

Demographic history and genetic structure in pre-Hispanic Central Mexico

Viridiana Villa-Islas¹, Alan Izarraras-Gomez¹, Maximilian Larena², Elizabeth Mejía Perez Campos³, Marcela Sandoval-Velasco⁴, Juan Esteban Rodríguez-Rodríguez¹, Miriam Bravo-Lopez¹, Barbara Moguel^{1,5}, Rosa Fregel⁶, Jazeps Medina Tretmanis⁷, David Alberto Velázquez-Ramírez⁸, Alberto Herrera-Muñoz³, Karla Sandoval⁹, María A. Nieves-Colón^{9,10}, Gabriela Zepeda¹¹, Fernando A Villanea¹², Eugenia Fernández Villanueva Medina¹³, Ramiro Aguayo-Haro¹³, Cristina Valdiosera¹⁴, Alexander Ioannidis¹⁵, Andrés Moreno-Estrada⁹, Flora Jay¹⁶, Emilia Huerta-Sanchez⁷, Federico Sánchez-Quinto¹⁷, María C. Ávila-Arcos^{1*}.

Affiliations:

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

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

³National Institute of Anthropology and History, Querétaro, Mexico

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

⁵Centro de Geociencias, UNAM Juriquilla

⁶Department of Biochemistry, Microbiology, Cell Biology and Genetics, Universidad de La Laguna, San Cristóbal de La Laguna, Spain

⁷Center for Computational Molecular Biology, Brown University, Providence, RI, USA

⁸LOEWE Center for Synthetic Microbiology–SYNMIKRO, Philipps-Universität Marburg, Marburg, Germany

⁹Unit of Advanced Genomics, National Laboratory of Genomics for Biodiversity (LANGEBIO), CINVESTAV, Irapuato, Guanajuato, Mexico

¹⁰Department of Anthropology, University of Minnesota Twin Cities, USA

¹¹National Institute of Anthropology and History, Guanajuato, Mexico

¹²Department of Anthropology, University of Colorado Boulder, USA

¹³National Institute of Anthropology and History, Michoacán, Mexico

¹⁴Department of Archaeology and History, La Trobe University, Melbourne, Australia

¹⁵Institute for Computational and Mathematical Engineering, Stanford University, USA

¹⁶Université Paris-Saclay, Centre National de la Recherche Scientifique, Inria, Laboratoire de Recherche en Informatique UMR 8623, Bâtiment 650 Ada Lovelace, 91405 Orsay Cedex, France

¹⁷Computational Genomics, Instituto Nacional de Medicina Genómica, Ciudad de México, Mexico

*Correspondence to: mavila@liigh.unam.mx

Abstract

Aridoamerica and Mesoamerica are two distinct cultural areas that hosted numerous pre-Hispanic civilizations between 2,500 BCE and 1,521 CE. The division between these regions shifted southward due to severe droughts ca. 1,100 years ago, allegedly driving demographic changes and population replacement in some sites in central Mexico. Here, we present shotgun genome-wide data from 12 individuals and 26 mitochondrial genomes from eight pre-Hispanic archaeological sites across Mexico, including two at the shifting border of Aridoamerica and Mesoamerica. We find population continuity spanning the climate change episode and a broad preservation of the genetic structure across present-day Mexico for the last 2,300 years. Lastly, we identify a contribution to pre-Hispanic populations of northern and central Mexico from an ancient unsampled ‘ghost’ population.

Introduction

Before European colonization, the territory comprised by present-day Mexico was home to numerous civilizations that occupied two main cultural areas: Aridoamerica in northern Mexico, inhabited mainly by hunter-gatherers, and Mesoamerica in central and southern Mexico (Fig. 1), where some of the largest agriculture-based civilizations flourished between 2,500 BCE and 1,521 CE (1, 2). The distinction between Aridoamerica and Mesoamerica is based on the cultural characteristics and subsistence strategies of the peoples that inhabited them, as well as the ecological features of each region (3, 4). Archaeological evidence indicates that the border between these two areas shifted southward between 900 and 1,300 CE following multi-decadal droughts (5). This period is also known as the Medieval Warm Period in other regions of the world (6). Moreover, this climate change period allegedly led to population replacements in the Northern Frontier of Mesoamerica (NFM) by semi-nomadic hunter-gatherers (“Chichimecas”) from Aridoamerica (7), and the fall of some pre-Hispanic societies and the abandonment of Mesoamerican cities in central (5, 8) and southeast (8–10) Mexico.

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

Here, we generate the most extensive set of complete mitogenomes and shotgun genome-wide data to date from pre-Hispanic individuals from Mexico spanning a time transect of 830 years: 7 genomes with coverage ranging from 0.01x to 0.14x, and five genomes with coverage ranging from 0.5x to 4.7x depth. Moreover, we sequenced 26 mitogenomes that span a time transect of 1,600 years. We include three previously published ancient genomes from northern Mexico as well as nine ancient mitogenomes yielding a total of 15 ancient genomes and 35 ancient mitogenomes from the Mexican territory. We analyze our data jointly with publicly available datasets of other ancient Native Americans outside of Mexico and present-day Indigenous populations from Mexico. This large compendium of ancient genomic data allows us to study the pre-Hispanic genetic structure in the territory occupied by Mexico, answering longstanding questions regarding population dynamics

2 at the NFM, disentangling the complex structure of central Mexico populations, and revealing a
3 previously unknown ancient contribution of an unsampled genetic lineage to Mexican populations.

4 **Results**

7 **Ancient genomic data from pre-Hispanic individuals from Northern and Central Mexico**

8 To gain insights into the population dynamics of Mexico in the last 2,300 years, we investigated
9 the genetic diversity, population structure, and demographic history of the populations inhabiting
0 the Mexican territory prior to European colonization. We screened archeological samples from 39
1 pre-Hispanic individuals. Thirty-seven out of the 39 pre-Hispanic individuals were sampled from
2 sites in Mesoamerica. These samples were taken from individuals located in seven archaeological
3 sites from central Mexico (n=37). The central Mexico sites include three in the Sierra Gorda (SG)
4 in Querétaro state, namely Toluquilla (TOL, n=9), Ranas (R, n=3) and a cave in Cadereyta de
5 Montes (CCM, n=1); Cañada de la Virgen (CdV) in Guanajuato state (n=16), and the Zaragoza
6 (n=4), Tanhuato (n=1) and La Mina (n=3) sites in Michoacán state (Mich). Furthermore, we
7 produced additional genome-wide data for two mummies from the Sierra Tarahumara, northern
8 Mexico (Aridoamerica) (Table S1), which were sequenced at lower depths in a previous study (14)
9 (Fig. 1).

0 For each sample, we evaluated the extent of endogenous ancient DNA (aDNA) via shotgun
1 sequencing. Out of the 39 samples, only three had endogenous content above 10% (13 – 39%)
2 (Table S2). We then generated additional sequence data for these three samples using shotgun
3 sequencing, while for the remaining libraries, we performed whole-genome or mitochondrial DNA
4 (mtDNA) capture (16, 17) (Table S2). For all of these, we verified the presence of the characteristic
5 misincorporation patterns of aDNA using mapDamage (18) (Fig. S1-S4). We obtained informative
6 amounts of genome-wide data for 12 individuals (0.01 – 4.7 x). Ten were from Mesoamerica; SG
7 (n=4), CdV (n=4), and Michoacán (n=2), and two from the Sierra Tarahumara in Aridoamerica
8 (Table S1). We integrated these data with previously reported low-depth genomes (0.09 – 0.3 x)
9 from three ancient Pericúes from Mexico (Aridoamerica) reported in (14, 15), resulting in a total of
0 15 pre-Hispanic individuals from Mexico for which we had genome-wide data. Our analysis also
1 include ancient genomic data from previous studies for 21 individuals from across the Americas
2 (Table S3)(14, 15, 19–24). Additionally, we reconstructed the mitochondrial genomes for 26
3 individuals across Mexico (5.7 – 1284.8x), spanning 1,600 years (320 BCE to 1,351 CE). Twenty-
4 four from Mesoamerica; SG (n=13), CdV (n=6), and Michoacán (n=5), and 2 individuals from
5 Aridoamerica; Sierra Tarahumara (n=2) (14) (Table S1 and S2). These mitochondrial genomes
6 were analyzed along with previously published mitochondrial genomes from Mexico (14, 15, 25)
7 (Table S4).

8
9 Individuals' names were stated according to burial and individual numbers provided by
0 archaeologists. Additionally, we added two labels, one consisting of 1-3 letters that refer to the
1 archaeological site and the other corresponding to one letter that represents the period of the
2 individuals: a letter 'b' for before the longstanding droughts and 'a' for those from the period during
3 or after the longstanding droughts. An underscore character '_' separates these three pieces of
4 information. Throughout the text we refer to all individuals using these IDs.

Genetic structure and diversity in uniparental markers

All individuals carried one of the mtDNA haplogroups found in Indigenous populations of the Americas: A (n=10), B (n=8), C (n=4), and D (n=4) (26) (Table S1) (see supplementary text). The pre-Hispanic mtDNA haplogroup distribution resembles the one observed in present-day Mexico (27–29), showing an overall continuity of matrilineal genetic structure for at least 2,300 years. We observed haplogroup C to be predominant in Aridoamerica in the Sierra Tarahumara, with decreasing frequency toward Mesoamerica. This gradient has been previously observed in ancient and contemporary populations from Mexico (25, 30–36). Specifically for central Mexico, we found the presence of all four autochthonous haplogroups, similar to previous observations on ancient and present-day populations inhabiting this region (27–31, 37), which highlights the higher genetic diversity and heterogeneity in this particular area (38, 39), and which is in agreement with larger civilizations and greater population mobility than in Aridoamerica. Furthermore, we detected unreported variants in the mtDNA PhyloTree (40) for seven individuals assigned to the sub-haplogroups A2d (n=4), B2c (n=2), and B2l (n=1). The absence of these variants could reflect a poor sampling of Native American mitochondrial genomes in the PhyloTree database or a loss of these haplotypes as a consequence of drift that includes the effect of a population decline following European colonization as reported in other Native American populations (41, 42). Accordingly, we propose adding these new unreported variants to the PhyloTree database (Table S5).

Notably, in SG we found the sub-haplogroup A2d in six out of 13 pre-Hispanic individuals, including the oldest in our dataset, P_CCM_b (320 BCE) (Table S1). Out of this six A2d individuals, four belong to the pre-drought period according to dating based either on C14 or the archaeological context, and two to the drought periods. A haplotype median-joining network for the haplogroup A2 shows a clustering of the SG individuals carrying the sub-haplogroup A2d despite a time transect of 1,480 years (Fig. 2A). This observation supports a population continuity, at least of the female population despite the severe droughts in 900 – 1,300 CE in this region of the NFM. For the other haplogroups B, C, and D, when individuals from the same site share the same haplogroup their sub-haplogroups tend to cluster together (Fig. S5).

To estimate the past female effective population size, we merged our dataset of pre-Hispanic mtDNA sequences (n=16, selected after alignments, see methods) with 164 publicly available ancient (n=9) and modern (n=155) mtDNA sequences from Mexico (Table S4 and S6). We generated an extended Bayesian Skyline Plot (EBSP) for all present-day haplogroups, for all pre-Hispanic haplogroups, for each haplogroup (present-day + pre-Hispanic), and for all haplogroups together merging present-day and pre-Hispanic (Fig. 2B and Fig. S6A-F). We ran the EBSP twice to observe the basic statistics of the runs, and both showed similar results (we report here the plots of the first replicate) (Table S7), we plotted the median value (black line) for all the EBS. The EBSP of all the present-day haplogroups together reconstructs a constant population size after the peopling of the Americas until 5,000 years ago, when a population expansion is observed. This expansion is followed by a population decline starting 2,000 years ago and becoming more pronounced 500 years ago (Fig. 2B). A similar trend is observed in the median value of the EBSP for haplogroups B and C (from modern + pre-Hispanic) but with higher variance in the 95% CI (Figs. S6C-S6D). While the EBSP from all the pre-Hispanic haplogroups together, haplogroup A (modern + pre-Hispanic), haplogroup D (modern + pre-Hispanic) and all haplogroups together (modern + pre-Hispanic) do not show any population decline in the median values observed when present-day mtDNA is used for the analysis (Fig. S6A, S6B, S6E and S6F). Similar demographic scenarios of population decline as in Fig.2B have been observed in other regions of the Americas colonized by Europeans in the 16th century, such as Puerto Rico and Peru (41, 43). However, we

9 found a larger population size and a less drastic decline in the effective female population size than
0 in these two regions (based on the 95% CI) (41, 43). This finding is consistent with a previous study
1 that reconstructed the N_e based on exome and whole-genome data from present-day populations
2 and found that in Mexico the N_e was higher than in Puerto Rico and Peru (44).

3
4 Determination of biological sex was made using the tool reported in (45). This approach computes
5 the proportion of reads mapped to the Y chromosome with respect to the reads mapping to the X
6 chromosome and Y chromosomes (R_y). According to the method, $R_y > 0.075$ corresponds to XY,
7 while $R_y < 0.016$ corresponds to XX. We could assign a sex for 16/39 individuals (females=3,
8 males= 13). For another four individuals, although the R_y values were above the required threshold
9 for XX assignment, they were very close (E2_Mich_b: 0.0218, E4_Mich_b: 0.022, E7_CdV_b:
0 0.0339 and E19_CdV_b: 0.0312). Sex assignment for these individuals was possible after applying
1 a filter with pmdtools (46) with thresholds of 0 – 3 (see Table S1), even though contamination
2 estimates based on the mitochondrial genome were $<3\%$. After pmdtools, all these four individuals
3 were assigned as XX (Table S1). Interestingly, we performed whole-genome in-solution capture
4 for these female individuals. According to a previous study, captured libraries tend to be enriched
5 in pseudo-autosomal and repetitive regions, due to an artifact caused by the use of DNA of male
6 individuals for preparing the RNA baits, which could lead to a bias in the estimate of the R_y value
7 (47).

8
9 Likewise, we analyzed pre-Hispanic male sub-lineages on the Y chromosome (Y-DNA). The
0 identification of Y-DNA haplogroups was only possible in five out of the thirteen individuals
1 genetically assigned as males, due to the low coverage (Table S1). All individuals presented the
2 Native American Q lineage as expected and in agreement with previous studies on ancient and
3 present-day Indigenous Mexican individuals (29, 34, 37, 48–50). Specifically, we identified the
4 sub-haplogroups Q1a2a1-L54 in the Sierra Tarahumara (n=1) and Ranas (n=1) and Q1a2a1a1a1-
5 M3 in the Sierra Tarahumara (n=1), Toluquilla (n=1) and CdV (n=1), with no apparent differences
6 between Aridoamerica and Mesoamerica (Table S1).

8 **Autosomal genetic structure in pre-Hispanic Mexico**

9 We performed ADMIXTURE(51) and Principal Components Analysis (PCA) (52, 53) to visualize
0 the genetic relationship and genetic structure at the autosomal level between the pre-Hispanic
1 individuals and other ancient and present-day Native American and continental populations (Table
2 S8) (14, 34, 50, 54, 55).

3 All pre-Hispanic individuals from Mexico, as well as all ancient individuals from California, Belize
4 and Patagonia clustered together in PCA with present-day Indigenous populations from Mexico
5 (Fig. S7) and share similar genetic composition between themselves as evidenced in ADMIXTURE
6 analysis (Fig. S8). The same analyses were repeated including only the pre-Hispanic individuals
7 and the present-day Indigenous populations from Mexico (14, 34, 50, 54) as a reference. When
8 projecting the pre-Hispanic individuals onto the principal component space, these individuals
9 clustered closely with present-day populations from the same geographical region, respectively. In
0 particular, the individual F9_ST_a from northern Mexico clusters closely with present-day
1 Rarámuri and Wixárika Indigenous populations (from north Mexico). Pericúes from Baja California
2 (14, 15) cluster closely with Wixárika, central Mexico Nahuas from Jalisco and Purépechas.
3 Regarding the individuals from central Mexico (SG, CdV, and Michoacán), these cluster closely
4 with present-day Indigenous populations from this same region (Nahua from Jalisco, Purépecha,
5 Totonac, Nahua from Puebla, and Nahua) (Fig. 3A). Notably, pre-, and post- mega-drought SG

6 individuals, cluster together with central Mexico present-day populations, contrary to what would
7 be expected if a population replacement from northern Mexico in the Sierra Gorda would have
8 taken place due to the climate change (Fig. 3A). For the ancient individual MOM6_ST_a from
9 northern Mexico, he clustered closely with present-day Indigenous populations from central-west
0 Mexico (Purépecha and Nahuas from Jalisco).

1 We contrasted the placement of individuals in PC space with that obtained when applying a
2 Temporal Factor Analysis (TFA) (56) to correct the ancestral relationships using the dates of the
3 pre-Hispanic individuals. For this, we included only four pre-Hispanic individuals with >1x
4 genome-wide coverage (F9_ST_a, 2417Q_TOL_b, 2417J_TOL_a, and 333B_TOL_a). This
5 analysis confirmed the closer genetic affinity of individual F9_ST_b with populations from
6 northern Mexico. Toluquilla (SG) individuals slightly shifted towards the central-east populations
7 in contrast with their position in the PCA plot that places them between central-west and central-
8 east populations. Furthermore, despite the different periods of the individuals from Toluquilla, they
9 keep clustering together as in the PCA (Fig.S9). To further support this structure, we applied a
0 novel approach named “missing data” PCA (mdPCA, see methods) on the 15 pre-Hispanic
1 individuals from Mexico. We observed a similar clustering of the ancient individuals as in the
2 standard PC projection, with the subtle difference that F9_ST_a clustered within present-day
3 Rarámuri and Wixárika populations, instead of in the vicinity as with the other two methods (Fig.
4 S10).

