Whole mitogenome analysis highlights demographic history and shared connections among distal Indigenous groups of Mexico

Complete mitogenome sequencing from 60 Mexican Native American groups

Marlen Flores-Huacuja[1], Meradeth Snow[2], Jazmín Ramos-Madrigal[3], Cecilia Contreras-Cubas[1], Francisco Barajas-Olmos[1], Angélica González-Oliver[4], Elvia Mendoza-Caamal[1], Isabel Cicerón-Arellano[1], Federico Centeno-Cruz[1], Emilio J. Córdova[1], Paulina Baca[1], Silvia Esperanza Flores-Martínez[5], Rocío Ortiz-López[6,7], Austin W. Reynolds[8], Aleksandar David Kostic[9], José Rafael Villafan-Bernal[1], Carlos Galaviz-Hernández[10], Alejandra Guadalupe García-Zapién[11], Haydee Miranda-Ortíz[1], Blanca Patricia Lazalde-Ramos[12], Francisco Loeza-Becerra[13], Alessandra Carnevale[1], Héctor Rangel-Villalobos[14], Martha Sosa-Macías[10], Augusto Rojas-Martinez[6,7], Angélica Martínez-Hernández[1], Humberto García-Ortiz[1*] & Lorena Orozco[1*]

1. Instituto Nacional de Medicina Genómica, Tlalpan, Mexico City, Mexico.
2. Department of Anthropology, University of Montana, Missoula, MT, USA.
3. Center for Evolutionary Hologenomics. Section for HoloGenomics, The Globe Institute, University of Copenhagen, Øster Farimagsgade 5A, 1352, Copenhagen, Denmark.
4. Departamento de Biología Celular, Facultad de Ciencias, Universidad Nacional Autonóma de México, Mexico City, Mexico.
5. División de Medicina Molecular, Centro de Investigación Biomédica de Occidente, Instituto Mexicano del Seguro Social (IMSS), Guadalajara, Mexico.
6. Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud and Insitute for Obesity Research. Monterrey, Mexico.
7. Universidad Autonoma de Nuevo Leon, Centro de Investigacion y Desarrollo en Ciencias de la Salud. Monterrey, Mexico.
8. Department of Anthropology, Baylor University, Waco, TX, USA.
9. Harvard Medical School, Joslin Diabetes Center, Boston, MA, USA.
10. Instituto Politécnico Nacional, CIIDIR-Durango, Durango, Mexico.
11. Departamento de Farmacobiología, Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara, Mexico.
12. Unidad Académica de Ciencias Químicas, Universidad Autónoma de Zacatecas, Zacatecas, Mexico.
13. Universidad Michoacana de San Nicolás de Hidalgo, Michoacán, Mexico.
14. Instituto de Investigación en Genética Molecular, Universidad de Guadalajara Ocotlán, Jalisco, Mexico.

*Corresponding authors
E-mail: hgarcia@inmegen.gob.mx (HGO) & lorozco@inmegen.gob.mx (LO)

Abstract

The study of mitochondrial DNA is a valuable tool to delve into the demographic history of human populations. Particularly in the Americas, five widespread Native American specific mitochondrial lineages have been identified. Here we included the complete mitogenome sequencing of 572 Indigenous individuals belonging to 60 populations spanning the Mexican territory. Our results show a great diversity of matrilineages widespread across the country, revealing shared mtDNA haplogroups in populations from distant regions. We identified all the five main Native American haplogroups clades, including 83 different subhaplogroups, from which nine are novel. The most frequent of the novel haplogroups was A2+64. A phylogenetic inference suggests that A2+64 comes from an ancestral maternal lineage that spread into the Caribbean islands. Additionally, a demographic reconstruction from whole mitogenomes showed an exponential increase in female Ne around 10 Ka ago in all the tested regions. All these findings suggest a genetic persistence through Mexico and possibly the Americas, in agreement with the model of the Mesoamerican-related expansion into the Caribbean and South America.

Introduction

Mitochondrial DNA (mtDNA) is an extranuclear 16.6 kb double-stranded DNA encoding for 37 genes. It is maternally inherited without recombination, possesses a higher mutational rate in comparison with the nuclear genome and it has a high copy number per cell [1]. Due to a process of random genetic drift and selection, several specific mtDNA lineages are widespread and fixed among human populations, making it a valuable tool to elucidate migration routes, human populations distribution patterns and evolutionary reconstructions [1,2].

The study of Native American mitochondrial lineages has contributed to elucidate how the complex demographic history, repeated and rapid populations movements throughout the continent, bottlenecks, and social conflicts have shaped the genetic structure of past and present day Native American populations [3]. To date, six specific Native American mitochondrial haplogroups (A2, B2, C1, D1, D4h3a and X2a) distributed in the American continent have been described. It has been suggested that these matrilineages come from a founding population that split from Northern Eurasian lineages around 25 ka ago and remained stranded in Beringia for ~6 ka before they moved into the Americas ~16 ka ago [4–7]. The distribution and frequency of these lineages varies across the continent and within local populations [8].

To better understand the distribution of these mitochondrial haplogroups, most investigations have focused on studying mtDNA hypervariable I & II regions, or RFLP haplogroups [9–18]. These studies have also identified specific regional admixture patterns and high overall haplogroup/haplotype diversity between populations, mainly between

94 those from the North (Aridoamerica), and those from the Center and South (Mesoamerica)
95 of Mexico [15,17,19].

