C) E8\_CdV\_b, and (D) E19\_CdV\_b compared with ancient individuals from Mexico and the Americas. Individuals from Cañada de la Virgen are highlighted in bold. Higher values of  $f_3$  indicate higher shared genetic drift. Point estimates and one standard error are shown.

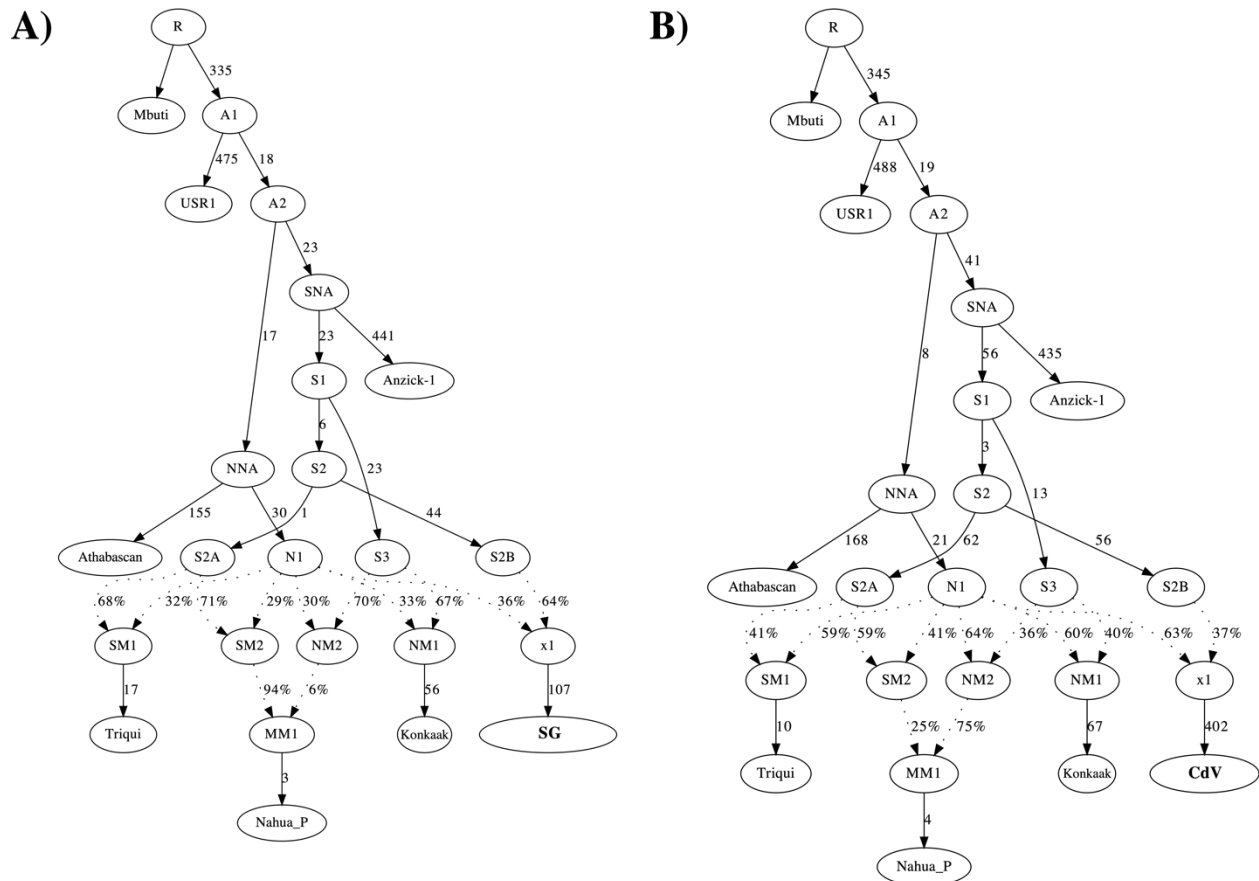


**Fig. S24. D statistic values between the pre-Hispanic individuals from Sierra Gorda in the form D (SG\_b, SG\_a; Pop3, YRI).** SG\_b is a pre-drought pre-Hispanic individual from Sierra Gorda, SG\_a is a post-drought pre-Hispanic individual from Sierra Gorda. Pop3 is any ancient individual from Mexico 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.