5 In agreement, ADMIXTURE analysis (K=6) reveals that pre-Hispanic individuals from northern
6 Mexico (Sierra Tarahumara and Pericúes) have a bigger proportion of ancestry shared with present-
7 day northern populations Akimel O’odham and Rarámuri (yellow component in Fig. 3B). Pre-
8 Hispanic individuals from central Mexico also display a similar genetic composition to present-day
9 central-west and central-east Mexico populations (Nahua from Jalisco, Purépecha, Totonac and,
0 Nahua from Puebla). Interestingly, individuals from SG show a homogeneous genetic composition
1 independently of the period in which they lived. The genetic composition of individuals from CdV
2 is similar to that in SG, but CdV individuals have a slightly higher proportion of the yellow
3 component shared with northern populations. In addition, we found the two individuals from
4 Michoacán have a higher proportion of a different genetic component shared with northern
5 populations (red component observed in Konkaak), E2_Mich_b presents the yellow component
6 while E4_Mich_b presents the red one.

7

8 We further explored the structure by computing outgroup-f3 and D-statistics. Outgroup-f3 statistics
9 were of the form $f3(\text{Test}, \text{Source1}; \text{YRI})$, where Test refers to the pre-Hispanic individual under
0 analysis, and Source1 being a present-day Indigenous population. D-statistics were of the form
1 $D(\text{Pop1}, \text{Pop2}; \text{Test}, \text{YRI})$, where Test refers to the pre-Hispanic individual under analysis, while
2 Pop1 and Pop2 were all possible combinations of present-day Indigenous populations. For both
3 analysis we considered only the combinations that have >1,000 overlapping Single Nucleotide
4 Variants (SNVs).

5

6 We found F9_ST_a to share higher genetic drift with present-day northern Mexico populations:
7 Rarámuri, Akimel O’odham and Konkaak (0.291,0.282,0.277), than with any other present-day
8 population (Fig. S11) (Table S9). Consistently, D-statistics showed F9_ST_a to be significantly
9 more related to the present-day Rarámuri than to any other population (Pop1) when tested in the
0 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)
1 (Table S10).

2 In contrast, the genetic ancestries of pre-Hispanic individuals from central Mexico are more
3 complex. They do not seem to have a higher shared genetic drift with any present-day Indigenous
4 population according to the f_3 outgroup values, where almost all standard errors overlap.
5 Furthermore, in D-statistics analysis, we find mostly $D=0$ in all possible combinations of Pop1 and
6 Pop2 (Fig. S13 – S18) (Table S9-S10). These results are consistent with populations from central
7 Mexico having extensive gene flow between them and not completely diverged from each other.

8 9 **Genetic diversity in pre-Hispanic Mexico**

0 The ancient genomic data we generated also enabled us to make inferences about past genetic
1 diversity and effective population sizes at the sites of study. This is relevant because it allows us to
2 get insights into the differences between the pre-Hispanic populations, contrast past and present-
3 day genetic diversity, and investigate changes in patterns of genetic variation that may have arisen
4 due to the drought period. To this end, we estimated the conditional nucleotide diversity (CND)
5 (57) and runs of homozygosity (ROH) (58), and contrasted these between the different pre-Hispanic
6 individuals and those observed in present-day Indigenous populations from Mexico (34, 50). CND,
7 a relative measure of genetic diversity, was estimated in pairs of individuals at genomic positions
8 that are variable in African individuals (57) and grouping individuals per archeological site. This
9 analysis included 16 pre-Hispanic individuals (Pericúes, Sierra Tarahumara, SG, CdV and
0 Patagonia) with at least 10,000 SNVs in common with the other individuals from the same
1 archaeological site to perform the analysis by pair of individuals (Table S11 and S12). We found
2 similar values of CND in pairs of pre-Hispanic individuals to those observed in present-day
3 Indigenous populations from Mexico, except for Pericúes from Baja California, who had the lowest
4 CND values (Fig. S19) (Table S12). This is in line with Pericúes being hunter gatherers, likely with
5 small population sizes and living in isolation in Baja California (59).

6 On the other hand, the ROHs reflect the proportion of the genome in a homozygous state, which is
7 inversely proportional to the heterozygosity, and can be the result of demographic scenarios that
8 lead to low population sizes and endogamy, thus it can be used to infer such processes (60–64). We
9 performed this analysis for six pre-Hispanic individuals with the highest coverage; namely two
0 from northern Mexico (the Pericú B03 and F9_ST_a from Sierra Tarahumara) and four from central
1 Mexico (three from Toluquilla and one from CdV) (Table S13). Of the three Toluquilla individuals
2 included in this analysis, one was dated pre-drought (2417Q_TOL_b), and two post-drought
3 (2417J_TOL_a and 333B_TOL_a). Pre-Hispanic individuals display fewer segments in ROH than
4 some present-day Indigenous individuals (especially present-day individuals from northern
5 Mexico). The segment size distribution of ROHs suggests that the pre-Hispanic individuals studied
6 here belonged to populations with small effective population (N_e) sizes ($2N_e= 1,600-6,400$) (Fig.
7 S20), which is in agreement with N_e previously calculated for present-day northern (Rarámuri, $N_e=$
8 2,419; Wixárika, $N_e= 2,193$) and southern (Triqui, $N_e= 2,407$; Maya, $N_e= 2,750$) populations (38).
9 We contrasted the CND values and ROH estimations of Pericúes from Mexico with another isolated
0 pre-Hispanic population from South America, Patagonia (1,275 – 895 BP) (24), expecting low
1 levels of genetic diversity in the latter too. We found that two pairs of Pericúes had similar CND
2 levels as Patagonia individuals ($\sim 0.16-0.175$), and the third pair of Pericúes presented a lower CND
3 value (~ 0.14). However, in ROH analysis, we found the four individuals from Patagonia have a
4 higher sum of fragments in ROH than the B03 Pericú individual, suggesting that the Pericú
5 population had a bigger population size than that of South American Patagonian individuals (Fig.
6 S20). In summary, the CND and ROH information showed that pre-Hispanic and present-day
7 individuals from northern Mexico (Pericúes and Akimel O’odham, respectively) present the lowest
8 CND values and the highest sum of inferred ROH >4 cM compared to other populations of their
9 respective period. Individuals from the Sierra Tarahumara, Toluquilla and Cañada de la Virgen
0 show similar values of genetic diversity and sum of inferred ROH segments. Furthermore, in the

1 case of Toluquilla, where we have three individuals from different dates, we found that the *CND*
2 value increase with the date difference between pairs of individuals, probably reflective of the
3 accumulation of new mutations or genetic drift through time during the 489-680 years of difference
4 between 2417Q_TOL_b and the individuals of a later period (2417J_TOL_a and 333B_TOL_a).

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

7 To formally test the hypothesis of population replacement in the SG from the NFM during the 900
8 – 1,300 CE droughts (7), we applied different combinations of outgroup-f3 and D-statistics using
9 individuals with more than 1000 SNVs overlapped. For the former, we used the individuals with
0 the highest depth from SG as the test populations in outgroup-f3 of the form f3(Test, Source1; YRI),
1 with individuals either representing before (TOL_b and R_b) or after (TOL_a) the climate change
2 episode. Source1 represents another pre-Hispanic individual, while YRI represents the outgroup
3 population. This allowed us to test if individuals from the SG shared higher genetic drift between
4 them than with any other pre-Hispanic individual from other archaeological sites, and if this
5 difference has a temporal significance.

6
7 We observed that the genetic affinities of the pre-Hispanic individuals from the SG before and after
8 the climate change episode share higher genetic drift between them than with any other pre-
9 Hispanic individual (Fig. 4A), except for the individual 11R_R_b, who shows higher genetic drift
0 with the individual E8_CdV_b from CdV (Fig. S21) (Table S14). However, standard errors for
1 11R_R_b are high, probably due to a low coverage genome-wide for this individual (Fig. S21).
2 This is in contrast with the outgroup-f3 values obtained when pre-drought CdV individuals were
3 used as the Test population f3(CdV, Source1; YRI) (Fig. S22). Here, the highest outgroup-f3 value
4 systematically corresponds to another CdV individual but corresponds to the second-grade relative
5 according to the 'READ' analysis (see methods). The following closest individuals do not belong to
6 CdV. In the case of the two individuals from Sierra Tarahumara, the higher value of genetic drift is
7 between both (Fig. S23).

8 D-statistics were calculated in the form (Pop1, Pop2; Test, YRI), where Test refers to the pre-
9 Hispanic individual under analysis, Pop1 representing any pre-Hispanic individual, and Pop2
0 representing a pre-Hispanic individual from the same region (Sierra Tarahumara, SG or CdV) as
1 Test (Fig. S24 – S26) (Table S15).

2 Individuals from SG are generally more closely related between them ($D < 0$) in D tests of the form
3 $D(\text{Pop1, SG1; SG2, YRI})$ (Fig. 4B), except when analyzing 11R_R_b (Fig. S25). We do not find a
4 higher relationship between the individual 11R_R_b from Ranas and the other individuals from the
5 Sierra Gorda (Toluquilla). We caution that error bars in $D(\text{Pop1, SG; 11R_R_b, YRI})$ are wide (Fig.
6 S25). In most of the cases, D statistics in form $D(\text{Pop1, SG; 2417Q_TOL_b, YRI})$ show a
7 significantly higher relationship ($|Z| > 3$) between individuals from Toluquilla than with any other
8 North America or South America individual. Indeed, the two individuals who lived after the
9 megadrought (2417J_TOL_a and 333B_TOL_a) are significantly more closely related than to any
0 other pre-Hispanic individual, except when compared with the individuals MOM6_ST_a (from the
1 north), 11R_R_b (from SG), E4_CdV_b (from CdV) and E4_Mich_b (from Michoacán) (Fig. 4B),
2 for which the D-statistic is not significant probably due to the low-coverage in these genomes which
3 increases the standard error. In a scenario of a population replacement, we would expect the
4 Toluquillan individuals who lived after the drought to be more closely related to pre-Hispanic
5 individuals with higher proportions of northern ancestry (Pericúes and F9_ST_a) than to
6 Toluquillan individuals who lived before the drought (2417Q_TOL_b), since the alleged population
7 replacement occurred from the north. Therefore, our results suggest genetic continuity in Toluquilla
8 before and after the reported climate change event. In some cases, D-statistic values were not

9 significant when another pre-Hispanic central Mexico individual was used as Pop1. These results
0 could be related to the low coverage mentioned for some pre-Hispanic individuals, or they might
1 speak about the previously reported heterogeneity in central Mexico (38, 39).

2 We then used f-statistic-based admixture graphs (qpGraph) (65) to characterize the ancestry
3 proportions of the pre-Hispanic individuals of SG (Fig. 5). For the model, we used the 11,500-kyo
4 USR1 (21) as the surrogate for ancient Beringian, 10,700-kyo Anzick (22) individual as the
5 surrogate for ancestral Southern Native American ‘SNA’(14) and the 4,200-kyo Ancient
6 Southwestern Ontario or ASO individual (15) as the surrogate for Northern Native American
7 ‘NNA’(14). In the analysis, we included the pre-Hispanic individuals 2417Q_TOL_b,
8 2417J_TOL_a, and 333B_TOL_a and present-day Indigenous populations in Mexico (54)
9 representing the northern (Konkaak), central (Purépecha or Nahuatl from Puebla), and southern
0 (Triqui) regions. The results show that the oldest individual 2417Q_TOL_b seems to have a higher
1 contribution of the ancestral B branch than the individuals after the megadrought (2417J_TOL_a
2 and 333B_TOL_a) (Fig. 5). However, overall results (PCA, TFA, mdPCA, Admixture, f3-outgroup
3 and D-statistics) point to a genetic continuity for the Toluquillan individuals.

5 **Admixture model for Central Mexico populations**

6 Given the complex relationships observed for ancient individuals in central Mexico, we tested
7 additional demographic models using qpGraph. The base model that fitted the pre-Hispanic
8 populations includes Anzick (22) individual as the ancestral SNA (12), present-day Athabascan
9 (14) as the representative for NNA (14), and present-day Indigenous populations in Mexico (44)
0 representing the northern (Konkaak), central (Nahuatl from Puebla), and southern (Triqui) regions.
1 Furthermore, this model involves a split of the Southern Native American into S2A and S2B
2 sources. The pre-Hispanic populations from SG and CdV show different levels of ancestry from
3 Southern Native American and Northern Native American. We found that the population from SG
4 had a higher percentage of the Southern Native American branch from S2B (64%) than the
5 population from CdV (37%) who had 62% ancestry shared with the Northern Native American
6 branch, indicating a different demographic history for each of these populations despite its relative
7 geographic proximity (Fig. S27). This model supports previous studies that report several admixture
8 events between the two branches that have given rise to Central and South American populations
9 (15). For the Michoacán population, we obtained terminal branches with no support (zeros) and,
0 because of that, we did not make conclusions with this model for the Michoacán pre-Hispanic
1 population (Fig. S27).

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

4 We decided to evaluate the possible scenario of population continuity in the Sierra Tarahumara
5 from northern Mexico, considering that the individual F9_ST_a was significantly more related to
6 present-day Rarámuri than to any other indigenous population from Mexico in f3 and D statistics.
7 Thus, we used the population continuity test described in (66), in such a way that the null hypothesis
8 of population continuity is tested using read counts in the ancient individual and allele frequencies
9 in the reference population (Rarámuri). Using this method, we found no population continuity
0 between F9_ST_a and present-day Rarámuri and rejected the null hypothesis with a p-value of $10^{-499.3}$.

5 **Ghost population contribution to pre-Hispanic Mexico**

6 Earlier reports have demonstrated the detection of ancestry from an unsampled group, designated
7 UpopA, among the present-day Mixe from Mexico and Lagoa Santa from Brazil populations (67).
8 We tested the presence of UpopA in the pre-Hispanic individuals, using an admixture graph model
9 with the same topological framework but some modifications compared to (67). We used Mbuti as
0 an outgroup, Ami as the East Asian, MA1 as the ancient north Eurasian, USR1 as the ancient
1 Beringian, Anzick-1 (22) and Spirit Cave (67) as the NNA, and Athabaskan as the SNA.
2 Interestingly, the Mummy F9_ST_a from the north and the individuals from CdV from central
3 Mexico have ancestry from a ‘ghost’ population, at 28% and 17%, respectively ($|Z\text{-score}|$: 2.142
4 and 1.655, respectively) (Fig. 6), consistent with the “ghost” UpopA ancestry reported in Mixe. We
5 repeated this test with the Mixe population to verify the contribution of the ghost population in our
6 tested admixture graph structures, and found Mixe to have an ancestry coming from an unsampled
7 UpopA population at 13% ($|Z\text{-score}|$: 2.34) (Fig. S28), which is similar to previous estimates of
8 11% (67). Models including individuals from other archaeological sites had $|Z\text{-score}|$ values over 3
9 and branches with inner zeros, thus we rejected them.

1 **Discussion**

2 We have generated the largest available ancient genome data set from Mexico to address long-
3 standing questions about population dynamics between Aridoamerica and Mesoamerica. The data
4 allowed us to i) describe levels of population structure and genetic diversity in Mexico prior to
5 European colonization, ii) test a previous hypothesis of population replacement in central Mexico
6 following a drastic climate change 1,100-700 years ago, iii) demonstrate a complex admixture
7 model for central Mexico populations, and iv) detect a contribution from unsampled population A
8 (UpopA) to some northern and central Mexico populations.

9 We find that present-day Indigenous populations from Mexico exhibit substantial geographical
0 structure following a northwest-southeast cline (54). The pre-Hispanic individuals from this study
1 also follow this structure. They cluster in proximity to present-day Indigenous populations from the
2 same geographical area based on PCA, TFA, and mdPCA, and exhibit similar ancestral components
3 based on ADMIXTURE analysis, except for the individual MOM6_ST_a (Fig. 3 and S9-S10). This
4 reflects an overall conservation of the genetic structure of the populations inhabiting the Mexican
5 territory (54) for at least 1,400 years (considering that the most ancient individual used for
6 autosomal genomic analysis is 1,400 years old), which is consistent with demographic models
7 inferred from exome (38) and genome-wide genotype (68) data from present-day Indigenous
8 populations from Mexico. These models propose a northern/southern population split between
9 4,000 and 10,000 years ago followed by multiple waves of admixture events between them (68).
0 The exome data further supports subsequent splits within each of these regions 6,523 and 5,715
1 years ago for north and south, respectively (38).

2 This geographical structure is also reflected in the maternal lineages. Even though we identified
3 previously unreported variants in the ancient mitogenomes (Table S5), the spatial distribution of
4 the haplogroups found, namely A, B, C and D, closely resembles that of present-day Mexico (27,
5 36). The information on the paternal lineage was limited, since, despite confirming the presence of
6 the Native American chromosomal haplogroup Q in the pre-Hispanic males, there was no clear
7 differential spatial distribution of the two identified sub-haplogroups, Q1a2a1-L54 and
8 Q1a2a1a1a1-M3, in both Aridoamerica and Mesoamerica. Few studies have recovered information
9 of ancient Y chromosome haplogroups across the Americas and, given the sex-biased admixture
0 events after European colonization, the frequency of haplogroup Q in the present-day population is
1 low. Thus, further studies on more ancient individuals from the Americas at higher coverage will
2 likely shed light on the demography of the male lineages.

3

4 We delved further into the past demography of the Indigenous population by performing a Bayesian
5 coalescent analysis on the mitogenome data. The resulting Extended Bayesian skyline plot (EBSP)
6 showed a larger female population size and a less drastic population decline upon European
7 colonization in Mexico compared to other regions in the Americas (based on the 95% CI) (41, 69).
8 This is intriguing considering previous reports of massive population decline during European
9 conquest (70, 71). However, these estimates were based primarily on historical accounts of the early
0 colonial Mexican population through taxation lists of most settlements across Mexico but not
1 including all of them (71). The mitogenomes used for the EBSP belong to individuals across the
2 Mexican territory ranging from the north to central Mexico. Thus, we suggest that the demographic
3 impact of European colonization on the female population was variable across Mexico, which is
4 generally less drastic than in insular and isolated regions of the Americas. EBSP, including only
5 modern mitochondrial genomes, shows the female population bottleneck experienced after
6 European contact (Fig. 2B). In contrast the EBSP that includes only pre-Hispanic mitochondrial
7 genomes does not show the female population decline (Fig. S6A).