96 Recently, to deepen our understanding in the peopling of the Americas, whole
97 ancient and modern mitogenomes have been utilized [6]. These studies have demonstrated
98 that the demographic bottleneck that the first settlers experienced, along with their rapid
99 expansion through the continent and limited intracontinental geneflow, have resulted in a
100 marked phylogeographic populations structure, which has persisted through time [3–6]. It
101 has been suggested that the first wave of people entering into the current Mexican territory
102 brought with them the A2 and B2 haplogroups around 15-18 ka ago, followed by a series of
103 expansions that brought additional diversity with the haplogroups C1 and D1 [20]. Previous
104 studies analyzing whole mitogenomes from Mayans, Mazahuas, and Zapotecs, all of them
105 Mexican Native Americans, suggested two main paleo-groupings: i) Centro-Mesoamerican
106 which remained within Mesoamerica, and ii) Pan-American, which moved to South America
107 [19,21].
108 Although mtDNA studies have addressed the genetic structure and migration of Mexican
109 Native Americans, most of them have often been limited in scope and sampling [3]. The
110 Mexican territory encompasses 68 Native American populations with unique cultural and
111 demographic histories, such as the establishment of sedentary agriculture in Mesoamerica,
112 or the European contact [22,23]. To gain insights into the evolutionary history of the mtDNA
113 lineages in Mexican population, we analyzed the complete mitogenomes of 572 Native
114 American individuals representing 60 populations spanning the country. The significant
115 increase in the number of whole mitogenomes sequenced and number of Native American

groups presented here, along with ancient mitogenomes from the Americas make it possible to look in greater detail at the regional variation present in Mexico and too deep in the genetic relationships of these groups, identifying novel mitochondrial subhaplogroups as well as shared matrilineages that span the entire country, helping to delve into the genetic differences between populations that belong to the same linguistic family. These findings are indicative of either shared ancestral maternal lineages or potentially population movements due to different demographic phenomena. The current data also brings insights into the genetic and cultural exchange that have taken place in the Americas through time. This is the first study that incorporates a large proportion of mitogenomes in a great diversity of Native American populations.

Results

Variants Call and Haplogroup Identification

To minimize the impact of admixture, especially with Europeans, we only included individuals with a Native American ancestry greater than or equal to 90%, estimated with ADMIXTURE v1.30 [24] (S1 Fig). Mitogenome sequencing was carried out in a total of 572 Mexican Native Americans encompassing 60 populations from 75 indigenous communities belonging to the Metabolic Analysis in an Indigenous Sample (MAIS) cohort (S1 Table). Raw reads were mapped to the Cambridge reference sequence for human mitochondrial DNA [25] and variant discovery was performed using HaplotypeCaller from GATK v3.7, setting

the ploidy parameter to 1 [26,27]. Based on this analysis, 952 variants passed the quality filters: 923 SNVs and 29 indels.

Distribution and frequency of mitochondrial haplogroups

Subsequently, the identification of haplogroups was carried out using Haplogrep2 v2.1.18 [28,29]. Three non-Native American haplogroups were identified in four individuals, which were excluded from the subsequent analyses (S2 Table). In the final data set of 568 mitogenomes, we identified five Native American haplogroups A2, B2, C1, D1 and D4h3a, where the most prevalent was A2 (53.9%), followed by B2 (24.0%), C1 (17.6%), D1 (4.3%) and D4h3a (0.2%). This general pattern follows previous findings of haplogroup frequencies in Mexico [15,16,30].

To get a deeper insight into the haplogroups distribution, we compared their geographic distribution using only those population with a sample size greater than or equal to 5 individuals (532 mitogenomes distributed in 52 indigenous populations). Although the distribution of the four main haplogroups (A2, B2, C1, D1) was highly heterogeneous throughout the Mexican territory, they were present in all regions of the country, with different frequencies among regions and populations (Fig 1). For example, we observed the highest frequency of A2 haplogroup in populations from the Southeast, Center and South of Mexico (Table 1) and notably, Huave (South, n = 10), Lacandon (Southeast, n = 11) and Kaqchikel individuals (Southeast, n = 10) only carried this haplogroup (Fig 1 and S3 Table). Moreover, the B2 and D1 haplogroups were more frequent in the Northwest (40.5% and 13.5%, respectively), while the C1 in the North (36.6 %) (Table 1). We also identified that

haplogroup B2 exhibits higher frequency than A2 in eight populations from four linguistic families (Wixarika, Mexicanero, Tepehuano, Mazahua, Mazateco Puebla, Totonaco Puebla, Zapoteco and Mocho). While, the haplogroup C1 exhibits higher frequency than A2 in Seri, Purepecha and Raramuri (Fig 1 and S3 Table).

Table 1.- Haplogroups count and frequencies by geographic region (n and %).

Region	n	A2		B2		C1		D1		D4h3	
North	101	32	31.7	27	26.7	37	36.6	5	5.0	0	0.0
Northwest	37	10	27.0	15	40.5	7	18.9	5	13.5	0	0.0
Center	185	112	60.5	38	20.5	31	16.8	4	2.2	0	0.0
South	92	53	57.6	25	27.2	9	9.8	4	4.3	1	1.1
Southeast	153	103	67.3	29	19.0	15	9.8	6	3.9	0	0.0

Fig 1.- Haplogroup distribution in 52 indigenous populations. The points in the graph show the approximate location and the pie charts show the proportion of the main haplogroups in each studied population.