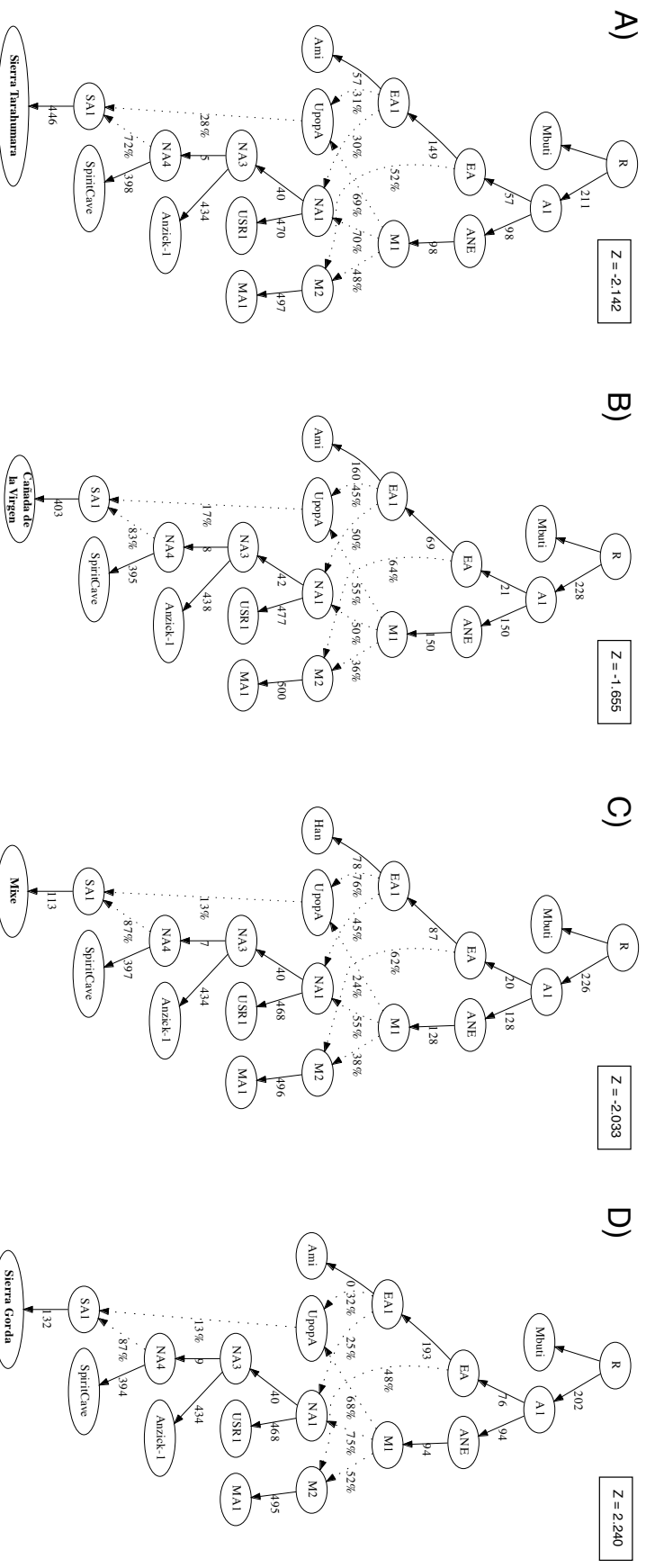


**Fig. S25. D statistic values between the pre-Hispanic individuals from Sierra Gorda and pre-Hispanic/present-day Indigenous populations from Mexico.** D statistics in the form  $D(\text{SG}_b, \text{SG}_a; \text{Pop3}, \text{YRI})$ .  $\text{SG}_b$  includes the pre-Hispanic individuals 2417Q\_TOL\_b and 11R\_R\_b, while  $\text{SG}_a$  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  $D=0$  ( $|Z|>3$ ).