8

9 The characterization of past genetic diversity revealed that ancient Aridoamerica (Sierra
0 Tarahumara) and NFM populations (Toluquilla and CdV) have similar levels of conditional
1 nucleotide diversity (CND) between them and to those observed in some present-day Indigenous
2 populations from northern (Akimel O’odham) and southern Mexico (Mixe and some Mayan) (Fig.
3 S19). However, we caution that we lack genome-wide genetic information of present-day
4 individuals from the same site as the pre-Hispanic individuals analyzed here to make a direct
5 temporal comparison. We observed, however, that the lowest CND and the highest total number of
6 segments in ROH were found in the ancient hunter gatherer Pericúes, and the present-day Akimel
7 O’odham population, both from northern Mexico. In contrast, pre-Hispanic individuals from the
8 NFM and present-day populations from central, southern, and southeastern Mexico have the lowest
9 total number of segments in ROH (Fig. S20). Altogether, these results are reflective of the lower
0 population sizes maintained in isolated Aridoamerica populations and larger population sizes in
1 pre-Hispanic and present-day populations in Mesoamerican (central to southeastern Mexico).
2 Nonetheless, it is important to note that the pre-Hispanic individuals from this study come from
3 small-sized villages (72, 73) and do not necessarily reflect the demography of other pre-Hispanic
4 populations from larger and multiethnic pre-Hispanic metropolises (e.g., Teotihuacán, Tenochtitlán,
5 or Palenque) (74, 75) where higher levels of diversity and lower ROH would be expected. Ancient
6 DNA studies on individuals from these sites could reveal more accurately the extent of genetic
7 diversity in these megalopolises prior to the population collapse resulting from European
8 colonization.

9

0 To directly test the hypothesis of a population replacement of Mesoamerican populations at the
1 NFM by Aridoamerican hunter gatherers because of a drastic climate change in the years 900-1,300
2 CE, we studied individuals from the Sierra Gorda (Toluquilla and Ranas sites), for which we had
3 access to remains from the time before (671-867 CE) and after ($1,160 \pm 60$ and $1,351 \pm 50$ CE) the
4 resulting megadrought. Using D statistics, we observed that the pre-drought individual
5 2417Q_TOL_b and the post-drought individuals 2417J_TOL_a and 333B_TOL_a, all have the
6 same genetic relationships to present-day Indigenous populations. In other words, 2417J_TOL_a
7 and 333B_TOL_a are not significantly more related to northern (Aridoamerica) populations than
8 2417Q_TOL_b, as it would be expected under a scenario of population replacement in the NFM.
9 This population continuity is further supported by the longstanding presence of the mitochondrial
0 lineage A2d in the region for at least 1,480 years (Fig. 2A). Notably, we found shorter ROH

1 segments in the pre-drought pre-Hispanic individual 2417Q_TOL_b than in the post-drought
2 individuals 2417J_TOL_a and 333B_TOL_a, which might indicate a reduction in the population
3 size post-climate change period. Altogether, these results suggest population continuity in the Sierra
4 Gorda of the NFM, before and after the climate change episode. Additional assessment of this
5 hypothesis at other NFM sites will shed light into the population migrations in other regions of the
6 NFM.

7
8 The observed population continuity in the Sierra Gorda might have been supported by the favorable
9 climatic conditions of the northern Sierra Gorda Mountain range, where both Toluquilla and Ranas
0 sites are located. The landscape could have helped maintain higher humidity conditions than in
1 other regions of the NFM. In this part of the Sierra, the humid winds from the Gulf of Mexico
2 course trough. The area is characterized by coniferous and mixed forests, where the average
3 maximum temperature is 25°C, and the minimum temperature is 6-7°C (76). This landscape is very
4 different from the conditions found in Cañada de la Virgen, Guanajuato, where the climate is
5 considerably more arid. CdV is located north of the Lerma River, where rainfall is known to be
6 scarce. The inhabitants of that region had to develop systems for the capture and maintenance of
7 water needed for cultivation, which probably became impossible after the prolonged droughts. This
8 forced them to abandon the site during 1,000-1,100 CE and migrate to areas with better climatic
9 conditions suitable for agriculture, which was the core of their primary subsistence activity (73). In
0 contrast, the main subsistence strategy in Toluquilla and Ranas was mining. Residents at these sites
1 exploited cinnabar for trade purposes with other villages and metropolises (72, 77). Cinnabar was
2 of great sacred value to pre-Hispanic cultures. Some ancient communities used it to anoint their
3 dead, especially in cases of highly regarded individuals, and as part of ritual offerings (77, 78). We
4 hypothesize that the cinnabar trade and the landscape of the Sierra Gorda allowed the peoples of
5 Toluquilla and Ranas to subsist despite low rainfall conditions during the drought.

6
7 On the other hand, we were also interested in exploring putative scenarios of population continuity
8 between pre-Hispanic and present-day populations. The only pre-Hispanic individual found to be
9 significantly more related to a present-day population was the individual F9_ST_a, who was
0 significantly more related to present-day Rarámuri than to any other present-day population from
1 Mexico. This relationship was expected since F9_ST_a was found in the Sierra Tarahumara, where
2 present-day Rarámuri currently reside. The northern region of Mexico was one of the most difficult
3 to access and the last areas to be invaded by the Europeans in the seventeenth century, due to its
4 geography and the resistance movements by the local people. Nowadays, Rarámuri still preserve
5 their pre-Hispanic cultural elements in language, music, dresses, and handcraft (79). However, the
6 genetic continuity between F9_ST_a and the Rarámuri was rejected despite the high genetic
7 affinities between them. We hypothesize that this genetic population discontinuity observed from
8 this analysis can result from the changes in allele frequencies derived from the admixture events,
9 population bottleneck, and isolation experienced by Rarámuri since pre-Hispanic times and after
0 the European colonization (68).

1
2 We were also interested in obtaining insights into the demography of central Mexico populations
3 since previous attempts of demographic modeling have been hampered by the high genetic
4 heterogeneity observed in these populations (32, 33). Our outgroup-f3 and D statistics results show
5 that, although pre-Hispanic individuals from central Mexico share higher genetic drift with present-
6 day populations from that region, none of these relationships were statistically significant. We thus
7 leveraged the ancient genome data to propose an admixture model for central Mexico populations.
8 Using qpGraph we found that the pre-Hispanic populations from central Mexico, all have different
9 ancestry sharing with the NNA and SNA branches. This was expected given previous studies on

0 ancient genomes from Central and South America reporting multiple admixture events between
1 these two branches after their split ca. 15,000 years ago (15, 68). Together, these observations point
2 to a scenario in which populations from central Mexico have not completely diverged from one
3 another, possibly due to extensive gene flow as expected based on the active commercial exchange
4 between different Mesoamerican populations for centuries (80). This interaction was mainly
5 through trade routes and alliances between different nations (55), as has been evidenced in present-
6 day Indigenous populations who share identity-by-descent (IBD) segments (43, 54) and through
7 the analysis of STR loci (56). In addition, there was also gene flow between Mesoamerican and
8 Aridoamerican populations in pre-Hispanic times, as evidenced by the study of IBD segments in
9 present-day Indigenous populations from Mexico (68). However, this gene flow occurred less
0 frequently than between Mesoamerican populations. The individual MOM6_ST_a from northern
1 Mexico may have been the result of such admixture between Mesoamerican and Aridoamerican
2 ancestors. This individual has a higher shared genetic drift with F9_ST_a (also from Aridoamerica),
3 as evidenced by outgroup-f3 tests. However, PCA and ADMIXTURE analyses show that
4 MOM6_ST_a has ancestry components found in central west present-day populations. Also, using
5 D statistics, we found MOM6_ST_a to be equally related to all individuals from Aridoamerica and
6 Mesoamerica, supporting the idea of gene flow between these two areas.

7

8 Furthermore, the qpGraph admixture models explored for the pre-Hispanic populations showed
9 F9_ST_a and the ancient individuals from CdV have ancestry from an unsampled population that
0 we hypothesize is the ghost population previously found named UpopA (67). The contribution from
1 UpopA was identified in present-day Mixe from Mexico and Lagoa Santa, Brazil, through qpGraph
2 models and estimated to have diverged ~24,700 years ago from Native Americans (67). Also, a
3 more recent study through D statistics analysis has suggested that UpopA contributed to present-
4 day populations from north Mexico as Akimel O’odham, Guarijo, Rarámuri, Cora, Mexicanero,
5 and Tepehuano, as well as in present-day populations from central and south Mexico as Mixe,
6 Totonac, Nahuas from Puebla, Otomi from Hidalgo, Chocholteco and, Mocho (68). Our study is
7 the first to identify this UpopA ancestry in admixture models involving pre-Hispanic individuals
8 from Mexico. Further aDNA studies from the Americas could help better characterize the source
9 of this UpopA ghost population contributing to many present-day Indigenous populations from
0 Mexico.

1

2 In summary, we find that the pre-Hispanic population structure from over a thousand years ago,
3 can still be observed today. Our work, together with previous studies (15, 38, 39, 54, 68), show that
4 the demographic events that gave rise to Aridoamerican and Mesoamerican populations are more
5 complex than previously thought. Commercial trade routes may have contributed to increased
6 mobility, facilitating gene flow between different populations within and between various cultural
7 areas. Furthermore, we found genetic continuity in the Sierra Gorda region at the NFM which
8 suggest the local population stayed in their homeland despite the longstanding droughts that forced
9 other NFM populations to abandon their cities. This study opens the door for further research to
0 address the questions of the unknown genetic past and population dynamics of Mexican pre-
1 Hispanic populations, whose genetic legacy is retained today among Indigenous and admixed
2 populations.

3

4 **Materials and Methods**

5 Laboratory procedures

6 DNA extraction and library preparation (before amplification) were performed in the Human
7 Paleogenomics Laboratory, a clean lab facility at the International Laboratory for Human Genome

8 Research from the Universidad Nacional Autónoma de México (LIIGH-UNAM). Bones and teeth
9 surfaces were cleaned with a 1% sodium hypochlorite solution, followed by a solution of 75%
0 ethanol. Then, the surface was UV irradiated (256 nm) for 1.5 minutes on each side using a UVP
1 CL-1000 crosslinker. A Dremel tool was used to remove the outer surfaces of bones and teeth.
2 Bones were cut to get an inner sample of around 100-200 mg for DNA extraction. Teeth were cut
3 at the cemento-enamel junction, the roots were wrapped in aluminum foil and pulverized using a
4 hammer. Around 100-200 mg of the root was used for DNA extraction. DNA extraction was
5 performed using the methods described in (81, 82), as described in Table S2.

6

7 DNA extraction from individual P_CCM_b was performed taking 150 mg of mummified skin. The
8 sample was washed in deionized sterile water and dried. Then, the sample was UV irradiated (256
9 nm) for 1.5 minutes using a UVP CL-1000 crosslinker. Epithelium was cut in small pieces and
0 incubated at 50°C for 24 hours in lysis buffer (10mM Tris-HCl, 10mM NaCl, 5mM CaCl₂, 2.5 mM
1 EDTA, 1% SDS, 10 mg/ml proteinase K, 10 mg/mL DTT). Followed by centrifugation at 16,100 x
2 g for 5 minutes and recovery of the supernatant. Then, DNA purification was performed according
3 to the protocol in (82).

4

5 Double-stranded libraries for Illumina sequencing were prepared according to (83), using single-
6 indexed adapters with 6-bp barcodes or double-indexed with 7-bp barcodes, depending on the
7 sequencing platform, NextSeq550 or NovaSeq, respectively. Libraries were analyzed with qPCR
8 to determine the optimum number of cycles during the indexing PCR step. Barcoded libraries were
9 sequenced for a first screening on the NextSeq550 equipment from Illumina using a 2x75 run at
0 either LANGEBIO's genomics core facility (National Laboratory of Genomics for Biodiversity,
1 Irapuato, Guanajuato.) or INMEGEN genomics facilities (National Institute of Genomic Medicine,
2 Mexico City). Sequence data were processed as described in the next section. Depending on the
3 quality of the libraries (% endogenous and % clonality), as revealed by the screening sequence data,
4 some were chosen to be subjected to mitochondrial genome capture or whole-genome capture using
5 Daicel Arbor Biosciences (Ann Arbor, MI, USA) commercial kits (Table S2). Captured libraries
6 were sequenced to assess complexity and yield with the tool preseq (84) and sequenced to higher
7 depth using the NextSeq 550 (2x75 cycles) at LANGEBIO's core facility or in the NovaSeq at the
8 Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT) with an S1, 2x100 cycles run
9 (Table S2). Sequencing runs not reported previously for Sierra Tarahumara individuals were carried
0 out at the Sci LifeLab Uppsala Sequencing Center and at the Bustamante Lab in the Department of
1 Genetics at Stanford University (Table S2).

2

3 Sequence data processing

4 Raw reads were processed using Adapterremoval (85) for trimming Illumina adapter sequences and
5 collapsing of pairs of reads with an overlap of at least 11 bp (--trimns --trimqualities --qualitybase
6 33 --minlength 30 --collapse). Collapsed reads with >30bp and quality above 33 were retained for
7 downstream processing.

8

9 The retained reads were mapped to the human reference genome b37 (hg19), with the mitochondrial
0 sequence replaced by the revised Cambridge reference sequence (rCRS, NC_012920). Mapping
1 was done with bwa 0.7.13, aln algorithm, seed disabled (-l 500), and keeping reads with mapping
2 quality >25. Clonal duplicates and reads mapping to more than one place in the genome were
3 removed with SAMtools rmdup (86) and eliminating reads with labels 'XT:A:R' and 'XA:Z',
4 respectively. Then, reads were realigned using GATK with RealignerTargetCreator and
5 IndelRealigner (87). To authenticate ancient DNA sequence data, mapDamage2 (18) was used to

6 assess the damage patterns and distribution of reads length, using default parameters (Fig S1-S4).
7 Alignments to rCRS were separated into new bam files using SAMtools (86) to analyze the
8 mitochondrial genome for haplogroup assignment, haplotype network analysis and Extended
9 Bayesian Skyline Plot (EBSP) (see respective sections below for details). For the analysis of
0 autosomal variants, base qualities were rescaled with mapDamage2 (18), this step reduces the base
1 quality of the bases in the reads that have a misincorporation from C to T and G to A in the
2 alignment with the reference genome, since this changes might be present due to ancient DNA
3 damage. Depth of coverage was estimated using ATLAS (88).

4 mtDNA analysis

7 Reads aligned to the Cambridge reference sequence (rCRS, NC_012920) were analyzed with
8 Schmutzi (89) to estimate contamination and obtain a consensus mitochondrial sequence for each
9 individual. Script contDeam.pl from Schmutzi was run with the parameters --library double and the
0 default --uselength, and schmutzi.pl was run with 8 threads. Haplogrep 2 (90) was used to assign a
1 mitochondrial haplogroup and quality of assignment for each library using the fasta file with the
2 consensus mitochondrial genome generated by Schmutzi (Table S1).

4 mtDNA alignments. Consensus fasta sequences of the whole mtDNA of ancient individuals from
5 this and previous studies were aligned together with previously published modern sequences (14,
6 15, 32, 34, 40, 41, 91–100) (Table S4 and S6) using MEGA X (101). Multiple sequence alignments
7 of mitochondrial genomes of the same haplogroup (A, B, C or D) were further visually inspected
8 (i.e. removing common indels, as well as to confirm variants and making sure sequences remain at
9 the same length), to assure an accurate alignment. The same process was also done for a sequence
0 alignment with the combination of all four haplogroups. Common hypervariable sites that are
1 phylogenetically uninformative were excluded from the analysis (309, 315, 515-522 AC indels,
2 3107, 16182-16183, 16193, and 16519) (40). Both sequence alignments for haplotype networks
3 and Extended Bayesian Skyline plots were done following this description, although they contain
4 different samples.

7 Haplotype networks. Haplotype median-joining networks were constructed in Popart (102) after
8 collapsing aligned full mtDNA sequences into haplotypes with DnaSP . Haplotype networks were
9 constructed with both ancient (from this study and previously published) (n=25) and present-day
0 (n=78) mtDNA of individuals in the Mexican territory (Table S4 and S6). All present-day samples
1 were retrieved from the PhyloTree (version 17) database (40, 91). Table S4 reports the ancient
2 individuals for which the mtDNA sequences were used. Only sequences that were reported as being
3 of the same sub haplogroup or neighboring sub-haplogroup as our ancient samples were chosen for
4 sequence comparison.

6 Extended Bayesian skyline. Past female effective population sizes were reconstructed using a
7 Bayesian skyline approach (104) in BEAST 2 (105). The ancient mtDNA sequences, mentioned
8 above were merged with present-day mtDNA sequences collected from published public databases
9 (Table S4 and S6). Alignments were partitioned into five concatenated regions in the following
0 order (Dloop, Coding, rRNA, and tRNA) as proposed by (41). HKY+G was found to be the best
1 substitution model for this arrangement of the data using PartitionFinder 2 software (106).
2 Substitution rates were estimated following a strict molecular clock model starting from point
3 estimations as in (107). Additionally, we used tip calibrations for ancient samples using dating
4 estimates (Supplementary Table S1). We ran Markov Chain Monte Carlo (MCMC) chains of 100
5 million steps for each haplogroup alignment with a sampling of parameters every 10,000

6 generations, discarding the first 10 million steps as burn-in. Extended Bayesian skyline analysis
7 was plotted using SkyViz reported in (42). Two independent runs of EBSP were run, showing a
8 similar behavior (see Table S7 for statistics of each run). The first run of each EBSP was used to
9 be reported.