On the other hand, 82 mitochondrial subhaplogroups were identified, 34 related to A2, 29 to B2, 12 to C1, 7 to D1, all of them with a variable frequency ranging from 0.2-5.6 % for all branches (Table 2). Notably, we found nine subhaplogroups non-previously reported in the present-day Native American populations, five related to the A2 haplogroup, three derived from B2 haplogroup, and one from the C1 haplogroup (Table 2, bold text). Moreover, 50 additional subhaplogroups, previously reported in other Native American populations across the continent, were also identified for the first time in the examined Mexican Native American individuals. All of them had a low frequency, ranging from 0.2%

to 5.6%, among them A2+64 showed the highest frequency (5.6%), mainly in those populations related to Mixe-Zoque (19.5%, n = 41) and Mayan linguistic families (18.9%, n=132), both from the Southeast region (S4 Table). To investigate the A2+64 divergence time, we analyzed 50 Mexican Native American individuals and seven ancient samples from Puerto Rico previously reported as carrying the A2+64 haplogroup [31]. To calibrate the tree, we included eight ancient samples and five present-day individuals harboring other haplogroups. The Bayesian species calibrated tree inferred under a strict molecular clock showed that the ancient and modern A2+64 lineages fall in the same clade, suggesting the same origin for both samples (Fig 2, S2 Fig). Moreover, tip dating estimation showed that the origin of A2+64 mitochondrial lineage is between 20 and 10 ka (mean = 14,100 years). These times broadly overlap with the estimated entrance of modern humans into the Americas, suggesting that this lineage was originated with the entrance of the first settlers to the continent.

202 **Table 2. Haplogroup diversity in the data set (Haplogroup, global frequency).**
203
204

A2		B2		C1		D1		D4h3a	
A2	4.8	B2	4.8	C1	5.3	D1	1.8	D4h3a	0.2
A2+(64)	5.6	**B2+16278**	1.4	C1b10	0.2	D1d2	0.2		
A2+(64)+@153	1.6	**B2+16279**	0.2	C1b5a	0.9	D1f3	0.5		
A2+(64)+@16111	1.4	B2a	0.5	C1c	1.2	D1h1	0.4		
A2+(64)+@16112	0.2	B2a1	0.4	C1c1a	2.5	D1i2	0.5		
A2ae	0.4	B2a1a1	0.2	C1c1b	0.7	D1k	0.2		
A2af1b	0.7	B2a4	0.4	C1c2	0.4	D1m	0.7		
A2ak	0.2	B2a4a	0.2	C1c4	1.1				
A2ao	0.4	B2a4a1	1.1	C1c5	2.1				
A2c	3.2	B2a5	0.4	C1c6	0.2				
A2d	3.0	B2b	0.9	**C1d+194**	0.2				
A2f2	0.4	**B2b+152**	0.2	C1d1	2.8				
A2g	3.0	B2b2	0.4						
A2h1	1.1	B2c	0.7						
A2j	3.0	B2c1	1.6						
A2j1	0.2	B2c2	1.2						
A2l	0.2	B2c2a	0.9						
A2m	4.0	B2c2b	0.2						
A2o	0.9	B2f	0.7						
A2p	1.6	B2g1	0.5						
A2q	1.8	B2g2	0.2						
A2q1	0.5	B2k	0.4						
A2r	2.0	B2l	0.7						
A2r1	1.1	B2o	0.9						
A2s	0.5	B2s	0.2						
A2t	1.6	B2t	2.5						
A2u	2.3	B2w	0.5						
A2u1	0.5	B2x	1.2						
A2u2	0.2	B2y1	0.5						
A2v1	0.4								
A2v1+152	2.1								
A2v1a	0.5								
A2w	4.1								
A2x	0.4								

205

206 **Fig 2.- A2+64 haplogroup tip dating tree.** A phylogenetic tree was constructed with the
207 whole mitogenome sequences of modern A2+64 carriers using a Bayesian approach under
208 a strict clock model. We included ancient genomes as calibrating information for the tree.
209 Purple bar plots illustrate the 95% HPD values of estimated mean ages for the diversification
210 of A2+64 mitogenomes.
211

212 ## Inferring Evolutionary Relationships

213 The evolutionary relationship was explored in the 52 populations data set. A Nei's
214 genetic distances matrix was constructed based on complete mitogenome sequences (S5
215 Table) and a PCoA projection was used to investigate if the populations grouped according
216 to the North, Northwest, Center, South, and Southeast regions from Mexico as previously
217 described [22,32] or linguistic filiation according to INALI classification [33]. These results
218 showed that the populations do not group by region or linguistic affiliation. In contrast,
219 some northern populations are more similar to those from the Southeast. For example, the
220 Wixarika (Northwest) population is close to the Mocho (Southeast) population, and Guarijio
221 (North) is closest to Tojolabal (Southeast). Otherwise, at regional level, a specific linguistic
222 group as Otomi-Pame from the Oto-mangue family, shows a clear subdivision between the
223 closely related Matlatzinca and Mazahua and the Otomi and Pame. Altogether, suggest that
224 they share a common ancestor or high levels of gene flow have shaped this genetic structure
225 (Fig 3 and S6 Table).
226
227

228 **Fig 3. Principal Coordinate Analysis.** Graph depicting the matrix of Nei's genetic distances
229 based on the variable sites of the mitochondrial genome. 52 indigenous populations from
230 Mexico are shown divided by region. Colors indicate the linguistic family and shapes the
231 geographic region in which the populations were classified: North (squares), Northwest
232 (circles), Center (Up-pointing Triangles) South (diamonds) and Southeast (Down-pointing
233 Triangles).