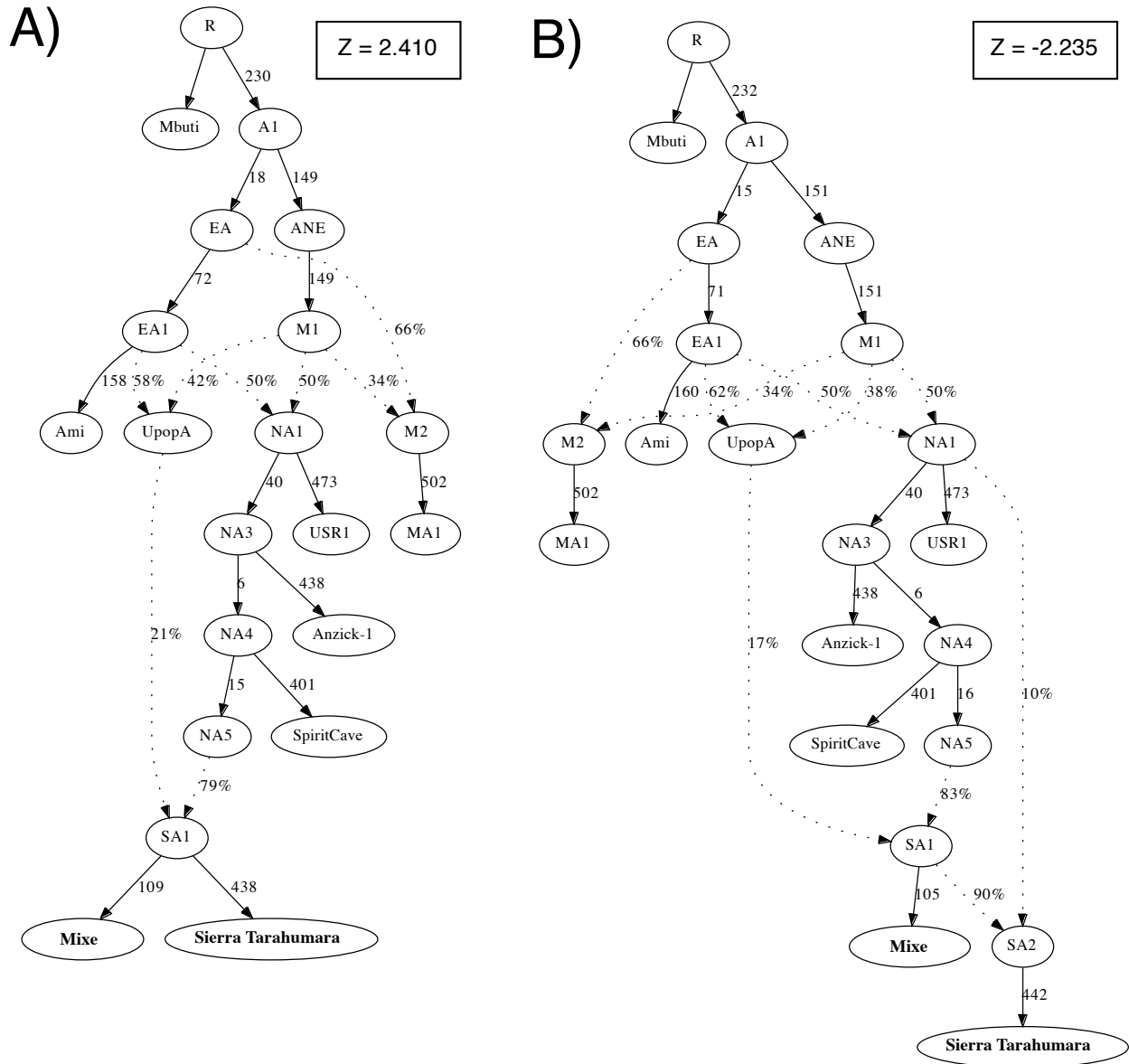


**Fig. S26. Genetic ancestry proportions for the individuals from Central Mexico, inferred with qpGraph.** Base model includes Anzick as a reference for the SNA and Athabaskan as a reference for the NNA branch. Konkaak was used a proxy for Northern Indigenous population from Mexico, Triqui as South proxy 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  $|Z\text{-score}|$  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  $|Z\text{-score}|$  of 2.821. All f statistics are within 1.75 standard error.