1 Sex assignment

2 Determination of biological sex was made using the tool reported in (45). This approach computes
3 the proportion of reads mapped to the Y chromosome with respect to the reads mapping to the X
4 chromosome (R_y) and Y chromosomes. According to the method, $R_y > 0.075$ corresponds to XY,
5 while $R_y < 0.016$ corresponds to XX (Table S1).

7 Y-chromosome haplogroup inference

8 Y chromosome genotype calling and haplogroup assignment were made as reported in (108).
9 Genotype calling was performed sampling one random base at each site of the Y chromosome
0 covered at least once, with ANGSD (-dohaplocall 1 -doCounts 1 -r chrY: -minMinor 0 -maxMis
1 4). Then, haplogroup assignments were made using the phylogenetic tree of Y chromosome single
2 nucleotide polymorphisms constructed from the 1000 Genomes Project Phase 3, as in (108). The
3 most derived haplogroup was assigned.

5 Autosomal reference panels

6 For analysis at a continental level, we constructed a reference panel (Supplementary Table S5) that
7 includes genetic information of: Yoruba (YRI), European Ancestry (CEU), and Chinese (CHB)
8 from the 1000 Genomes Project Phase 3 (55, 109), and available genome-wide information for
9 Indigenous populations from Mexico previously published by the Human Genome Diversity
0 Project (HGDP) (50, 110). These genome-wide data were intersected with previously published
1 reference panels of genotype information of Indigenous populations from Mexico (54), after
2 masking non-Native American sites as in (111) and keeping individuals with $> 85\%$ genotype
3 information (mind 0.15) and SNVs with a minor allele frequency of 0.01 (maf 0.01), and with a
4 90% genotyping rate (geno 0.1) using the software plink (112). The final reference panel includes
5 genotype information of 576,409 SNVs from 596 individuals from 4 continental populations (YRI,
6 CEU, CHB, and Indigenous from Mexico) (Table S8).

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

3 Both reference panels (continental and within Mexico) were pseudohaplotyped as described in the
4 next section.

6 Autosomal genotype calling and population genomic analysis

8 Pseudo-haploid calls of the ancient genomic data were made for the positions included in the
9 reference panel. This was done by randomly sampling one read (when more than one read covered
0 the site) at each site present in the reference panel and keeping the observed base at that site if it
1 had a minimum base quality of 30. If only one read covered the site, the observed base was only
2 kept if it had a base quality above 30. Then, the calls were turned into homozygous genotypes.
3 Similarly, for the genotype data in the reference panel, one allele was randomly sampled at
4 heterozygous sites and turned into a homozygous genotype. For downstream analyses, we kept only
5 ancient individuals with $>1,000$ SNVs intersected with the panel.

6

7 We carried out the analysis of autosomal genomes of 13 pre-Hispanic individuals with the highest
8 depth of coverage per archaeological site: Sierra Tarahumara (n=2, 0.01-1.8x), Sierra Gorda (n=5,
9 0.005-4.7x), Cañada de la Virgen (n=4, 0.03-0.5x), and Michoacán (n=2, 0.04-0.05x) (Table S1).
0 As a first approach, ancient genomes were compared with present-day Indigenous populations from
1 Mexico and three continental populations (YRI, CEU, and CHB), using a set of 561,3279 SNVs.

2

3 Relatedness. Relatedness between ancient individuals from the same site was assessed using READ
4 (113). A normalization step with a panel of non-related individuals is required previous to
5 estimating the relationship in the ancient individuals, this step was performed with all Nahuas from
6 the reference panel as they are the biggest group in the reference panel. Then, the normalization
7 value was used to run READ with the ancient individuals. Two individuals from Toluquilla
8 (333C_TOL_a and 333B_TOL_a) were identified as first-degree relatives. Thus, the one with the
9 highest coverage (333B_TOL_a) was used for downstream analyses. Furthermore, individuals
0 E4_CdV_b and E8_CdV_b were identified as second-degree relatives, as well as E7_CdV_b and
1 E19_CdV_b. All of these individuals from CdV were kept in the downstream analyses.

2

3 ADMIXTURE analysis. A first ADMIXTURE (51) analysis was carried out on the genotype
4 reference panel including Indigenous individuals from Mexico and the three continental
5 populations (YRI, CEU, and CHB) to identify present-day Indigenous individuals with <90%
6 Native American ancestry. One hundred replicates were run for each k, from k=2 to k=9 with the
7 cross-validation error (-cv) parameter. We found k=4 to be the k with the lowest mean cv error,
8 separating YRI, CEU, CHB, and Native American ancestry. Then the run of k=4 with the best
9 likelihood was chosen and the Indigenous individuals with <90% Native American ancestry were
0 removed from the panel. A second ADMIXTURE analysis was carried out with only the present-
1 day Indigenous populations with >90% Native American ancestry, from k=2 to k=8 using cv error
2 parameter to identify the best value of k, being k=2 the one with the lowest cv error, and separating
3 populations with northern and southern ancestries. Then, a third ADMIXTURE run was carried out
4 including the present-day Indigenous populations and pre-Hispanic individuals. In all cases, one
5 hundred replicates were run with randomly generated seed values; the run with the best likelihood
6 for each k was plotted using pong software (114).

7

8 Principal Components Analysis. We performed principal component analyses (PCA) using
9 smartpca from the software eigensoft v6.0.1 (52) and projecting the individuals (lsqproject: YES)
0 on the PCs estimated for the modern populations. The PCA (Fig. 3A) with only the ancient
1 individuals as well as the Indigenous populations from Mexico, only included present-day
2 individuals masked and with >90% Native American ancestry.

3

4 Temporal Factor Analysis. Temporal Factor Analysis (TFA) (56) were run with the reference panel
5 that includes only the Indigenous populations from Mexico. The ancient samples included in the
6 TFA consisted exclusively of individuals from Mexican territory from this study with >1x genome-
7 wide coverage: one individual from Sierra Tarahumara (F9_ST_b), and three from Toluquilla
8 (2417Q_TOL_b, 2417J_TOL_a, and 333B_TOL_a). Genotyping data was manipulated with Plink
9 version 1.90 beta (112). The dataset was converted from Plink format to Eigenstrat's geno format
0 with the command convertf from Eigensoft version 6.0.1.1 (52). Missing genotypes were imputed
1 with the snmf function from the LEA package, using the parameters K = 2, entropy = TRUE, and
2 repetitions = 5. The runs with the lowest cross-entropy value were considered for posterior steps.
3 Genotypes were corrected according to the autosomal coverage depth of each sample. We assigned
4 the average depth from the HGDP dataset mentioned in the publication: 35X (50). We assigned
5 the respective coverage read depth for all 11 individuals from SGDP. Modern samples genotyped

with microarrays (54) did not have a coverage depth value as they represent a different technology without sequencing reads, therefore, we assigned the maximum value from the dataset to all microarray samples, i.e. 56.19X. TFA was applied to these imputed and corrected genotypes with the following parameters: $\lambda = 5e-1$, and $K = 2$. Imputation, coverage correction, TFA, and the manipulation of the genotyping matrix in geno format were performed with R version 4.0.2. Variance explained by λ did not change considerably across logarithmic values, thus we chose the default λ value $5e-1$. We plotted factor 1 and 2 as negative when required to match geography and facilitate the comparison between TFA plots.

mdPCA. Missing DNA PCA (mdPCA) is a principal component analysis that corrects for genotypes missing due to ancestry masking or degradation in ancient DNA samples. The correction is performed by comparing genetic distances between all samples. The dataset included 15 ancient samples and 231 modern samples. Modern indigenous populations included Northern Mexico (Pima, Rarámuri, and Wixárika), Central-West Mexico (Purépecha, and Nahuatl from Jalisco), Central-East Mexico (Nahuatl, and Totonac), South Mexico (Mazatec, Zapotec, and Mixe), and Southeast Mexico (Maya, Tzotzil, and Tojolabal). The ancient samples included in the mdPCA consisted exclusively of individuals from Mexican territory: two individuals from Sierra Tarahumara (La Ventana: F9_ST_a, and MOM6_ST_a), three Pericú individuals (Pericu_B03, Pericu_BC25, and Pericu_BC30), four individuals from Sierra Gorda (Ranas: 11R_R_b; Toluquilla: 333B_TOL_a, 2417J_TOL_a, and 2417Q_TOL_b), four individuals from Cañada de la Virgen (E4_CdV_b, E7_CdV_b, E8_CdV_b, and E19_CdV_b), and two individuals from Michoacán (La Mina: E2_Mich_b, and E4_Mich_b). Genotyping data in plink format was converted to an unphased VCF format with Plink version 1.90 beta (112). The method used this VCF file as an input with the following parameters: each individual was plotted as an average of both parental haplotypes (AVERAGE_PARENTS=True), weights for each individual were used inversely proportional to the number of samples from the corresponding population (IS_WEIGHTED=True), the simple weighted of covariance PCA without any optimization was used (METHOD=1), no genotype was masked based on local ancestry calls (IS_MASKED=False), and only the three first PC's were calculated (NUM_DIMS=3). Individuals with $<0.15x$ whole genome coverage (19,851 SNP markers or less) were further down-weighted to 0.1 to avoid having the principal components be selected based on noise variance stemming from these samples. These included BC30, MOM6_ST_a, 11R_R_b, E4_CdV_b, E7_CdV_b, E19_CdV_b, E2_Mich_b, and E4_Mich_b.

Outgroup f3. To estimate each pre-Hispanic individual's genetic relatedness with a particular pre-Hispanic or present-day Indigenous population, we performed outgroup f3 statistics using ADMIXTOOLS v5.0 (65). Outgroup f3 was calculated in the form (Test, Source1; YRI), where the "Test" is the pre-Hispanic individual under analysis, and Source1 being another pre-Hispanic individual or present-day Indigenous population. We used all YRI from the 1000 Genomes Project (55, 109) as an outgroup population. Assuming there was no admixture in the tree, the f3 value is proportional to the shared genetic drift between Test and Source1.

D statistics. We performed D statistics using ADMIXTOOLS v5.0 (65) to identify whether a pre-Hispanic individual had higher genetic affinities with: i) other pre-Hispanic individuals from the same archaeological site than to pre-Hispanic individuals from other sites; or ii) a specific present-day indigenous population than to other present-day population. D values were computed in the form (Pop1, Pop2; Test, YRI). The "Test" refers to the pre-Hispanic individual under analysis, while Pop1 and Pop2 were all possible combinations between pre-Hispanic individuals and present-day Indigenous populations. Under the null hypothesis, Test individual would be equally related to

6 Pop1 and Pop2, and we expect no significant deviations from $D=0$. Significant $D<0$ indicates closer
7 relation between Test individual and Pop2. Moreover, a significant $D>0$ indicates closer relation
8 between Test individual and Pop1. Only absolute values of Z-score > 3 were considered as
9 significant (which corresponds to a p-value of ~ 0.0027).

1 qpGraph. We utilized the qpGraph tool of the ADMIXTOOLS v5.0 software package (65) to
2 construct admixture graphs and test which various models of demographic history best fit the data.
3 With the reference panel including Indigenous populations from Mexico previously published (50,
4 54, 110), we included in the baseline model the following individuals/populations: Mbuti as an
5 African outgroup, USR-1 (21) as the reference Beringian, Anzick-1 (22) as the reference ancestral
6 Southern Native American, and Athabascan (14) or ASO (15) as the reference ancestral Northern
7 Native American branch. Present-day Indigenous populations from Mexico and pre-Hispanic
8 individuals were subsequently added to the model in various combinations to test the fit. The
9 following parameters were applied: outpop (NULL), diag (0.0001), hires (YES), blgsize (0.05), and
0 lsqmode (YES). A worst fitting f statistical test of $|Z|$ score >3 is set as a threshold for model
1 rejection.

2
3 Population continuity test. According to the PCA, TFA, ADMIXTURE, outgroup-f3 and D
4 statistics, individual F9 from north Mexico is significantly closer to present-day Rarámuri. Thus,
5 we analyzed whether individual F9 belonged to the present-day Rarámuri ancestral population
6 using the method published in (66). Derived alleles in individual F9 and present-day Rarámuri were
7 identified using ancestral alleles' information from
8 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/pilot_data/technical/reference/ancestral_alignments).
9 Then, allele counts were performed for each site in the present-day Rarámuri (reference
0 population), excluding alleles with frequency 0 or 1. We performed allele counts for the pre-
1 Hispanic individual F9. Parameters alpha and beta necessary for the analysis were estimated using
2 $\text{min_samples} = 14$, which refers to the minimum number of individuals required in the reference
3 panel that have information for each SNV. Under the null hypothesis, it is assumed that reference
4 population and ancient individual belong to a continuous population, rejection of the null hypothesis
5 would mean no population continuity.

6
7 Conditional nucleotide diversity. We estimated relative levels of genetic diversity for pre-Hispanic
8 and present-day indigenous individuals from Mexico as previously reported in (50, 110). The
9 analysis was made according to the approach reported in (57). This consists of counting the
0 differences between the ascertained alleles present in one pair of individuals from the same
1 population. For this analysis, we used a panel of 1,938,919 transversions SNVs with a minor allele
2 frequency of 0.01 in the YRI population. In all cases, pseudohaploid genotype calls were performed
3 before the comparisons. CNV values were calculated for all pairs of individuals for each present-
4 day Indigenous population and for all pair of pre-Hispanic individuals for each archaeological
5 region. For present-day populations we masked the sites with no Native American ancestry to avoid
6 counting levels of diversity from a different ancestry. Mean and standard error values were
7 estimated 100 times using a window size of 50 Kb even for pairs of ancient individuals. Results are
8 reported in a violin plot, where each dot inside the violin corresponds to the CNV value for a unique
9 pair of individuals (Fig. S19).

0
1 Runs of homozygosity. We used the tool hapROH (58) to detect runs of homozygosity in pre-
2 Hispanic individuals. The distribution of ROHs was used as an indirect way of measuring genetic
3 diversity and effective population size. This method detects the ROH's lengths longer than four
4 centimorgans(cM) and up to 300 cM. The sum of the ROH detected gives insights into the
5 relatedness of the parents of each individual and informs about population sizes. Length of ROH

6 and relatedness between parents are positively correlated. While the higher the number of ROH
7 segments, the smaller the population size and the smaller the number of heterozygote sites, which
8 means less genetic diversity. hapROH uses a built-in set of 1240K SNVs from (115) to perform
9 pseudohaploid genotype calls and a reference panel of 5,008 present-day human haplotypes from
0 (54). We performed a pseudohaploid variant calling for the 1240K SNVs for each pre-Hispanic
1 genome. hapROH's manual recommends using samples with a minimum of 400K SNVs intersected
2 with the 1240K SNVs. However, we also included the individual E8_CdV_b, which intersected
3 369,504 SNVs, and the Pericú individual B03, which intersected 370,026 SNVs, to have at least
4 one individual of these sites in the analyses and considering the number of SNVs is not too far from
5 400K (Table S13). Detection of runs of homozygosity was made using the set of 5,008 haplotypes
6 mentioned above.

9 References:

- 0 1. A. L. Austin, L. L. Luján, *El pasado indígena* (Fondo de Cultura Económica, 2019).
- 1 2. B. B. Cornejo, *La frontera septentrional de Mesoamérica* (Instituto Nacional de
2 Antropología e Historia, México, ed. 2, 2000), vol. 1 of *Historia Antigua de México*.
- 3 3. P. Kirchhoff, Mesoamerica: its geographic limits, ethnic composition and cultural
4 characteristics. *Heritage of conquest*, 17–30 (1952).
- 5 4. M. del C. Solanes, E. Vela, Atlas del México prehispánico. *Arqueología Mexicana*. **5**, 5–
6 18 (2000).
- 7 5. D. W. Stahle, J. V. Diaz, D. J. Burnette, J. C. Paredes, R. R. Heim, F. K. Fye, R. Acuna
8 Soto, M. D. Therrell, M. K. Cleaveland, D. K. Stahle, Major Mesoamerican droughts of the past
9 millennium: MESOAMERICAN DROUGHTS. *Geophysical Research Letters*. **38** (2011)
0 (available at <http://doi.wiley.com/10.1029/2010GL046472>).
- 1 6. M. K. Hughes, H. F. Diaz, Was there a 'medieval warm period', and if so, where and
2 when? *Climatic Change*. **26**, 109–142 (1994).
- 3 7. P. Armillas, Condiciones ambientales y movimientos de pueblos en la frontera
4 septentrional de Mesoamérica. *Homenaje a Fernando Márquez-Miranda*, 62–82 (1964).
- 5 8. K. H. Wogau, H. W. Arz, H. N. Böhnelt, N. R. Nowaczyk, J. Park, High resolution
6 paleoclimate and paleoenvironmental reconstruction in the Northern Mesoamerican Frontier for
7 Prehistory to Historical times. *Quaternary Science Reviews*. **226**, 106001 (2019).
- 8 9. D. J. Kennett, S. F. M. Breitenbach, V. V. Aquino, Y. Asmerom, J. Awe, J. U. L. Baldini,
9 P. Bartlein, B. J. Culleton, C. Ebert, C. Jazwa, M. J. Macri, N. Marwan, V. Polyak, K. M. Prufer,
0 H. E. Ridley, H. Sodemann, B. Winterhalder, G. H. Haug, Development and Disintegration of
1 Maya Political Systems in Response to Climate Change. *Science*. **338**, 788–791 (2012).
- 2 10. J. A. Hoggarth, S. F. M. Breitenbach, B. J. Culleton, C. E. Ebert, M. A. Masson, D. J.
3 Kennett, The political collapse of Chichén Itzá in climatic and cultural context. *Global and
4 Planetary Change*, 18 (2016).
- 5 11. S. L. O'Hara, S. E. Metcalfe, F. A. Street-Perrott, On the arid margin: The relationship
6 between climate, humans and the environment. A review of evidence from the highlands of
7 central Mexico. *Chemosphere*. **29**, 965–981 (1994).
- 8 12. B. Braniff, Oscilación de la frontera norte mesoamericana: un nuevo ensayo. *Arqueología,
9 segunda época*, 99–114 (1989).
- 0 13. A. Gómez Aí-za, S. Sánchez Vázquez, La Frontera Cultural Meso-Aridoamericana:
1 Construcción De Imaginarios Nacionalistas En La Historia Mexicana. *Xih*. **9** (2014),
2 doi:10.37646/xihmai.v9i18.244.
- 3 14. M. Raghavan, M. Steinrucken, K. Harris, S. Schiffels, S. Rasmussen, M. DeGiorgio, A.
4 Albrechtsen, C. Valdiosera, M. C. Avila-Arcos, A.-S. Malaspinas, A. Eriksson, I. Moltke, M.