234

235 Furthermore, we constructed Median Joining Haplotype Networks for each
236 haplogroup included in the 568 mitogenome data set using PopArt v1.7.2 [34]. These
237 networks showed that samples clustered according to their haplogroup regardless of the
238 geographic region. In general, this pattern is consistent with a prior lineage expansion
239 followed by a local differentiation of the derived haplotypes and pursued by gene flow
240 between different regions (Fig 4). Moreover, a Mantel test showed no correlation between
241 matrilineal population relationships and geographic distance ($r^2 = 0.0167$ and $p = 0.081$; S3
242 Fig, S5 and S7 Tables).

243

244

245 **Fig 4.- Median Joining Haplotype Networks of whole sequenced mtDNA.** (A) A2
246 haplogroup (B) B2 Haplogroup (C) C1 Haplogroup (D) D1 Haplogroup.

247

248

249

Estimating Past Population Dynamics

We estimated the past population dynamics and the effective population size (Ne) through time in the set of 568 individuals, using Bayesian Skyline plots (BSP). The runs were broken down by the five main geographical patterns from Mexico. These demographic reconstructions showed a bottleneck resulting in a decline in female Ne in the last 20 generations in all regions, in accordance with previous observations [22,35,36]. Interestingly, the BSP showed an increase of female Ne around 10 ka that was constant until the bottleneck in all the studied regions, suggesting a population expansion of all clades in the early settlement of the present-day Mexican territory (Fig 5).

Fig 5. Extended Bayesian skyline plot of female effective population size with calibrating information from ancient sequences. Populations were grouped according to the geographic region. (A) North; (B) Northwest; (C) Center; (D) South; (E) Southeast

Discussion

Since pre-Hispanic times, different migrations, wars, conquests, territorial removals, and displacements occurred between the numerous Mexican populations. Each of these events translates into bottlenecks, genetic drift, and gene flow that have affected the distribution of the maternal haplogroups of current populations. Several studies have

271 shown that many Mesoamerican populations exhibit the haplogroup pattern ABCD,
272 according to their relative frequency. The extensive geographic sampling included here
273 shows that Native American populations matrilineages have persisted over time with a high
274 haplogroup diversity that vary across Mexico and the Americas. Herein, we identified 5
275 haplogroups (A2, B2, C1, D1 and D4h3a) following the expected frequency pattern.
276 However, several populations exhibited haplogroups with higher frequency than A2 (Fig 1)
277 The geographic location of individuals showing higher frequencies of B2 or C1 than A2,
278 suggests that A2, B2 and C1 were dispersed mainly by a Pacific coast route and were spread
279 latitudinally 800 km along the Neovolcanic axis region, then southward by the Pacific coast
280 rout and to Yucatan Peninsula via Gulf coast route. Two important rivers, Lerma and Balsas,
281 originate in the Neovolcanic axis which acts as natural barriers to inclement weather from
282 the Pacific, could have also contributed to spreading of the haplogroups (S3 Table).

283 In addition, 82 different subhaplogroups were also identified, all of them related to
284 the main four haplogroups (A2, B2, C1, D1). From these, 24 subhaplogroups were previously
285 described in the Mexican population, here we report 58 found for the first time in present
286 day Mexican Natives Americans (Table 2). The most frequent among them was the A2+64,
287 which was found with the highest frequency in the populations from the Southeast region
288 belonging to the Mayan (18.9%) and Mixe-Zoque (19.7%) linguistic families. This
289 subhaplogroup was also present in other linguistic families, like the Oto-Mangue (4.1%)
290 from the central region of the country (S4 Table). The tip dating analysis showed that the
291 frequency of this maternal lineage increased in some southern Mexican groups between 28
292 and 16 ka (Fig 2). This haplogroup was previously identified in Pre-Columbian individuals

(7%) from Cholula archaeological site [37] and in 7 of 45 (15.6%) ancient settlers of Puerto Rico before the European conquest [31]. The Bayesian inference showed a common origin from those individuals and our modern samples carrying the A2+64 haplogroup (Fig 2), suggesting that it has an origin in an ancestral lineage related to the early settlement of the Americas that was subsequently dispersed into the Caribbean islands through a coastal and/or marine route. Perhaps, the presence of this subhaplogroup in the Center of Mexico is also due to the gene flow between populations from both regions, as previously documented by anthropological and nuclear DNA studies [22,38–42].

Regarding D4h3a, which is a one of the founding pan-American subhaplogroups, it has been present in the ~12.8 ka old Anzick-1 child from the Clovis burial and the ~10 ka old Sumidouro5 from Lagoa Santa, Brazil [43,44]. This haplogroup has been previously found in low frequency in some Native American groups from the Center and South of Mexico and other regions of America, mainly distributed over the Pacific Coast [18,21,45]. Actually, we identified this haplogroup only in a Tacuate female from the Pacific Coast in the Southern region.