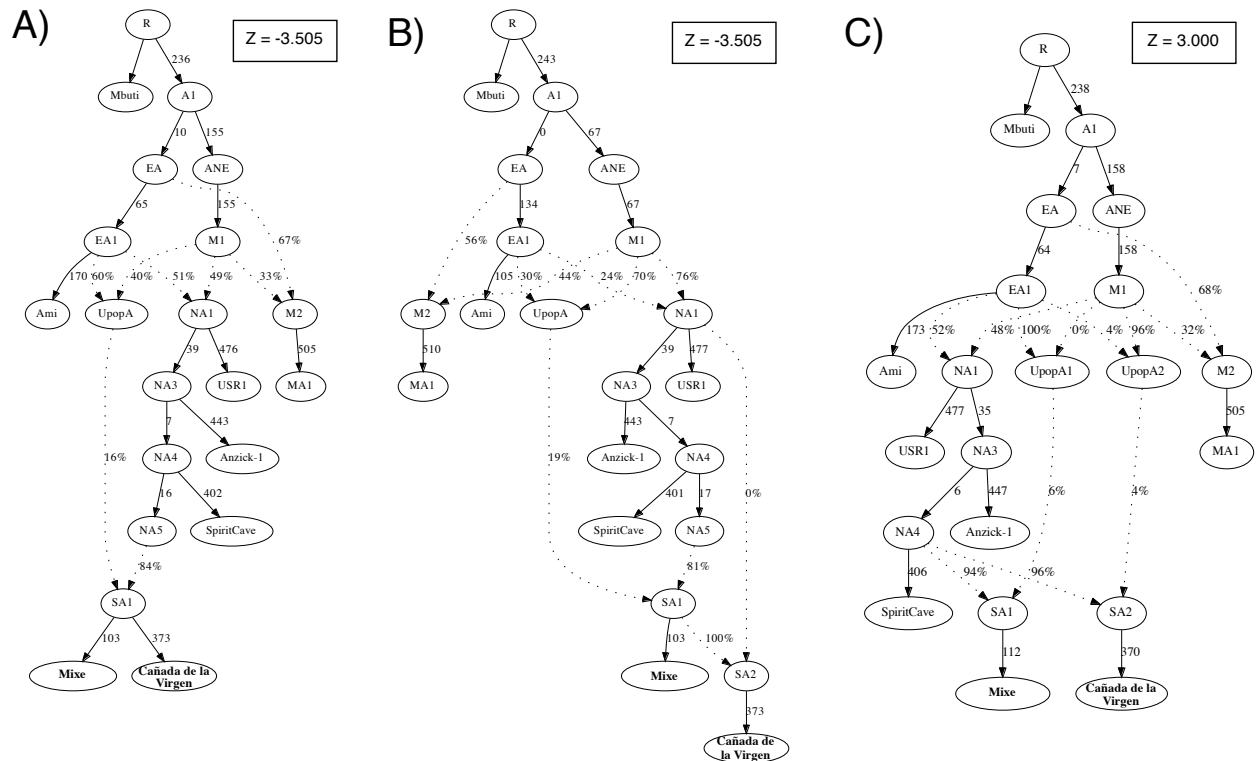




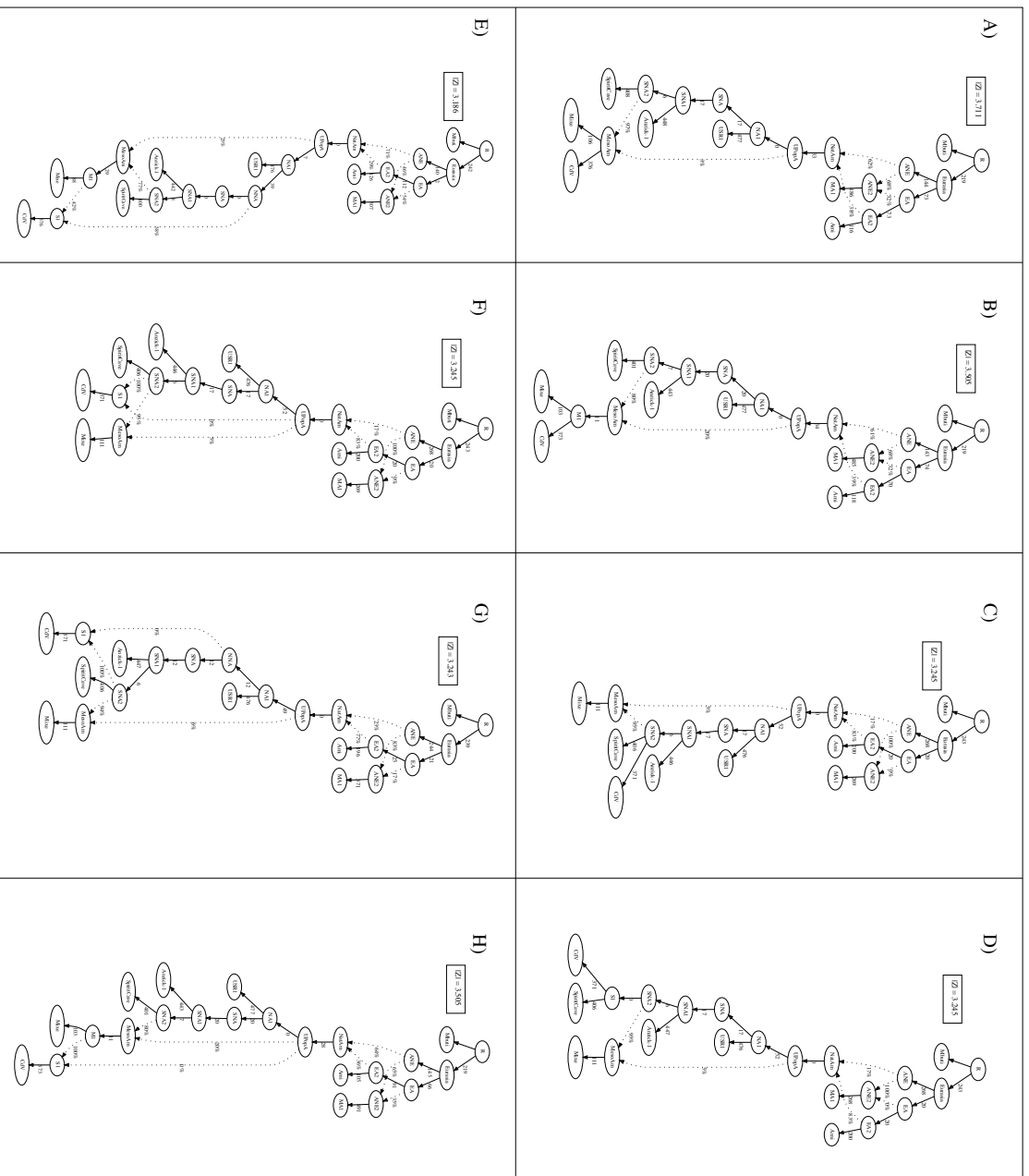
**Fig. S27. Presence of ghost population genetic ancestry in pre-Hispanic from Sierra Tarahumara and Cañada de la Virgen, and present-day Mixe population inferred with qpGraph.