- 5 Metspalu, J. R. Homburger, J. Wall, O. E. Cornejo, J. V. Moreno-Mayar, T. S. Korneliussen, T.
6 Pierre, M. Rasmussen, P. F. Campos, P. d. B. Damgaard, M. E. Allentoft, J. Lindo, E. Metspalu,
7 R. Rodriguez-Varela, J. Mansilla, C. Henrickson, A. Seguin-Orlando, H. Malmstrom, T. Stafford,
8 S. S. Shringarpure, A. Moreno-Estrada, M. Karmin, K. Tambets, A. Bergstrom, Y. Xue, V.
9 Warmuth, A. D. Friend, J. Singarayer, P. Valdes, F. Balloux, I. Lebreiro, J. L. Vera, H. Rangel-
0 Villalobos, D. Pettener, D. Luiselli, L. G. Davis, E. Heyer, C. P. E. Zollikofer, M. S. Ponce de
1 Leon, C. I. Smith, V. Grimes, K.-A. Pike, M. Deal, B. T. Fuller, B. Arriaza, V. Standen, M. F.
2 Luz, F. Ricaut, N. Guidon, L. Osipova, M. I. Voevoda, O. L. Posukh, O. Balanovsky, M.
3 Lavryashina, Y. Bogunov, E. Khusnutdinova, M. Gubina, E. Balanovska, S. Fedorova, S.
4 Litvinov, B. Malyarchuk, M. Derenko, M. J. Mosher, D. Archer, J. Cybulski, B. Petzelt, J.
5 Mitchell, R. Worl, P. J. Norman, P. Parham, B. M. Kemp, T. Kivisild, C. Tyler-Smith, M. S.
6 Sandhu, M. Crawford, R. Villems, D. G. Smith, M. R. Waters, T. Goebel, J. R. Johnson, R. S.
7 Malhi, M. Jakobsson, D. J. Meltzer, A. Manica, R. Durbin, C. D. Bustamante, Y. S. Song, R.
8 Nielsen, E. Willerslev, Genomic evidence for the Pleistocene and recent population history of
9 Native Americans. *Science*. **349**, aab3884–aab3884 (2015).
- 0 15. C. L. Scheib, H. Li, T. Desai, V. Link, C. Kendall, G. Dewar, P. W. Griffith, A.
1 Mörseburg, J. R. Johnson, A. Potter, S. L. Kerr, P. Endicott, J. Lindo, M. Haber, Y. Xue, C.
2 Tyler-Smith, M. S. Sandhu, J. G. Lorenz, T. D. Randall, Z. Faltyskova, L. Pagani, P. Danecek, T.
3 C. O’Connell, P. Martz, A. S. Boraas, B. F. Byrd, A. Leventhal, R. Cambra, R. Williamson, L.
4 Lesage, B. Holguin, E. Ygnacio-De Soto, J. Rosas, M. Metspalu, J. T. Stock, A. Manica, A.
5 Scally, D. Wegmann, R. S. Malhi, T. Kivisild, Ancient human parallel lineages within North
6 America contributed to a coastal expansion. *Science*. **360**, 1024–1027 (2018).
- 7 16. T. Maricic, M. Whitten, S. Pääbo, Multiplexed DNA Sequence Capture of Mitochondrial
8 Genomes Using PCR Products. *PLoS ONE*. **5**, e14004 (2010).
- 9 17. M. L. Carpenter, J. D. Buenrostro, C. Valdiosera, H. Schroeder, M. E. Allentoft, M.
0 Sikora, M. Rasmussen, S. Gravel, S. Guillén, G. Nekhrizov, K. Leshtakov, D. Dimitrova, N.
1 Theodossiev, D. Pettener, D. Luiselli, K. Sandoval, A. Moreno-Estrada, Y. Li, J. Wang, M. T. P.
2 Gilbert, E. Willerslev, W. J. Greenleaf, C. D. Bustamante, Pulling out the 1%: Whole-Genome
3 Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. *The American*
4 *Journal of Human Genetics*. **93**, 852–864 (2013).
- 5 18. H. Jónsson, A. Ginolhac, M. Schubert, P. L. F. Johnson, L. Orlando, mapDamage2.0: fast
6 approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. **29**, 1682–
7 1684 (2013).
- 8 19. M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, M.
9 Metspalu, E. Metspalu, T. Kivisild, R. Gupta, M. Bertalan, K. Nielsen, M. T. P. Gilbert, Y. Wang,
0 M. Raghavan, P. F. Campos, H. M. Kamp, A. S. Wilson, A. Gledhill, S. Tridico, M. Bunce, E. D.
1 Lorenzen, J. Binladen, X. Guo, J. Zhao, X. Zhang, H. Zhang, Z. Li, M. Chen, L. Orlando, K.
2 Kristiansen, M. Bak, N. Tommerup, C. Bendixen, T. L. Pierre, B. Grønnow, M. Meldgaard, C.
3 Andreasen, S. A. Fedorova, L. P. Osipova, T. F. G. Higham, C. B. Ramsey, T. v. O. Hansen, F. C.
4 Nielsen, M. H. Crawford, S. Brunak, T. Sicheritz-Pontén, R. Villems, R. Nielsen, A. Krogh, J.
5 Wang, E. Willerslev, Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature*. **463**,
6 757–762 (2010).
- 7 20. P. Flegontov, N. E. Altınışık, P. Changmai, N. Rohland, S. Mallick, N. Adamski, D. A.
8 Bolnick, N. Broomandkoshbacht, F. Candilio, B. J. Culleton, O. Flegontova, T. M. Friesen, C.
9 Jeong, T. K. Harper, D. Keating, D. J. Kennett, A. M. Kim, T. C. Lamnidis, A. M. Lawson, I.
0 Olalde, J. Oppenheimer, B. A. Potter, J. Raff, R. A. Sattler, P. Skoglund, K. Stewardson, E. J.
1 Vajda, S. Vasilyev, E. Veselovskaya, M. G. Hayes, D. H. O’Rourke, J. Krause, R. Pinhasi, D.
2 Reich, S. Schiffels, Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North
3 America. *Nature*. **570**, 236–240 (2019).
- 4 21. J. V. Moreno-Mayar, B. A. Potter, L. Vinner, M. Steinrücken, S. Rasmussen, J. Terhorst,

- 5 J. A. Kamm, A. Albrechtsen, A.-S. Malaspinas, M. Sikora, J. D. Reuther, J. D. Irish, R. S. Malhi,
6 L. Orlando, Y. S. Song, R. Nielsen, D. J. Meltzer, E. Willerslev, Terminal Pleistocene Alaskan
7 genome reveals first founding population of Native Americans. *Nature*. **553**, 203–207 (2018).
- 8 22. M. Rasmussen, S. L. Anzick, M. R. Waters, P. Skoglund, M. DeGiorgio, T. W. Stafford,
9 S. Rasmussen, I. Moltke, A. Albrechtsen, S. M. Doyle, G. D. Poznik, V. Gudmundsdottir, R.
0 Yadav, A.-S. Malaspinas, S. S. W. V, M. E. Allentoft, O. E. Cornejo, K. Tambets, A. Eriksson, P.
1 D. Heintzman, M. Karmin, T. S. Korneliussen, D. J. Meltzer, T. L. Pierre, J. Stenderup, L. Saag,
2 V. M. Warmuth, M. C. Lopes, R. S. Malhi, S. Brunak, T. Sicheritz-Ponten, I. Barnes, M. Collins,
3 L. Orlando, F. Balloux, A. Manica, R. Gupta, M. Metspalu, C. D. Bustamante, M. Jakobsson, R.
4 Nielsen, E. Willerslev, The genome of a Late Pleistocene human from a Clovis burial site in
5 western Montana. *Nature*. **506**, 225–229 (2014).
- 6 23. C. Posth, N. Nakatsuka, I. Lazaridis, P. Skoglund, S. Mallick, T. C. Lamnidis, N. Rohland,
7 K. Nägele, N. Adamski, E. Bertolini, N. Broomandkhoshbacht, A. Cooper, B. J. Culleton, T.
8 Ferraz, M. Ferry, A. Furtwängler, W. Haak, K. Harkins, T. K. Harper, T. Hünemeier, A. M.
9 Lawson, B. Llamas, M. Michel, E. Nelson, J. Oppenheimer, N. Patterson, S. Schiffels, J. Sedig,
0 K. Stewardson, S. Talamo, C.-C. Wang, J.-J. Hublin, M. Hubbe, K. Harvati, A. Nuevo Delaunay,
1 J. Beier, M. Francken, P. Kaulicke, H. Reyes-Centeno, K. Rademaker, W. R. Trask, M. Robinson,
2 S. M. Gutierrez, K. M. Prufer, D. C. Salazar-García, E. N. Chim, L. Müller Plumm Gomes, M. L.
3 Alves, A. Liryo, M. Inglez, R. E. Oliveira, D. V. Bernardo, A. Barioni, V. Wesolowski, N. A.
4 Scheifler, M. A. Rivera, C. R. Plens, P. G. Messineo, L. Figuti, D. Corach, C. Scabuzzo, S.
5 Eggers, P. DeBlasis, M. Reindel, C. Méndez, G. Politis, E. Tomasto-Cagigao, D. J. Kennett, A.
6 Strauss, L. Fehren-Schmitz, J. Krause, D. Reich, Reconstructing the Deep Population History of
7 Central and South America. *Cell*. **175**, 1185–1197.e22 (2018).
- 8 24. C. de la Fuente, M. C. Ávila-Arcos, J. Galimany, M. L. Carpenter, J. R. Homburger, A.
9 Blanco, P. Contreras, D. Cruz Dávalos, O. Reyes, M. San Roman, A. Moreno-Estrada, P. F.
0 Campos, C. Eng, S. Huntsman, E. G. Burchard, A.-S. Malaspinas, C. D. Bustamante, E.
1 Willerslev, E. Llop, R. A. Verdugo, M. Moraga, Genomic insights into the origin and
2 diversification of late maritime hunter-gatherers from the Chilean Patagonia. *Proceedings of the*
3 *National Academy of Sciences*. **115**, E4006–E4012 (2018).
- 4 25. A. Y. Morales-Arce, C. A. Hofman, A. T. Duggan, A. K. Benfer, M. A. Katzenberg, G.
5 McCafferty, C. Warinner, Successful reconstruction of whole mitochondrial genomes from
6 ancient Central America and Mexico. *Scientific Reports*. **7**, 18100 (2017).
- 7 26. H.-J. Bandelt, C. Herrnstadt, Y.-G. Yao, Q.-P. Kong, T. Kivisild, C. Rengo, R. Scozzari,
8 M. Richards, R. Villems, V. Macaulay, N. Howell, A. Torroni, Y.-P. Zhang, Identification of
9 Native American Founder mtDNAs Through the Analysis of Complete mtDNA Sequences: Some
0 Caveats. *Ann Human Genet*. **67**, 512–524 (2003).
- 1 27. A. Gorostiza, V. Acunha-Alonzo, L. Regalado-Liu, S. Tirado, J. Granados, D. Sámano, H.
2 Rangel-Villalobos, A. González-Martín, Reconstructing the History of Mesoamerican Populations
3 through the Study of the Mitochondrial DNA Control Region. *PLoS ONE*. **7**, e44666 (2012).
- 4 28. A. González-Martín, A. Gorostiza, L. Regalado-Liu, S. Arroyo-Peña, S. Tirado, I. Nuño-
5 Arana, R. Rubi-Castellanos, K. Sandoval, M. D. Coble, H. Rangel-Villalobos, Demographic
6 History of Indigenous Populations in Mesoamerica Based on mtDNA Sequence Data. *PLOS*
7 *ONE*. **10**, e0131791 (2015).
- 8 29. B. Z. González-Sobrino, A. P. Pintado-Cortina, L. Sebastián-Medina, F. Morales-
9 Mandujano, A. V. Contreras, Y. E. Aguilar, J. Chávez-Benavides, A. Carrillo-Rodríguez, I. Silva-
0 Zolezzi, L. Medrano-González, Genetic Diversity and Differentiation in Urban and Indigenous
1 Populations of Mexico: Patterns of Mitochondrial DNA and Y-Chromosome Lineages.
2 *Biodemography and Social Biology*. **62**, 53–72 (2016).
- 3 30. J. Mata-Míguez, L. Overholtzer, E. Rodríguez-Alegría, B. M. Kemp, D. A. Bolnick, The
4 genetic impact of aztec imperialism: Ancient mitochondrial DNA evidence from Xaltocan,

- 5 Mexico. *American Journal of Physical Anthropology*. **149**, 504–516 (2012).
- 6 31. A. Y. Morales-Arce, G. McCafferty, J. Hand, N. Schmill, K. McGrath, C. Speller, Ancient
7 mitochondrial DNA and population dynamics in postclassic Central Mexico: Tlatelolco (ad 1325–
8 1520) and Cholula (ad 900–1350). *Archaeological and Anthropological Sciences* (2019),
9 doi:10.1007/s12520-018-00771-7.
- 0 32. A. Y. Morales-Arce, M. H. Snow, J. H. Kelley, M. Anne Katzenberg, Ancient
1 mitochondrial DNA and ancestry of Paquimé inhabitants, Casas Grandes (A.D. 1200–1450).
2 *American Journal of Physical Anthropology*. **163**, 616–626 (2017).
- 3 33. A. González-Oliver, L. Márquez-Morfin, J. C. Jiménez, A. Torre-Blanco, Founding
4 Amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo: mtDNA
5 Lineages in Ancient Maya. *American Journal of Physical Anthropology*. **116**, 230–235 (2001).
- 6 34. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N.
7 Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N.
8 Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F.
9 Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C.
0 Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S.
1 Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M.
2 S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E.
3 Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G.
4 Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D.
5 Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B.
6 Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K.
7 Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300
8 genomes from 142 diverse populations. *Nature*. **538**, 201–206 (2016).
- 9 35. B. M. Kemp, A. Gonzalez-Oliver, R. S. Malhi, C. Monroe, K. B. Schroeder, J.
0 McDonough, G. Rhett, A. Resendez, R. I. Penaloza-Espinosa, L. Buentello-Malo, C. Gorodesky,
1 D. G. Smith, Evaluating the Farming/Language Dispersal Hypothesis with genetic variation
2 exhibited by populations in the Southwest and Mesoamerica. *Proceedings of the National*
3 *Academy of Sciences*. **107**, 6759–6764 (2010).
- 4 36. M. Bodner, U. A. Perego, J. E. Gomez, R. M. Cerda-Flores, N. Rambaldi Migliore, S. R.
5 Woodward, W. Parson, A. Achilli, The Mitochondrial DNA Landscape of Modern Mexico.
6 *Genes*. **12**, 1453 (2021).
- 7 37. I. De La Cruz, A. González-Oliver, B. M. Kemp, J. A. Román, D. G. Smith, A. Torre-
8 Blanco, Sex Identification of Children Sacrificed to the Ancient Aztec Rain Gods in Tlatelolco.
9 *Current Anthropology*. **49**, 519–526 (2008).
- 0 38. M. C. Ávila-Arcos, K. F. McManus, K. Sandoval, J. E. Rodríguez-Rodríguez, V. Villa-
1 Islas, A. R. Martin, P. Luisi, R. I. Peñaloza-Espinosa, C. Eng, S. Huntsman, E. G. Burchard, C. R.
2 Gignoux, C. D. Bustamante, A. Moreno-Estrada, Population History and Gene Divergence in
3 Native Mexicans Inferred from 76 Human Exomes. *Molecular Biology and Evolution*. **37**, 994–
4 1006 (2020).
- 5 39. S. Romero-Hidalgo, A. Ochoa-Leyva, A. Garcíaarrubio, V. Acuña-Alonzo, E. Antúnez-
6 Argüelles, M. Balcazar-Quintero, R. Barquera-Lozano, A. Carnevale, F. Cornejo-Granados, J. C.
7 Fernández-López, R. García-Herrera, H. García-Ortíz, Á. Granados-Silvestre, J. Granados, F.
8 Guerrero-Romero, E. Hernández-Lemus, P. León-Mimila, G. Macín-Pérez, A. Martínez-
9 Hernández, M. Menjivar, E. Morett, L. Orozco, G. Ortiz-López, F. Pérez-Villatoro, J. Rivera-
0 Morales, F. Riveros-McKay, M. Villalobos-Comparán, H. Villamil-Ramírez, T. Villarreal-
1 Molina, S. Canizales-Quinteros, X. Soberón, Demographic history and biologically relevant
2 genetic variation of Native Mexicans inferred from whole-genome sequencing. *Nature*
3 *Communications*. **8**, 1005 (2017).
- 4 40. M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human