In addition, our results show the presence of all four main mitochondrial haplogroup clades without a clear geographic or linguistic structure in all the tested Mexican indigenous groups. Gene flow among populations was evidenced by PCoA and Median Joining Haplotype Networks, in which populations from different regions are closer to each other and share the same matrilineages (Figs 3 and 4). This pattern can be attributed to the constant movements among populations from different regions [22,38]. For example, there is evidence that in prehispanic times, trade contributed to shape the relations between

315 populations, since there were several sites of commercial exchange across the country
316 [38,39,46]. Additionally, the spread of agriculture could have altered the dynamics and the
317 genetic landscape of populations by the dispersal of different haplotypes, mainly B2a5,
318 which emerged approximately 5 ka ago as a lineage derived from B2a [20]. However, at
319 regional level interesting genetic relationships can be found as the subdivision of Oto-
320 mangue speaking populations in Central Mexico in two clearly different groups, one
321 including the Matlatzinca and Mazahua, and the other the Otomi and Pame (Fig 3). The
322 Otomi-Pame linguistic group include only five populations, four were here analyzed. Their
323 linguistic affinity supports a distant common ancestry but separated early in two groups and
324 probably admixed with other populations before migrating to settle in Central Mexico [18].
325 This genetic difference subsisted despite their geographical proximity due to their marriage
326 customs related with the patrilocal land inheritance and a late urbanization in the region.
327 Another possibility is that the lack of haplogroup structure is also due to shared ancestral
328 lineages, which first entered and spread throughout Mexico before the establishment and
329 differentiation of these populations. In line with this hypothesis, our Bayesian demographic
330 reconstruction shows an exponential increase in female Ne around 10 Ka ago in all regions
331 tested at the same time (Fig 5). This is in accordance with recent studies that show that the
332 early peopling of the Americas was characterized by a rapid dispersal and early
333 diversification of the first settlers through the continent, which in combination with the
334 effect of distance, geographic, and social barriers led to complex population histories
335 [44,47]. Altogether, these data suggest that the indigenous populations were rapidly
336 dispersed through the Mexican territory allowing the distribution of all haplogroups

thorough Mexico at different proportions, before linguistic variation was firmly established, and that in conjunction with more recent expansions and contractions of large-scale societies [15], shaped the mitochondrial haplogroups landscape of Mexican Indigenous populations. Otherwise, the presence of ancient Native American haplogroups in the Mexican Indigenous individuals tested here, can be an indicative of genetic persistence of maternal lineages through the continent, in agreement with the model of the Mesoamerican-related expansion into the Caribbean and South America [3,6,7,18,21].

In brief, the mitogenomes reported here fill gaps about the mitochondrial haplogroups distribution, and our demographic inferences showed that the American Indigenous population's history was marked by admixture between the populations which allowed the spread of the main haplogroups in the region, potentially predating linguistic diversity.

Materials and Methods

Sample description

The samples included in this study belong to the Metabolic Analysis in an Indigenous Sample (MAIS) cohort that was described previously [22,32,48,49]. Briefly, the MAIS cohort was collected between 2011-2015, all individuals were self-recognized as Indigenous members of a specific population, and had parents and grandparents born in the same community. This study was designed in accordance with the Declaration of Helsinki and was approved by the local ethics, research and biosafety human committees of the Instituto

Nacional de Medicina Genómica (INMEGEN) in Mexico City (protocol numbers: 31/2011/I and 17/2013/I). All participants provided informed written consent. For some participants, informed consent was translated into their native language, and some individuals signed with their fingerprint.

Whole-blood samples were collected by venipuncture in tubes containing EDTA and genomic DNA was extracted using the QIAmp DNA Blood Maxi kit (Qiagen, Valencia CA, USA) according to the manufacturer's protocol. The DNA concentration was determined by spectrophotometry (Nanodrop 1000 Spectophotometer).

From the MAIS cohort, we included only those Mexican Indigenous individuals with at least 90% of Native-American ancestry determined by ADMIXTURE v1.30 [25], comprising a total of 572 Native Mexicans belonging to 60 populations. To assess this, the samples were previously genotyped with Affymetrix human array 6.0 or Illumina Infinium array. Next, this set was merged with a reference population panel composed of 50 Native Americans, previously identified without evidence of recent admixture with continental populations [22] and with 50 Europeans, and 50 Africans derived from the 1000 Genomes Project Phase 3 [50]. We then ran ADMIXTURE analyses assuming K=3 clusters, including a block relaxation algorithm as the optimization method, and 100 replicates.

Mitogenome sequencing/Massively Parallel Sequencing on the Ion Torrent PGM

Mitochondrial DNA was amplified by two long-range PCR reactions using two pairs of previously reported primers [51], producing overlapping 8.2 and 8.6 kb amplicons,

380 followed by subsequent fragmentation by sonication to appropriately sized fragments using
381 ShearTM Plus Reagents Kit. The fragments were then ligated to ion barcoded adaptors, and
382 the adaptor-ligated library was then size-selected for 400-500bp length. Finally, the
383 sequencing reaction was carried out in the Ion PGMTM Systems.
384 The mtDNA data was de-multiplexed and each sample's raw reads were mapped to the
385 revised Cambridge Reference Sequence (rCRS) [25] with the BWA algorithm [52] and
386 processed with the Genome Analysis Toolkit (GATK) to recalibrate base quality-scores and
387 perform local realignment around known insertion and deletion (indels).
388