** Base model includes Mbuti as an outgroup, Ami as the East Asian, MA1 as the ancient North Eurasian, USRI as the ancient Beringian,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

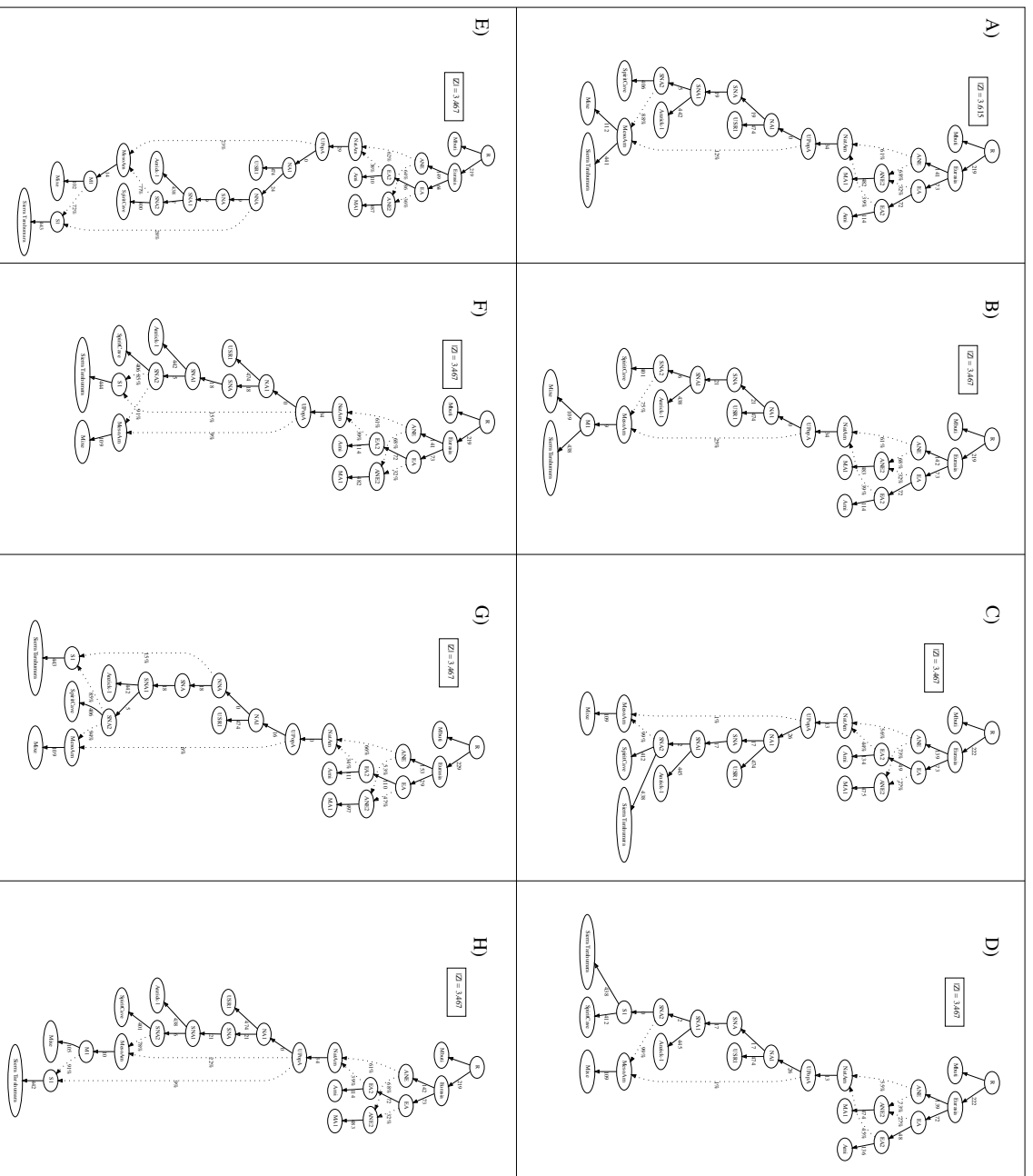


**Fig. S29. Presence of ghost population genetic ancestry in pre-Hispanic from Cañada de la Virgen.** Base model includes Mbuti as an outgroup, Ami as the East Asian, MA1 as the ancient North Eurasian, USR1 as the ancient Beringian, Anzick-1 and Spirit Cave as the NNA. (A) The model with UpopA contribution 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.