- 5 mitochondrial DNA variation. *Human Mutation*. **30**, E386–E394 (2009).
- 6 41. B. Llamas, L. Fehren-Schmitz, G. Valverde, J. Soubrier, S. Mallick, N. Rohland, S.
7 Nordenfelt, C. Valdiosera, S. M. Richards, A. Rohrlach, M. I. B. Romero, I. F. Espinoza, E. T.
8 Cagigao, L. W. Jiménez, K. Makowski, I. S. L. Reyna, J. M. Lory, J. A. B. Torrez, M. A. Rivera,
9 R. L. Burger, M. C. Ceruti, J. Reinhard, R. S. Wells, G. Politis, C. M. Santoro, V. G. Standen, C.
0 Smith, D. Reich, S. Y. W. Ho, A. Cooper, W. Haak, Ancient mitochondrial DNA provides high-
1 resolution time scale of the peopling of the Americas. *Science Advances*. **2**, e1501385 (2016).
- 2 42. B. D. O’Fallon, L. Fehren-Schmitz, Native Americans experienced a strong population
3 bottleneck coincident with European contact. *Proceedings of the National Academy of Sciences*.
4 **108**, 20444–20448 (2011).
- 5 43. M. A. Nieves-Colón, W. J. Pestle, A. W. Reynolds, B. Llamas, C. de la Fuente, K. Fowler,
6 K. M. Skerry, E. Crespo-Torres, C. D. Bustamante, A. C. Stone, Ancient DNA Reconstructs the
7 Genetic Legacies of Precontact Puerto Rico Communities. *Molecular Biology and Evolution*. **37**,
8 611–626 (2020).
- 9 44. S. Gravel, F. Zakharia, A. Moreno-Estrada, J. K. Byrnes, M. Muzzio, J. L. Rodriguez-
0 Flores, E. E. Kenny, C. R. Gignoux, B. K. Maples, W. Guiblet, J. Dutil, M. Via, K. Sandoval, G.
1 Bedoya, The 1000 Genomes Project, T. K. Oleksyk, A. Ruiz-Linares, E. G. Burchard, J. C.
2 Martinez-Cruzado, C. D. Bustamante, Reconstructing Native American Migrations from Whole-
3 Genome and Whole-Exome Data. *PLoS Genet*. **9**, e1004023 (2013).
- 4 45. P. Skoglund, J. Storå, A. Götherström, M. Jakobsson, Accurate sex identification of
5 ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science*. **40**,
6 4477–4482 (2013).
- 7 46. P. Skoglund, B. H. Northoff, M. V. Shunkov, A. P. Derevianko, S. Pääbo, J. Krause, M.
8 Jakobsson, Separating endogenous ancient DNA from modern day contamination in a Siberian
9 Neandertal. *Proceedings of the National Academy of Sciences*. **111**, 2229–2234 (2014).
- 0 47. R. Fregel, F. L. Méndez, Y. Bokbot, D. Martín-Socas, M. D. Camalich-Massieu, J.
1 Santana, J. Morales, M. C. Ávila-Arcos, P. A. Underhill, B. Shapiro, G. Wojcik, M. Rasmussen,
2 A. E. R. Soares, J. Kapp, A. Sockell, F. J. Rodríguez-Santos, A. Mikdad, A. Trujillo-Mederos, C.
3 D. Bustamante, Ancient genomes from North Africa evidence prehistoric migrations to the
4 Maghreb from both the Levant and Europe. *Proc Natl Acad Sci USA*. **115**, 6774–6779 (2018).
- 5 48. D. Perez-Benedico, J. La Salvia, Z. Zeng, G. A. Herrera, R. Garcia-Bertrand, R. J.
6 Herrera, Mayans: a Y chromosome perspective. *Eur J Hum Genet*. **24**, 1352–1358 (2016).
- 7 49. K. Sandoval, A. Moreno-Estrada, I. Mendizabal, P. A. Underhill, M. Lopez-Valenzuela,
8 R. Peñaloza-Espinosa, M. Lopez-Lopez, L. Buentello-Malo, H. Avelino, F. Calafell, D. Comas,
9 Y-chromosome diversity in Native Mexicans reveals continental transition of genetic structure in
0 the Americas. *Am. J. Phys. Anthropol.* **148**, 395–405 (2012).
- 1 50. A. Bergström, S. A. McCarthy, R. Hui, M. A. Almarri, Q. Ayub, P. Danecek, Y. Chen, S.
2 Felkel, P. Hallast, J. Kamm, H. Blanché, J.-F. Deleuze, H. Cann, S. Mallick, D. Reich, M. S.
3 Sandhu, P. Skoglund, A. Scally, Y. Xue, R. Durbin, C. Tyler-Smith, Insights into human genetic
4 variation and population history from 929 diverse genomes. *Science*. **367**, eaay5012 (2020).
- 5 51. D. H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in
6 unrelated individuals. *Genome Res*. **19**, 1655–1664 (2009).
- 7 52. N. Patterson, A. L. Price, D. Reich, Population Structure and Eigenanalysis. *PLoS*
8 *Genetics*. **2**, e190 (2006).
- 9 53. D. Reich, A. L. Price, N. Patterson, Principal component analysis of genetic data. *Nature*
0 *Genetics*. **40**, 491–492 (2008).
- 1 54. A. Moreno-Estrada, C. R. Gignoux, J. C. Fernandez-Lopez, F. Zakharia, M. Sikora, A. V.
2 Contreras, V. Acuna-Alonzo, K. Sandoval, C. Eng, S. Romero-Hidalgo, P. Ortiz-Tello, V. Robles,
3 E. E. Kenny, I. Nuno-Arana, R. Barquera-Lozano, G. Macin-Perez, J. Granados-Arriola, S.
4 Huntsman, J. M. Galanter, M. Via, J. G. Ford, R. Chapela, W. Rodriguez-Cintrón, J. R.

- 5 Rodriguez-Santana, I. Romieu, J. J. Sienna-Monge, B. d. R. Navarro, S. J. London, A. Ruiz-
6 Linares, R. Garcia-Herrera, K. Estrada, A. Hidalgo-Miranda, G. Jimenez-Sanchez, A. Carnevale,
7 X. Soberon, S. Canizales-Quinteros, H. Rangel-Villalobos, I. Silva-Zolezzi, E. G. Burchard, C. D.
8 Bustamante, The genetics of Mexico recapitulates Native American substructure and affects
9 biomedical traits. *Science*. **344**, 1280–1285 (2014).
- 0 55. The 1000 Genomes Project Consortium, A global reference for human genetic variation.
1 *Nature*. **526**, 68–74 (2015).
- 2 56. O. François, F. Jay, Factor analysis of ancient population genomic samples. *Nat Commun*.
3 **11**, 4661 (2020).
- 4 57. P. Skoglund, H. Malmstrom, A. Omrak, M. Raghavan, C. Valdiosera, T. Gunther, P. Hall,
5 K. Tambets, J. Parik, K.-G. Sjogren, J. Apel, E. Willerslev, J. Stora, A. Gotherstrom, M.
6 Jakobsson, Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and
7 Farmers. *Science*. **344**, 747–750 (2014).
- 8 58. H. Ringbauer, J. Novembre, M. Steinrücken, Parental relatedness through time revealed
9 by runs of homozygosity in ancient DNA. *Nat Commun*. **12**, 5425 (2021).
- 0 59. S. J. Macfarlan, C. N. Henrickson, Inferring Relationships Between Indigenous Baja
1 California Sur and Seri/Comcáac Populations Through Cultural Traits. *Journal of California and*
2 *Great Basin Anthropology*. **30**, 18 (2021).
- 3 60. M. Kirin, R. McQuillan, C. S. Franklin, H. Campbell, P. M. McKeigue, J. F. Wilson,
4 Genomic Runs of Homozygosity Record Population History and Consanguinity. *PLoS ONE*. **5**,
5 e13996 (2010).
- 6 61. T. J. Pemberton, D. Absher, M. W. Feldman, R. M. Myers, N. A. Rosenberg, J. Z. Li,
7 Genomic Patterns of Homozygosity in Worldwide Human Populations. *The American Journal of*
8 *Human Genetics*. **91**, 275–292 (2012).
- 9 62. R. McQuillan, A.-L. Leutenegger, R. Abdel-Rahman, C. S. Franklin, M. Pericic, L. Barac-
0 Lauc, N. Smolej-Narancic, B. Janicijevic, O. Polasek, A. Tenesa, A. K. MacLeod, S. M.
1 Farrington, P. Rudan, C. Hayward, V. Vitart, I. Rudan, S. H. Wild, M. G. Dunlop, A. F. Wright,
2 H. Campbell, J. F. Wilson, Runs of Homozygosity in European Populations. *The American*
3 *Journal of Human Genetics*. **83**, 359–372 (2008).
- 4 63. L. Yengo, N. R. Wray, P. M. Visscher, Extreme inbreeding in a European ancestry sample
5 from the contemporary UK population. *Nat Commun*. **10**, 3719 (2019).
- 6 64. F. C. Ceballos, P. K. Joshi, D. W. Clark, M. Ramsay, J. F. Wilson, Runs of homozygosity:
7 windows into population history and trait architecture. *Nature Reviews Genetics*. **19**, 220–234
8 (2018).
- 9 65. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T.
0 Webster, D. Reich, Ancient Admixture in Human History. *Genetics*. **192**, 1065–1093 (2012).
- 1 66. J. G. Schraiber, Assessing the Relationship of Ancient and Modern Populations. *Genetics*.
2 **208**, 383–398 (2018).
- 3 67. J. V. Moreno-Mayar, L. Vinner, P. de Barros Damgaard, C. de la Fuente, J. Chan, J. P.
4 Spence, M. E. Allentoft, T. Vimala, F. Racimo, T. Pinotti, S. Rasmussen, A. Margaryan, M. Iraeta
5 Orbegozo, D. Mylopotamitaki, M. Wooller, C. Bataille, L. Becerra-Valdivia, D. Chivall, D.
6 Comeskey, T. Devièse, D. K. Grayson, L. George, H. Harry, V. Alexandersen, C. Primeau, J.
7 Erlandson, C. Rodrigues-Carvalho, S. Reis, M. Q. R. Bastos, J. Cybulski, C. Vullo, F. Morello,
8 M. Vilar, S. Wells, K. Gregersen, K. L. Hansen, N. Lynnerup, M. Mirazón Lahr, K. Kjær, A.
9 Strauss, M. Alfonso-Durruty, A. Salas, H. Schroeder, T. Higham, R. S. Malhi, J. T. Rasic, L.
0 Souza, F. R. Santos, A.-S. Malaspinas, M. Sikora, R. Nielsen, Y. S. Song, D. J. Meltzer, E.
1 Willerslev, Early human dispersals within the Americas. *Science*. **362**, eaav2621 (2018).
- 2 68. H. García-Ortiz, F. Barajas-Olmos, C. Contreras-Cubas, M. Á. Cid-Soto, E. J. Córdova, F.
3 Centeno-Cruz, E. Mendoza-Caamal, I. Cicerón-Arellano, M. Flores-Huacuja, P. Baca, D. A.
4 Bolnick, M. Snow, S. E. Flores-Martínez, R. Ortiz-Lopez, A. W. Reynolds, A. Blanchet, M.

- 5 Morales-Marín, R. Velázquez-Cruz, A. D. Kostic, C. Galaviz-Hernández, A. G. García-Zapién, J.
6 C. Jiménez-López, G. León-Reyes, E. G. Salas-Bautista, B. P. Lazalde-Ramos, J. L. Jiménez-
7 Ruíz, G. Salas-Martínez, J. Ramos-Madrigal, E. Mirzaeicheshmeh, Y. Saldaña-Alvarez, M. del
8 Carmen Abrahantes-Pérez, F. Loeza-Becerra, R. Mojica-Espinosa, F. Sánchez-Quinto, H. Rangel-
9 Villalobos, M. Sosa-Macías, J. Sánchez-Corona, A. Rojas-Martínez, A. Martínez-Hernández, L.
0 Orozco, The genomic landscape of Mexican Indigenous populations brings insights into the
1 peopling of the Americas. *Nat Commun.* **12**, 5942 (2021).
- 2 69. M. A. Nieves-Colo, Ancient DNA Reconstructs the Genetic Legacies of Precontact Puerto
3 Rico Communities, 16.
- 4 70. R. Acuna-Soto, Megadrought and Megadeath in 16th Century Mexico. *Emerging*
5 *Infectious Diseases.* **8**, 360–362 (2002).
- 6 71. G. Kubler, *The Population of Central Mexico in the Sixteenth Century.* (JSTOR, 1948).
- 7 72. A. J. Herrera Muñoz, E. Mejía Campos, Un minero en la Sierra Gorda: caso de
8 contaminación ocupacional multielemental de metales pesados a finales del periodo Clásico.
9 *Arqueología (México, DF)*, 127–135 (2017).
- 0 73. E. Nalda, La arqueología de Guanajuato: trabajos recientes. *Arqueología mexicana.* **16**,
1 36–43 (2008).
- 2 74. M. E. Smith, Mesoamerica’s First World City: Teotihuacan in Comparative Context.
3 *Teotihuacan: The World Beyond the City* (2020).
- 4 75. B. A. Álvarez-Sandoval, L. R. Manzanilla, M. González-Ruiz, A. Malgosa, R. Montiel,
5 Genetic Evidence Supports the Multiethnic Character of Teopancazco, a Neighborhood Center of
6 Teotihuacan, Mexico (AD 200-600). *PLoS ONE.* **10**, e0132371 (2015).
- 7 76. E. M. P. Campos, INTERPRETACIÓN PRELIMINAR RESPECTO A LA
8 TEMPORALIDAD DE TOLUQUILLA, QUERÉTARO, 24.
- 9 77. E. Campos, J. Campos, A. Muñoz, Mercurialism Determination in Fetuses Bone Remains
0 from Toluquilla, Queretaro, Mexico. *JSRR.* **8**, 1–10 (2015).
- 1 78. E. M. Pérez Campos, A. J. Herrera Muñoz, MINAS Y MINEROS: PRESENCIA DE
2 METALES EN SEDIMENTOS Y RESTOS HUMANOS AL SUR DE LA SIERRA GORDA DE
3 QUERÉTARO EN MÉXICO. *Chungará (Arica).* **45**, 161–176 (2013).
- 4 79. C. Paciotto, The Tarahumara of Mexico. *life*, 35 (1991).
- 5 80. D. A. Corona, Expansión Territorial Comercial en Mesoamérica y Mesoamérica
6 Septentrional por medio del Intercambio. **19**, 10 (2015).
- 7 81. N. Rohland, M. Hofreiter, Ancient DNA extraction from bones and teeth. *Nat Protoc.* **2**,
8 1756–1762 (2007).
- 9 82. J. Dabney, M. Knapp, I. Glocke, M.-T. Gansauge, A. Weihmann, B. Nickel, C.
0 Valdiosera, N. Garcia, S. Paabo, J.-L. Arsuaga, M. Meyer, Complete mitochondrial genome
1 sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments.
2 *Proceedings of the National Academy of Sciences.* **110**, 15758–15763 (2013).
- 3 83. M. Meyer, M. Kircher, Illumina Sequencing Library Preparation for Highly Multiplexed
4 Target Capture and Sequencing. *Cold Spring Harbor Protocols.* **2010** (2010) (available at
5 <http://www.cshprotocols.org/cgi/doi/10.1101/pdb.prot5448>).
- 6 84. T. Daley, A. D. Smith, Predicting the molecular complexity of sequencing libraries.
7 *Nature Methods.* **10**, 325–327 (2013).
- 8 85. M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming,
9 identification, and read merging. *BMC Research Notes.* **9**, 88 (2016).
- 0 86. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis,
1 R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map
2 format and SAMtools. *Bioinformatics.* **25**, 2078–2079 (2009).
- 3 87. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K.
4 Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A