389 ## Variant Calling of mtDNA-Sequence Data/mtDNA Variant Analysis

390 Target coverage for each sample was computed with the GATK pipeline [26]. Single
391 nucleotide variants (SNVs) and indels were called with the Unified Genotype module of the
392 GATK and filtered to remove SNVs with annotations indicative of technical artifacts (such as
393 strand-bias, low variant call quality, or homopolymer runs). Variant discovery was
394 performed using HaplotypeCaller from GATK v3.7, setting the ploidy parameter to 1 and
395 next we applied hard filters to SNV and indel calls using the GATK's VariantFiltration [26,27].
396 Variants were annotated with The Ensembl Variant Effect Predictor (VEP) [53].
397

398 ## Haplogroups Identification

399 Aligned sequences were visualized and checked with MEGA7 v7.0.26 software [54]
400 prior haplogroup determination. The assignment of haplogroups for each sample was
401 assessed with the HaploGrep2 v2.1.0 software, which provides an automated way to

402 determine the haplogroup of the mtDNA profiles, based on Phylotree [28]. The classification
403 of haplogroups was based on pre-calculated phylogenetic weights that correspond to the
404 occurrence per position in Phylotree and show the stability of mutation of a variant [28].
405 Samples with non-Native American haplogroups, were excluded for subsequent analyses
406 (n=4).
407

408 Tip Dating

409 To date the A2+64 haplogroup we first transformed the genotypes from 50
410 individuals from MAIS carrying the A2+64 haplogroup into a fasta file using the
411 *FastaAlternateReferenceMaker* command from GATK v3.7 [26,27]. All fasta sequences were
412 merged into a single file and were aligned with MAFFT v7 [55]. Next, each sequence was
413 checked with MEGA7 v7.0.26. These sequences were loaded in PartitionFinder2 v2.1.1
414 [56,57] to choose the nucleotide substitution model. To calibrate the tree, we included four
415 ancient samples from Mexico carrying the B2 haplogroup, two contemporary sequences
416 from A4, two from Siberia carrying the B4 haplogroup [6] and seven ancient sequences with
417 the A2+64 haplogroup from Puerto Rico [31]. The Bayesian species calibrated tree was
418 constructed in BEAST2 v2.6.6 [58] with the following parameters: a constant population
419 model, a TN93 as nucleotide substitution model, a strict molecular clock, six partitions
420 (position by codon: 1spot, 2stpos, 3stpos; position by gene: tRNA, rRNA, HVR1-HVR2 and
421 ND6) [6], a mutation rate of $8x10^{-8}$, four chains of the MCMC algorithm for 500 million
422 iterations each, sampling every 500 steps with a burn-in of 25%. Convergence was assessed

423 based on the effective sample sizes estimates (>200) and loglikelihood distribution through
424 the run.
425

426 Distribution of mitochondrial haplogroups

427 In order to compare the haplogroup distribution in Mexican Native American, we
428 included only those populations with a sample size ≥ 5 giving a n = 532 samples distributed
429 in 52 indigenous populations. Next, a genetic distances matrix based on Nei's index were
430 constructed to establish approximate relationships among populations through a Principal
431 Coordinate analysis (PCoA) based on a dissimilarity matrix using the haplogroup
432 designations between populations. To test if mitogenome genetic and geographic distance
433 are related a Mantel test. All tests were carried out with the GenAlex v6.5 software [59].
434

435 Median Joining Haplotype Networks

436 To infer evolutionary relationships, we carried out Median Joining Haplotype
437 Networks with Population Analysis with Reticulate Trees using PopArt software with
438 standard specifications [34] in the set of 568 mitochondrial genomes. For regional
439 comparisons, we included the regional Mexican country division North, Northwest, Center,
440 South and Southeast proposed in Contreras-Cubas et al [32].
441
442
443

Demographic inference through Bayesian Skyline Plots

For the five regions, estimation of past population dynamics was carried out by Skyline Plots with BEAST2 (Bayesian Evolutionary Analysis by Sampling trees) v2.4.5 [60] in the set of 568 mitochondrial genomes. BEAST2 is based on Monte Carlo Chains Markov (MCMC) method [61,62]. A Coalescent Extended Bayesian Skyline was run using six partitions (position by codon: 1spot, 2stpos, 3stpos; position by gene: tRNA, rRNA, HVR1-HVR2 and ND6), the T93 model with a strict molecular clock and 250 millions of iterations [59], following Heled and Drummond 2008 [62] and Bouckaert et al 2014 [58]. Also, we used the sequence of Anzick-1 and five ancient sequences from Mexico, 20 modern sequences from Siberia [6], and eight Mexican sequences with African haplogroups with the purpose of calibrating the tree. Trees were plotted with R software [63].