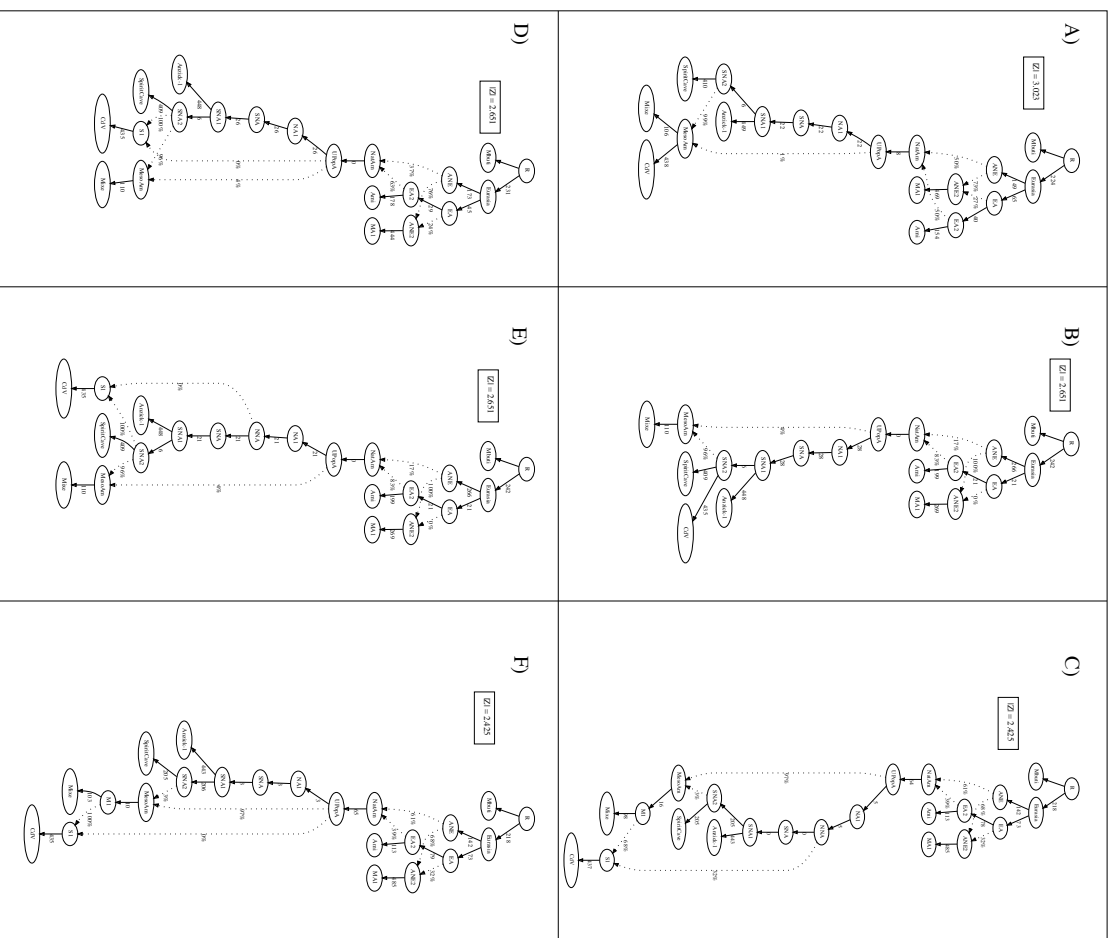
**Fig. S30. qpGraphs modeling the ghost genetic ancestry in Cañada de la Virgen (CdV) and UPopA modeled with Native American clade as its ancestor.** (A) Model with Cañada de la Virgen and Mixe in the same clade. (B) Same as A but including an intermediate lineage 'MI' between Mesamerica, Mixe and Cañada de la Virgen. (C) Model with Cañada de la Virgen and SpiritCave 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 SpiritCave in the same clade with a contribution of UPopA. (G) Model with Cañada de la Virgen and SpiritCave 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 U'Popá 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 'MI' between Mesamerica, Mixe and Sierra Tarahumara. (C) Model with

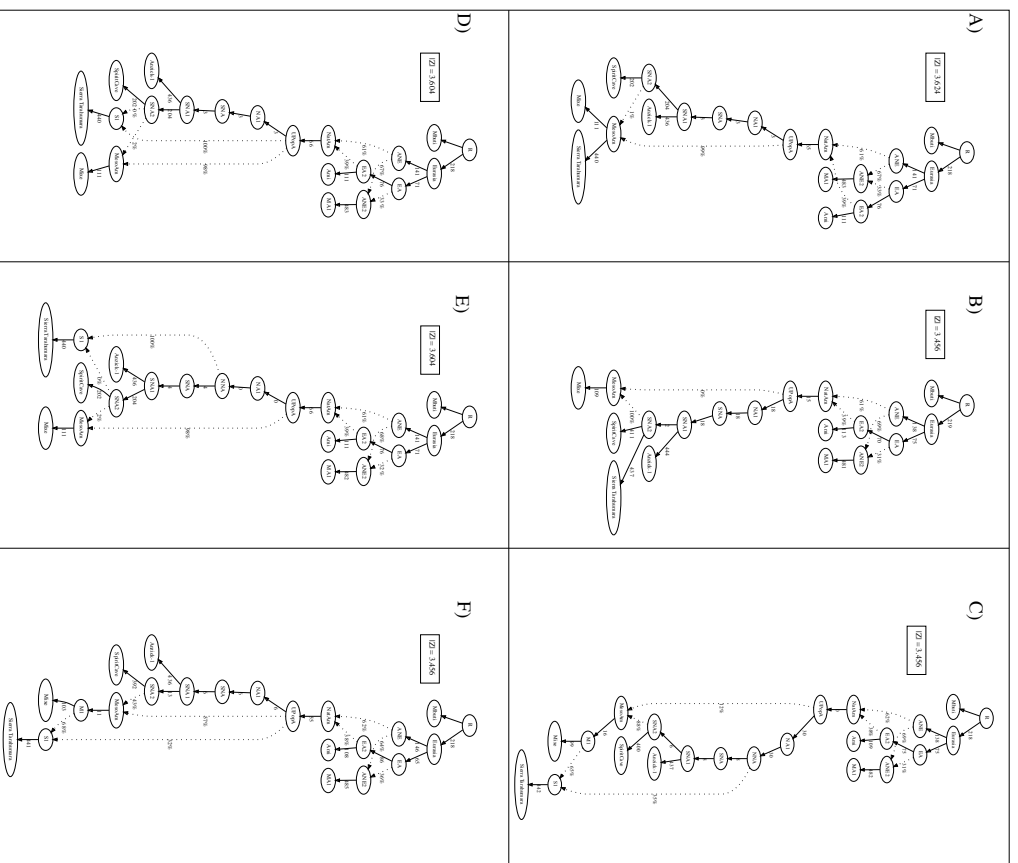
Sierra Tarahumara and SpiritCave in the same clade. (D) Same as C but including an intermediate lineage 'S1' between SNA2, SpiritCave and Sierra Tarahumara. (E) Model with Sierra Tarahumara coming from 'M1' with a contribution of 'NNA'. (F) Model with Sierra Tarahumara and SpiritCave in the same clade with a contribution of UPopA. (G) Model with Sierra Tarahumara and SpiritCave 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 SpiritCave in the same clade. (C) Cañada de la Virgen coming from 'MI' with a**



contribution of NNA. (D) Cañada de la Virgen and SpiritCave in the same clade with a contribution of UPopA. (E) Cañada de la Virgen and SpiritCave 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 SpiritCave in the same clade. **(C)** Sierra Tarahumara coming from 'M1' with a contribution of NNA. **(D)** Sierra Tarahumara and SpiritCave in the same clade with a contribution of UPopA. **(E)** Sierra Tarahumara and SpiritCave 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.