- 5 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*.
6 **20**, 1297–1303 (2010).
- 7 88. V. Link, A. Kousathanas, K. Veeramah, C. Sell, A. Scheu, D. Wegmann, “ATLAS:
8 Analysis Tools for Low-depth and Ancient Samples” (preprint, Bioinformatics, 2017), ,
9 doi:10.1101/105346.
- 0 89. G. Renaud, V. Slon, A. T. Duggan, J. Kelso, Schmutzi: estimation of contamination and
1 endogenous mitochondrial consensus calling for ancient DNA. *Genome Biology*. **16**, 224 (2015).
- 2 90. H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt, F.
3 Kronenberg, A. Salas, S. Schönherr, HaploGrep 2: mitochondrial haplogroup classification in the
4 era of high-throughput sequencing. *Nucleic Acids Res*. **44**, W58–W63 (2016).
- 5 91. M. Van Oven, PhyloTree Build 17: Growing the human mitochondrial DNA tree.
6 *Forensic Science International: Genetics Supplement Series*. **5**, e392–e394 (2015).
- 7 92. S. Kumar, C. Bellis, M. Zlojutro, P. E. Melton, J. Blangero, J. E. Curran, Large scale
8 mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American
9 origins. *BMC evolutionary biology*. **11**, 1–17 (2011).
- 0 93. U. A. Perego, A. Achilli, N. Angerhofer, M. Accetturo, M. Pala, A. Olivieri, B. H.
1 Kashani, K. H. Ritchie, R. Scozzari, Q.-P. Kong, N. M. Myres, A. Salas, O. Semino, H.-J.
2 Bandelt, S. R. Woodward, A. Torroni, Distinctive Paleo-Indian Migration Routes from Beringia
3 Marked by Two Rare mtDNA Haplogroups. *Current Biology*. **19**, 1–8 (2009).
- 4 94. A. Achilli, U. A. Perego, H. Lancioni, A. Olivieri, F. Gandini, B. H. Kashani, V. Battaglia,
5 V. Grugni, N. Angerhofer, M. P. Rogers, Reconciling migration models to the Americas with the
6 variation of North American native mitogenomes. *Proceedings of the National Academy of
7 Sciences*. **110**, 14308–14313 (2013).
- 8 95. A. Achilli, A. Olivieri, M. Pala, B. Hooshiar Kashani, V. Carossa, U. A. Perego, F.
9 Gandini, A. Santoro, V. Battaglia, V. Grugni, H. Lancioni, C. Sirolla, A. R. Bonfigli, A. Cormio,
0 M. Boemi, I. Testa, O. Semino, A. Ceriello, L. Spazzafumo, M. N. Gadaleta, M. Marra, R. Testa,
1 C. Franceschi, A. Torroni, Mitochondrial DNA Backgrounds Might Modulate Diabetes
2 Complications Rather than T2DM as a Whole. *PLoS ONE*. **6**, e21029 (2011).
- 3 96. U. A. Perego, N. Angerhofer, M. Pala, A. Olivieri, H. Lancioni, B. H. Kashani, V.
4 Carossa, J. E. Ekins, A. Gomez-Carballa, G. Huber, B. Zimmermann, D. Corach, N. Babudri, F.
5 Panara, N. M. Myres, W. Parson, O. Semino, A. Salas, S. R. Woodward, A. Achilli, A. Torroni,
6 The initial peopling of the Americas: A growing number of founding mitochondrial genomes
7 from Beringia. *Genome Research*. **20**, 1174–1179 (2010).
- 8 97. D. M. Behar, M. van Oven, S. Rosset, M. Metspalu, E.-L. Loogväli, N. M. Silva, T.
9 Kivisild, A. Torroni, R. Villems, A “Copernican” Reassessment of the Human Mitochondrial
0 DNA Tree from its Root. *The American Journal of Human Genetics*. **90**, 675–684 (2012).
- 1 98. A. Hartmann, M. Thieme, L. K. Nanduri, T. Stempf, C. Moehle, T. Kivisild, P. J. Oefner,
2 Validation of microarray-based resequencing of 93 worldwide mitochondrial genomes. *Hum.*
3 *Mutat*. **30**, 115–122 (2009).
- 4 99. R. S. Just, T. M. Diegoli, J. L. Saunier, J. A. Irwin, T. J. Parsons, Complete mitochondrial
5 genome sequences for 265 African American and U.S. “Hispanic” individuals. *Forensic Science
6 International: Genetics*. **2**, e45–e48 (2008).
- 7 100. M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, Mitochondrial genome variation and
8 the origin of modern humans. *Nature*. **408**, 708–713 (2000).
- 9 101. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary
0 genetics analysis across computing platforms. *Molecular biology and evolution*. **35**, 1547 (2018).
- 1 102. J. W. Leigh, D. Bryant, POPART : full-feature software for haplotype network construction.
2 *Methods Ecol Evol*. **6**, 1110–1116 (2015).
- 3 103. J. Rozas, A. Ferrer-Mata, J. C. Sánchez-DelBarrio, S. Guirao-Rico, P. Librado, S. E.
4 Ramos-Onsins, A. Sánchez-Gracia, DnaSP 6: DNA Sequence Polymorphism Analysis of Large

- 5 Data Sets. *Molecular Biology and Evolution*. **34**, 3299–3302 (2017).
- 6 104. J. Heled, A. J. Drummond, Bayesian inference of population size history from multiple
7 loci. *BMC Evol Biol*. **8**, 289 (2008).
- 8 105. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A.
9 Rambaut, A. J. Drummond, BEAST 2: A Software Platform for Bayesian Evolutionary Analysis.
0 *PLoS Comput Biol*. **10**, e1003537 (2014).
- 1 106. R. Lanfear, P. B. Frandsen, A. M. Wright, T. Senfeld, B. Calcott, PartitionFinder 2: New
2 Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological
3 Phylogenetic Analyses. *Mol Biol Evol*, msw260 (2016).
- 4 107. P. Soares, L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Röhl, A. Salas, S.
5 Oppenheimer, V. Macaulay, M. B. Richards, Correcting for Purifying Selection: An Improved
6 Human Mitochondrial Molecular Clock. *The American Journal of Human Genetics*. **84**, 740–759
7 (2009).
- 8 108. The 1000 Genomes Project Consortium, G. D. Poznik, Y. Xue, F. L. Mendez, T. F.
9 Willems, A. Massaia, M. A. Wilson Sayres, Q. Ayub, S. A. McCarthy, A. Narechania, S. Kashin,
0 Y. Chen, R. Banerjee, J. L. Rodriguez-Flores, M. Cerezo, H. Shao, M. Gymrek, A. Malhotra, S.
1 Louzada, R. Desalle, G. R. S. Ritchie, E. Cerveira, T. W. Fitzgerald, E. Garrison, A. Marcketta,
2 D. Mittelman, M. Romanovitch, C. Zhang, X. Zheng-Bradley, G. R. Abecasis, S. A. McCarroll,
3 P. Flicek, P. A. Underhill, L. Coin, D. R. Zerbino, F. Yang, C. Lee, L. Clarke, A. Auton, Y.
4 Erlich, R. E. Handsaker, C. D. Bustamante, C. Tyler-Smith, Punctuated bursts in human male
5 demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics*. **48**,
6 593–599 (2016).
- 7 109. The 1000 Genomes Project Consortium, An integrated map of genetic variation from
8 1,092 human genomes. *Nature*. **491**, 56–65 (2012).
- 9 110. S. Mallick, H. Li, M. Lipson, I. Mathieson, M. Gymrek, F. Racimo, M. Zhao, N.
0 Chennagiri, S. Nordenfelt, A. Tandon, P. Skoglund, I. Lazaridis, S. Sankararaman, Q. Fu, N.
1 Rohland, G. Renaud, Y. Erlich, T. Willems, C. Gallo, J. P. Spence, Y. S. Song, G. Poletti, F.
2 Balloux, G. van Driem, P. de Knijff, I. G. Romero, A. R. Jha, D. M. Behar, C. M. Bravi, C.
3 Capelli, T. Hervig, A. Moreno-Estrada, O. L. Posukh, E. Balanovska, O. Balanovsky, S.
4 Karachanak-Yankova, H. Sahakyan, D. Toncheva, L. Yepiskoposyan, C. Tyler-Smith, Y. Xue, M.
5 S. Abdullah, A. Ruiz-Linares, C. M. Beall, A. Di Rienzo, C. Jeong, E. B. Starikovskaya, E.
6 Metspalu, J. Parik, R. Villems, B. M. Henn, U. Hodoglugil, R. Mahley, A. Sajantila, G.
7 Stamatoyannopoulos, J. T. S. Wee, R. Khusainova, E. Khusnutdinova, S. Litvinov, G. Ayodo, D.
8 Comas, M. F. Hammer, T. Kivisild, W. Klitz, C. A. Winkler, D. Labuda, M. Bamshad, L. B.
9 Jorde, S. A. Tishkoff, W. S. Watkins, M. Metspalu, S. Dryomov, R. Sukernik, L. Singh, K.
0 Thangaraj, S. Pääbo, J. Kelso, N. Patterson, D. Reich, The Simons Genome Diversity Project: 300
1 genomes from 142 diverse populations. *Nature*. **538**, 201–206 (2016).
- 2 111. B. K. Maples, S. Gravel, E. E. Kenny, C. D. Bustamante, RFMix: A Discriminative
3 Modeling Approach for Rapid and Robust Local-Ancestry Inference. *The American Journal of*
4 *Human Genetics*. **93**, 278–288 (2013).
- 5 112. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller,
6 P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A Tool Set for Whole-Genome
7 Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*.
8 **81**, 559–575 (2007).
- 9 113. J. M. Monroy Kuhn, M. Jakobsson, T. Günther, Estimating genetic kin relationships in
0 prehistoric populations. *PLoS ONE*. **13**, e0195491 (2018).
- 1 114. A. A. Behr, K. Z. Liu, G. Liu-Fang, P. Nakka, S. Ramachandran, pong: fast analysis and
2 visualization of latent clusters in population genetic data. *Bioinformatics*. **32**, 2817–2823 (2016).
- 3 115. Q. Fu, M. Hajdinjak, O. T. Moldovan, S. Constantin, S. Mallick, P. Skoglund, N.
4 Patterson, N. Rohland, I. Lazaridis, B. Nickel, B. Viola, K. Prüfer, M. Meyer, J. Kelso, D. Reich,

5 S. Pääbo, An early modern human from Romania with a recent Neanderthal ancestor. *Nature*.
6 **524**, 216–219 (2015).

0 **Acknowledgments**

1
2 **Acknowledgments:** We greatly acknowledge the IT support of Luis Alberto Aguilar
3 Bautista, Alejandro de León Cuevas, Carlos Sair Flores Bautista and Jair Garcia Sotelo of
4 the Laboratorio Nacional de Visualización Científica Avanzada (LAVIS UNAM). We also
5 thank Alejandra Castillo Carbajal and Carina Uribe Díaz from LIIGH-UNAM for wetlab
6 technical support during the project. We thank Alexandra Sockell for technical support at
7 the Bustamante Lab. We thank Diego Ortega-Del Vecchyo and Angélica González Oliver
8 for enriching comments throughout the project on earlier versions of the manuscripts.
9

0 **Funding:** Viridiana Villa-Islas is a doctoral student from the Programa de Doctorado en Ciencias
1 Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received
2 CONACyT fellowship (Número de becaria: 282188). M.C.A.-A. received support from
3 UNAM-PAPIIT (Project numbers: IA201219 and IA203821) and the Wellcome Trust Seed
4 Award in Science (grant number: 208934/Z/17/Z), and the International Centre for Genetic
5 Engineering and Biotechnology (project number: CRP/MEX10-03). M.C.A.-A., E.H.-S.,
6 and F.J. receive support from the Human Frontiers Science Program (grant number:
7 RGY0075/201). R.F. was supported by a grant (PALEUCOL; PGC2018-094101-A-I00)
8 from FEDER/Ministerio de Ciencia e Innovación—Agencia Estatal de Investigación and
9 from CajaCanarias and La Caixa Banking foundations (GENPAC; 2018PATRI16). Data
0 generation for the Sierra Tarahumara Individuals was funded with a Trainee Research Grant
1 to M.C.A.-A. by Stanford Center for Computational, Evolutionary and Human Genomics.
2

3 **Author contributions:** M.C.A.-A. and V.V.-I. conceived and designed the study with input from
4 E.M.P.C. V.V.-I., M.S.-V., M.B.-L., B.M., R.F., K.S., and M.A.N.-C. performed the wetlab
5 work. V.V.-I. processed all sequence data. V.V.-I. carried out most analyses with help from,
6 A.I.-G., M.L., J.M.T., J.E.R.-R., D.A.V.-R., and F.S.-Q. E.M.P.C., A.H.-M., K.S., G.Z.,
7 M.V.-M, R.A.-H., A.G.-O., C.V. contributed samples, provided the archaeological context
8 and discussed results. F.A.V., A.M.-E., F.J., A.I., E.H.-S., and F.S.-Q. suggested analyses,
9 discussed results and provided feedback throughout the project. V.V.-I. and M.C.A.-A.
0 wrote and edited the manuscript with input from all authors.
1

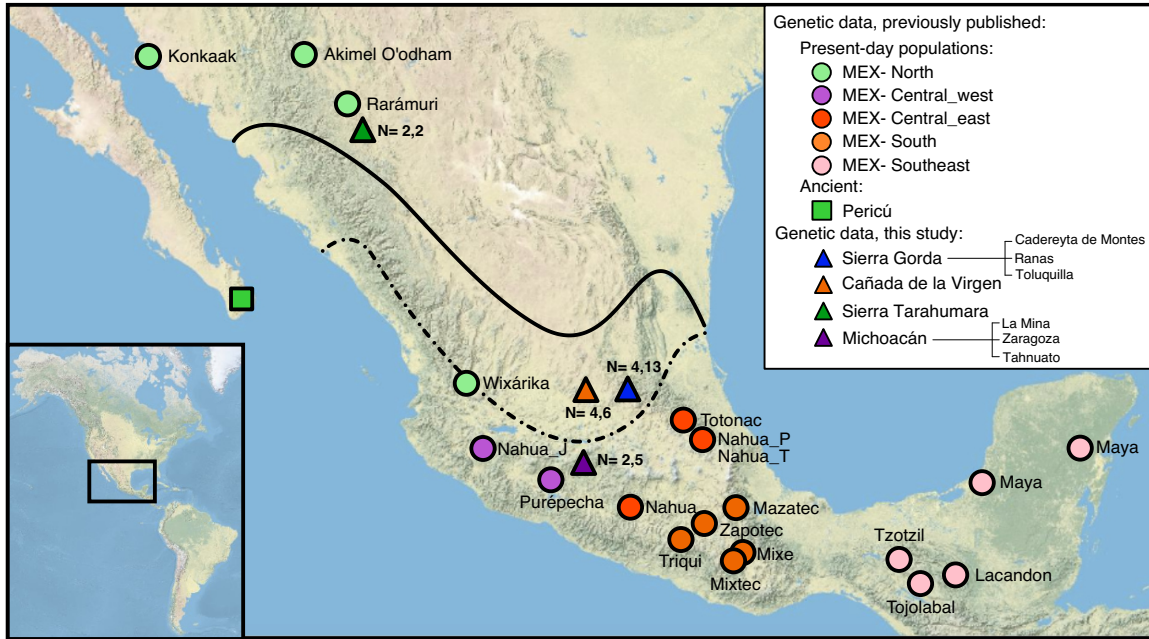
2 **Competing interests:** The authors declare no competing interests.
3

4 **Data and materials availability:** Data will be available from the European Nucleotide
5 Archive under the project accession PRJEB51440, upon publication.
6

7 **Ethics statement:** See supplementary material.
8
9
0
1
2
3
4

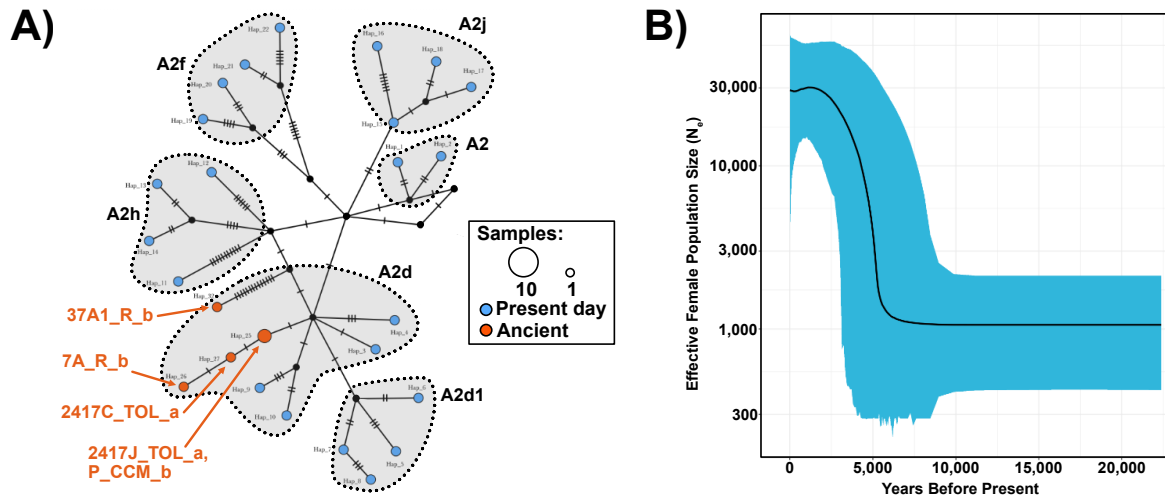
5
6
7
8
9
0
1
2

Figures and Tables



3
4
5
6
7
8
9
0
1
2
3
4
5

Fig.1. Map of Mexico indicating the location of the pre-Hispanic sites analyzed here and in previous studies (14, 15) and, the approximate locations of present-day Indigenous populations sampled in previous studies (14, 50, 54, 110). The continuous line indicates the border between Aridoamerica (northern) and Mesoamerica (central and southern) during the ninth century, while dashed line indicates this border in the sixteenth century 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).



6
7
8
9
0
1
2
3
4
5
6
7
8
9

Fig. 2. Haplotype network analysis for haplogroup A including ancient and modern mtDNA, and Extended Bayesian Skyline Plot (EBSP) for all present-day mtDNA haplogroups from Mexico. A) Haplotype network of mitochondrial haplogroup A including present-day and ancient mitochondrial haplotypes from Mexico. Individuals from SG within haplogroup A2d cluster together independently of their period. B) EBSP for all present-day mtDNA haplogroups from Mexico. EBSP shows a constant population size after the peopling of the Americas, followed by a population expansion 15,000 years ago. After that, it shows a population decline starting slowly at 2,000 years ago and accelerating at 0.5 years ago when contact with Europeans occurred. The Median is shown as a black line, and the credibility interval of 95% is shown in blue.

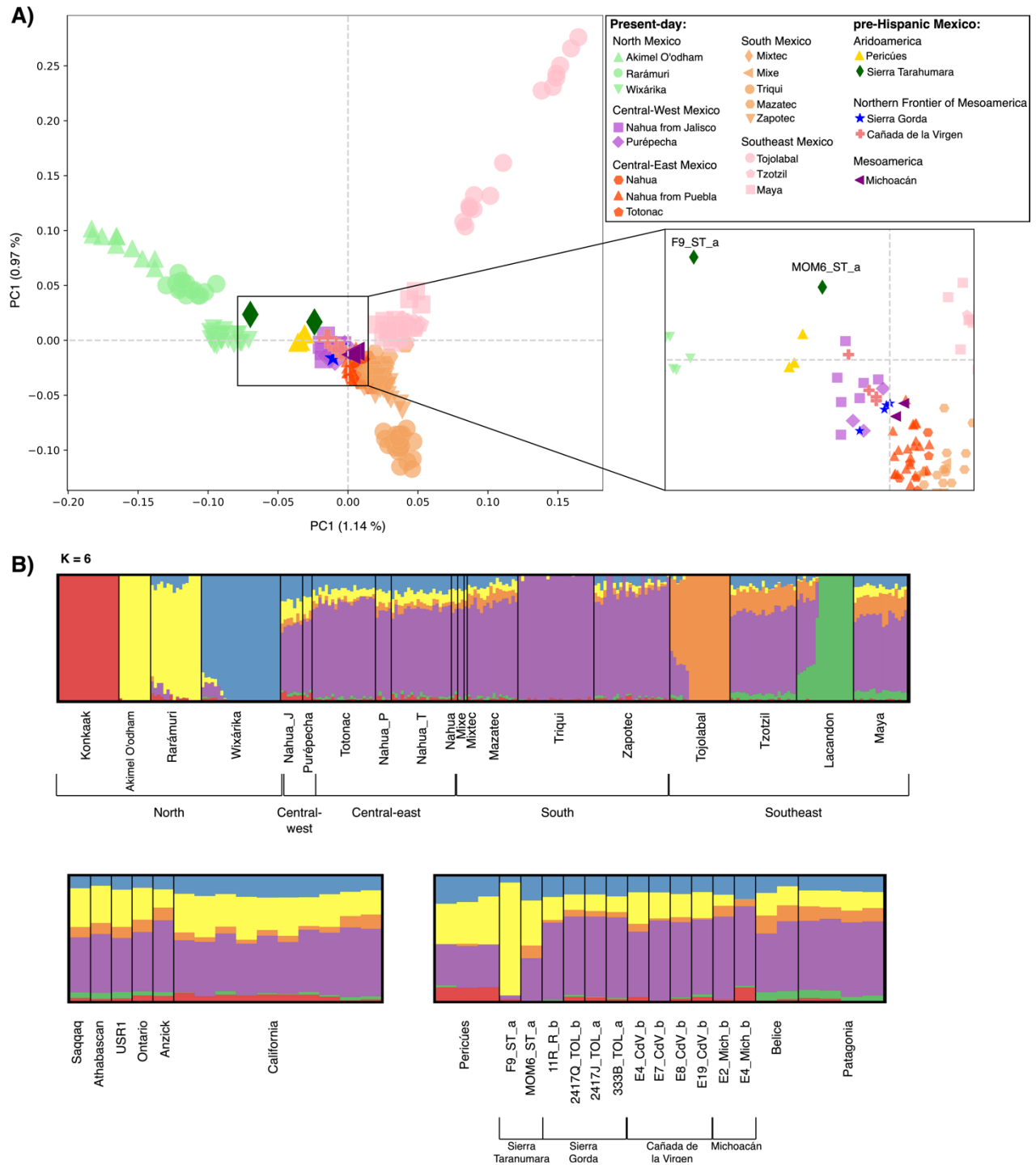


Fig.3. PCA and Admixture plots. a) Principal Component Analysis of the reference panel of present-day Indigenous Populations (see methods), the genomic data of ancient individuals from the Americas previously published and the pre-Hispanic individuals from Mexico were projected onto the panel with lsqproject from the eigensoft package version 6.0.1.1. The northern individual F9_ST_a clusters near current northern populations and the individual MOM6_ST_a clusters with central-west populations as the pre-Hispanic individuals from central Mexico. Pre-Hispanic individuals SG and CdV cluster close to present-day central-west indigenous populations, and Michoacán individuals cluster closely to central-east indigenous populations. b) Unsupervised clustering analysis using ADMIXTURE for k=6, showing component red, yellow and blue in northern populations, component green

and orange in southeast populations and component purple in south and central Mexico populations. Individuals from the Sierra Tarahumara and Pericúes, from northern Mexico have a bigger proportion of the yellow component shared with present-day Akimel O’odham and Rarámuri from northern Mexico. Prehispanic individuals from central Mexico have similar ancestral components as present-day central Mexico populations. Prehispanic individuals from the Sierra Gorda show similar genetic composition between them.

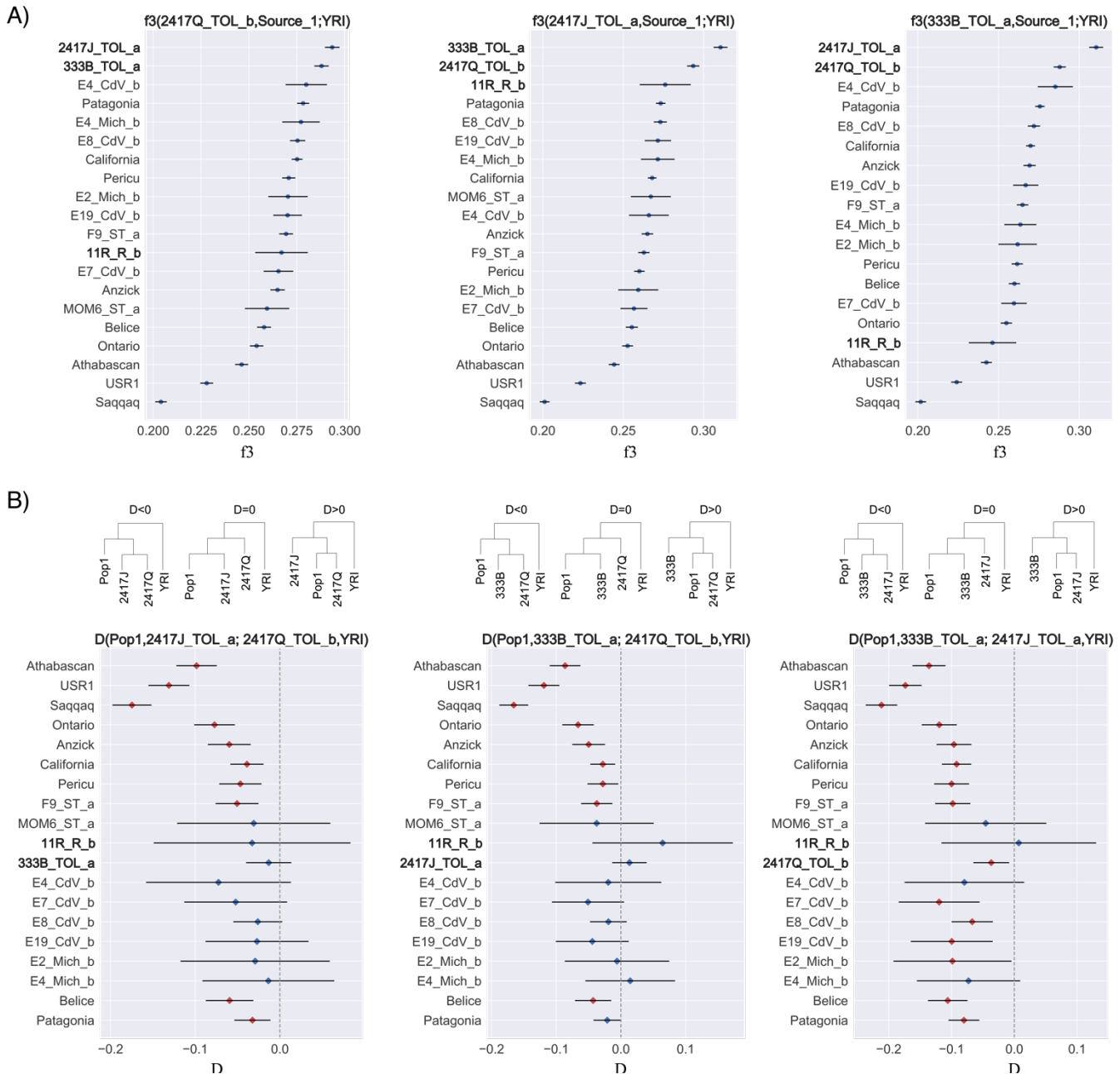
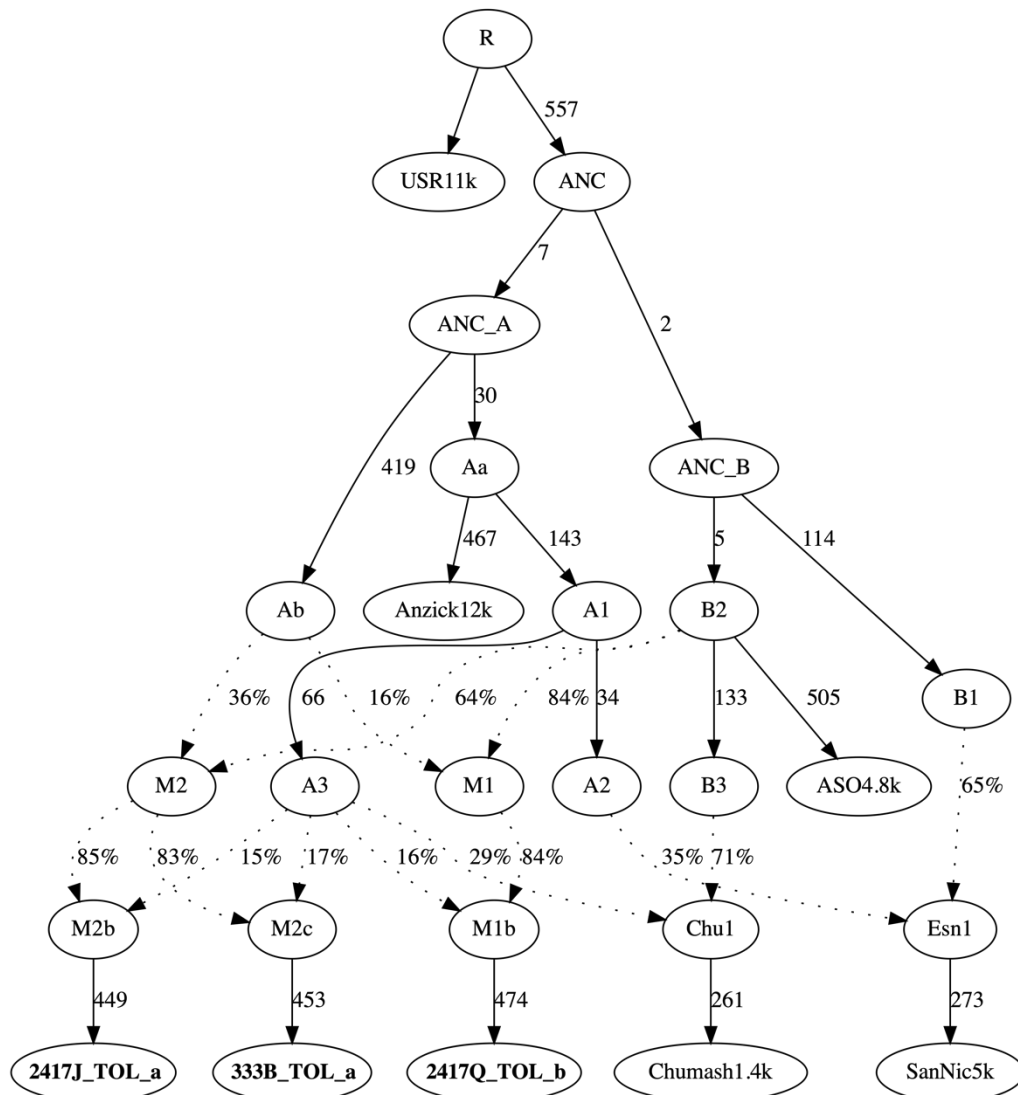


Fig.4. Population continuity in Toluquilla. A) Outgroup f_3 statistics for pre-Hispanic individuals from the Sierra Gorda, 2417Q_TOL_b, 2417J_TOL_a, and 333B_TOL_a compared with pre-Hispanic individuals from Mexico and the Americas. Higher values of f_3 indicate higher shared genetic drift. Point estimates and one standard error are shown. Individuals 2417Q_TOL_b,

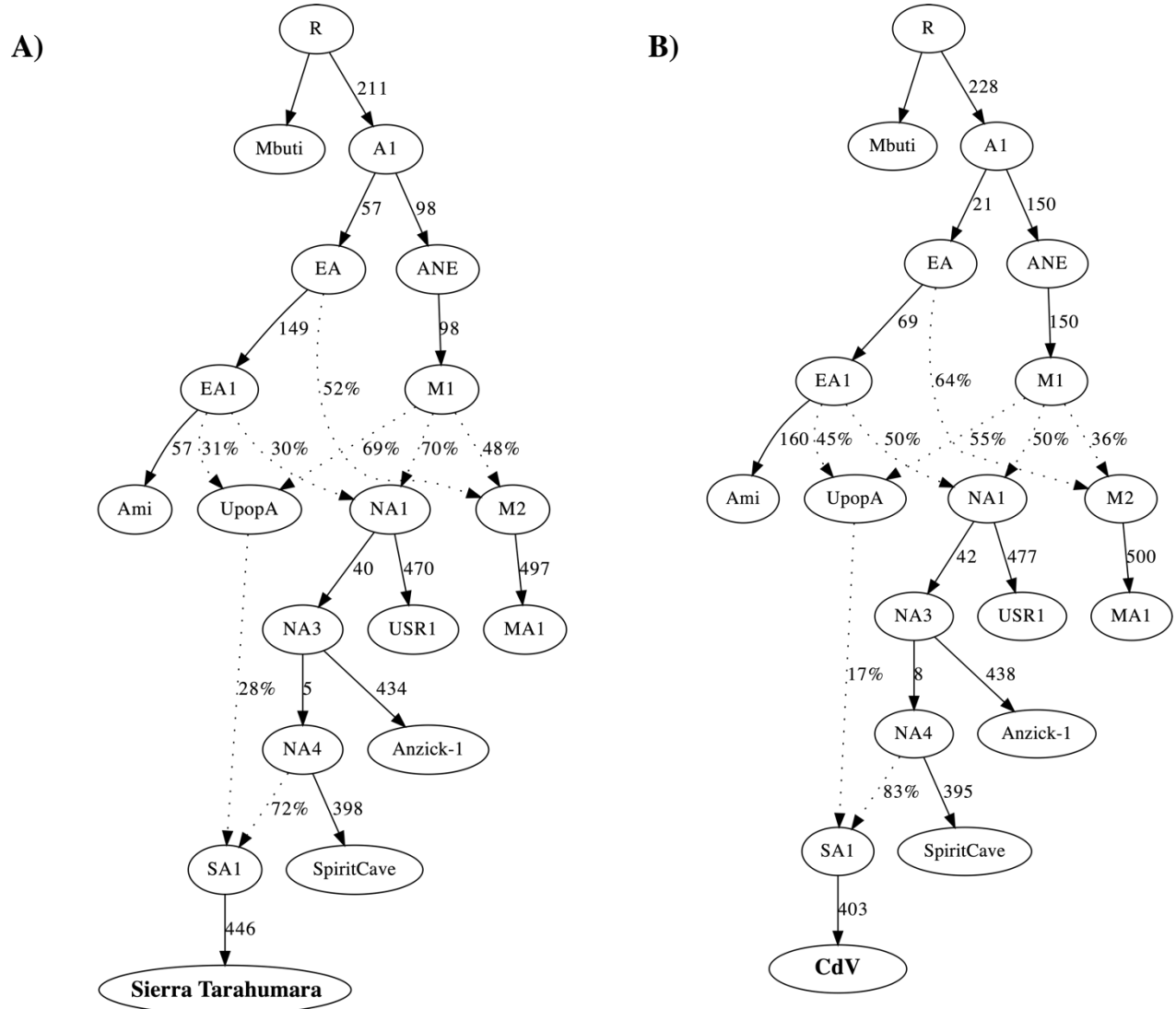
9 2417J_TOL_a, and 333B_TOL_a share higher genetic drift between them than with any other
 0 ancient Native American. B) D statistics analysis in the form
 1 $D(\text{Pop1}, \text{Toluquilla1}; \text{Toluquilla2}; \text{YRI})$. Pop1 is any of the other pre-Hispanic individuals from
 2 Mexico or America and shown in the Y axis. YRI corresponds to individuals from Africa from the
 3 1000 Genomes Project (55, 109). Expected tree topologies according to different D values are
 4 drawn on the top of the plot. Individual IDs in the trees are indicated with no suffixes. Red dots
 5 indicate significant deviations from $D=0$ ($|Z|>3$). When Toluquilla1 is the pre longstanding drought
 6 individual 2417Q_TOL_b, and Toluquilla2 is one of the post drought individuals (2417J_TOL_a
 7 or 333B_TOL_a), all Toluquilla individuals are equally related to the rest of pre-Hispanic
 8 individuals from central Mexico ($D=0$), but show a higher genetic relation when Pop1 is a
 9 population from northern Mexico. When Toluquilla1 and Toluquilla2 are both post drought
 0 individual, they tend to be more related to each other than to any other pre-Hispanic individual
 1 ($D<0$). Individuals from the same archaeological site (Sierra Gorda) are highlighted in bold.



7
 8
 9 **Fig.5.** Ancestry proportions for the individuals from Toluquilla (Sierra Gorda) before and
 0 after the megadrought analyzed through qpGraph. All individuals show a similar

1
2
3
4
5
6
7
8

percentage of ancestry from A3. The individual 2417Q_TOL_b seems to have a higher contribution of the ancestral B branch from the ancestor M1, than the individuals after the megadrought who had the ancestor M2. All f statistics are within 1.75 standard error. The tree has no inner 0 branches and obtained a |Z-score| of 1.757.



9
0
1
2
3
4
5
6
7

Fig.6. Ghost population in pre-Hispanic individuals from Mexico. A) qpGraph model for F9_ST_b. The tree has no inner 0 branches and obtained a |Z-score| of 2.142. B) qpGraph model for individuals from Cañada de la Virgen. The tree has no inner 0 branches and obtained a |Z-score| of 1.655. Models show different proportions of the ghost population UpopA in the pre-Hispanic individuals from Mexico, being 28% of UpopA for the Sierra Tarahumara individual and 17% of UpopA for CdV individuals. All f statistics are within 1.75 standard error.