Acknowledgements

Marlen Flores-Huacuja is a doctoral student from the Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and has received the CONAHCyT fellowship 596612. This study was supported by the Consejo Nacional de Ciencia y Tecnología, Mexico [233970]. The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease. Nature Reviews Genetics. Nature Publishing Group; 2015. pp. 530–542. doi:10.1038/nrg3966
2. Skoglund P, Reich D. A genomic view of the peopling of the Americas. Current Opinion in Genetics and Development. Elsevier Ltd; 2016. pp. 27–35. doi:10.1016/j.gde.2016.06.016
3. Llamas B, Harkins KM, Fehren-Schmitz L. Genetic studies of the peopling of the Americas: What insights do diachronic mitochondrial genome datasets provide? Quaternary International. 2017;444: 26–35. doi:10.1016/j.quaint.2017.04.040
4. Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, et al. Beringian standstill and spread of native American founders. PLoS One. 2007;2. doi:10.1371/journal.pone.0000829
5. Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, Kashani BH, et al. The initial peopling of the Americas: A growing number of founding mitochondrial genomes from Beringia. Genome Res. 2010;20: 1174–1179. doi:10.1101/gr.109231.110
6. Llamas B, Fehren-Schmitz L, Valverde G, Soubrier J, Mallick S, Rohland N, et al. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. Sci Adv. 2016;2. doi:10.1126/sciadv.1501385
7. Roca-Rada X, Souilmi Y, Teixeira JC, Llamas B. Ancient DNA studies in pre-Columbian Mesoamerica. Genes. MDPI AG; 2020. pp. 1–19. doi:10.3390/genes11111346
8. Bisso-Machado R, Fagundes NJR. Uniparental genetic markers in Native Americans: A summary of all available data from ancient and contemporary populations. Am J Phys Anthropol. 2021;176: 445–458. doi:10.1002/ajpa.24357
9. Malhi RS, Eshleman JA, Greenberg JA, Weiss DA, Shook BAS, Kaestle FA, et al. The Structure of Diversity within New World Mitochondrial DNA Haplogroups: Implications for the Prehistory of North America. Am J Hum Genet. 2002.
10. Peñaloza-Espinosa RI, Arenas-Aranda Diego, Cerda-Flores RM, Buentello-Malo Leonor, González-Valencia Gerardo, Torres Javier, et al. Characterization of mtDNA Haplogroups in 14 Mexican Indigenous Populations. Hum Biol. 2007;79: 313–320. doi:10.1353/hub.2007.0042
11. Sandoval K, Buentello-Malo L, Peñaloza-Espinosa R, Avelino H, Salas A, Calafell F, et al. Linguistic and maternal genetic diversity are not correlated in Native Mexicans. Hum Genet. 2009;126: 521–531. doi:10.1007/s00439-009-0693-y
12. Kemp BM, González-Oliver A, Malhi RS, Monroe C, Schroeder KB, McDonough J, et al. Evaluating the farming/language dispersal hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. Proc Natl Acad Sci U S A. 2010;107: 6759–6764. doi:10.1073/pnas.0905753107
13. Martínez-Cortés G, Salazar-Flores J, Haro-Guerrero J, Rubi-Castellanos R, Velarde-Félix JS, Muñoz-Valle JF, et al. Maternal admixture and population structure in Mexican-Mestizos based on mtDNA haplogroups. Am J Phys Anthropol. 2013;151: 526–537. doi:10.1002/ajpa.22293

14. Gorostiza A, Acunha-Alonzo V, Regalado-Liu L, Tirado S, Granados J, Sámano D, et al. Reconstructing the History of Mesoamerican Populations through the Study of the Mitochondrial DNA Control Region. PLoS One. 2012;7. doi:10.1371/journal.pone.0044666
15. González-Martín A, Gorostiza A, Regalado-Liu L, Arroyo-Peña S, Tirado S, Nuño-Arana I, et al. Demographic history of indigenous populations in Mesoamerica based on mtDNA sequence data. PLoS One. 2015;10. doi:10.1371/journal.pone.0131791
16. González-Sobrino BZ, Pintado-Cortina AP, Sebastián-Medina L, Morales-Mandujano F, Contreras A V., Aguilar YE, et al. Genetic Diversity and Differentiation in Urban and Indigenous Populations of Mexico: Patterns of Mitochondrial DNA and Y-Chromosome Lineages. Biodemography Soc Biol. 2016;62: 53–72. doi:10.1080/19485565.2015.1117938
17. Ochoa-Lugo MI, Muñoz M de L, Pérez-Ramírez G, Beaty KG, López-Armenta M, Cervini-Silva J, et al. Genetic affiliation of pre-hispanic and contemporary mayas through maternal linage. Hum Biol. 2016;88: 136–167. doi:10.13110/humanbiology.88.2.0136
18. González-Oliver A, Garfijias-Morales E, Smith DG, Quinto-Sánchez M. Mitochondrial DNA analysis of mazahua and otomi indigenous populations from Estado de México suggests a distant common ancestry. Hum Biol. 2017;89: 195–216. doi:10.13110/humanbiology.89.3.03
19. Mizuno F, Wang L, Sugiyama S, Kurosaki K, Granados J, Gomez-Trejo C, et al. Characterization of complete mitochondrial genomes of indigenous Mayans in Mexico. Ann Hum Biol. 2017;44: 652–658. doi:10.1080/03014460.2017.1358393
20. Achilli A, Perego UA, Lancioni H, Olivieri A, Gandini F, Kashani BH, et al. Reconciling migration models to the Americas with the variation of North American native mitogenomes. Proc Natl Acad Sci U S A. 2013;110: 14308–14313. doi:10.1073/pnas.1306290110
21. Mizuno F, Gojobori J, Wang L, Onishi K, Sugiyama S, Granados J, et al. Complete mitogenome analysis of indigenous populations in Mexico: Its relevance for the origin of Mesoamericans. J Hum Genet. 2014;59: 359–367. doi:10.1038/jhg.2014.35
22. García-Ortiz H, Barajas-Olmos F, Contreras-Cubas C, Cid-Soto MÁ, Córdova EJ, Centeno-Cruz F, et al. The genomic landscape of Mexican Indigenous populations brings insights into the peopling of the Americas. Nat Commun. 2021;12. doi:10.1038/s41467-021-26188-w
23. Quinto-Cortés CD, Arriola LA, García-Hughes G, García-López R, Molina DP, Flores M, et al. Genetic characterization of indigenous peoples from Oaxaca, Mexico, and its relation to linguistic and geographic isolation. Hum Biol. 2010;82: 409–432. doi:10.3378/027.082.0405
24. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19: 1655–1664. doi:10.1101/gr.094052.109
25. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet. 1999;23: 147–147. doi:10.1038/13779

26. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013. doi:10.1002/0471250953.bi1110s43
27. Depristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43: 491–501. doi:10.1038/ng.806
28. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt HJ, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. 2016;44: W58–W63. doi:10.1093/nar/gkw233
29. Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, et al. HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat. 2011;32: 25–32. doi:10.1002/humu.21382
30. Gojobori J, Mizuno F, Wang L, Onishi K, Granados J, Gomez-Trejo C, et al. mtDNA diversity of the Zapotec in Mexico suggests a population decline long before the first contact with Europeans. J Hum Genet. 2015;60: 557–559. doi:10.1038/jhg.2015.55
31. Nieves-Colón MA, Pestle WJ, Reynolds AW, Llamas B, De La Fuente C, Fowler K, et al. Ancient DNA Reconstructs the Genetic Legacies of Precontact Puerto Rico Communities. Mol Biol Evol. 2020;37: 611–626. doi:10.1093/molbev/msz267
32. Contreras-Cubas C, Sánchez-Hernández BE, García-Ortiz H, Martínez-Hernández A, Barajas-Olmos F, Cid M, et al. Heterogenous distribution of MTHFR gene variants among mestizos and diverse amerindian groups from Mexico. PLoS One. 2016;11. doi:10.1371/journal.pone.0163248
33. INALI. CATALOGO de las Lenguas Indígenas Nacionales: Variantes Lingüísticas de México con sus autodenominaciones y referencias geoestadísticas. Mexico; 2008.
34. Leigh JW, Bryant D. POPART: Full-feature software for haplotype network construction. Methods Ecol Evol. 2015;6: 1110–1116. doi:10.1111/2041-210X.12410
35. Lindo J, Rogers M, Mallott EK, Petzelt B, Mitchell J, Archer D, et al. Patterns of Genetic Coding Variation in a Native American Population before and after European Contact. Am J Hum Genet. 2018;102: 806–815. doi:10.1016/j.ajhg.2018.03.008
36. O'Fallon BD, Fehren-Schmitz L. Native Americans experienced a strong population bottleneck coincident with European contact. Proceedings of the National Academy of Sciences. 2011;108: 20444–20448. doi:10.1073/pnas.1112563108
37. González-Oliver A, Garfias-Morales E, Bravo-López MJ, De La Cruz-Laina MI. Genetic Relationships between Mesoamerican Ancient Populations and with American Greater Southwest and Caribbean Populations Close to Mesoamerican Borders. Hum Biol. 2022. doi:10.1353/hub.2017.0108
38. Ragsdale CS, Edgar HJH. Cultural interaction and biological distance in postclassic period Mexico. Am J Phys Anthropol. 2015;157: 121–133. doi:10.1002/ajpa.22701
39. Mata-Míguez J, Overholtzer L, Rodríguez-Alegría E, Kemp BM, Bolnick DA. The genetic impact of Aztec imperialism: Ancient mitochondrial DNA evidence from Xaltocan, Mexico. Am J Phys Anthropol. 2012;149: 504–516. doi:10.1002/ajpa.22152

40. Moreno-Estrada A, Gignoux CR, Fernández-López JC, Zakharia F, Sikora M, Contreras A V., et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science (1979). 2014;344: 1280–1285. doi:10.1126/science.1251688
41. Bellwood P. Farmers, Foragers, Languages, Genes: The Genesis of Agricultural Societies. 2002.
42. Bellwood P, Gamble C, Blanc SA Le, Pluciennik M, Richards MB, Terrell JE. First Farmers: The Origins of Agricultural Societies, by Peter Bellwood. Malden (MA): Blackwell, 2005; ISBN 0-631-20565-9. Cambridge Archaeological Journal. 2007;17: 87. doi:doi:10.1017/S0959774307000078
43. Rasmussen M, Anzick SL, Waters MR, Skoglund P, Degiorgio M, Stafford TW, et al. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature. 2014;506: 225–229. doi:10.1038/nature13025
44. Moreno-Mayar JV, Vinner L, De Barros Damgaard P, De La Fuente C, Chan J, Spence JP, et al. Early human dispersals within the Americas. Science. 2018;362. doi:10.1126/science.aav2621
45. Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, et al. Distinctive Paleo-Indian Migration Routes from Beringia Marked by Two Rare mtDNA Haplogroups. Current Biology. 2009;19: 1–8. doi:10.1016/j.cub.2008.11.058
46. Bolnick DA, Raff JA, Springs LC, Reynolds AW, Miró-Herrans AT. Native American Genomics and Population Histories. 2016. doi:10.1146/annurev-anthro-102215-100036
47. Posth C, Nakatsuka N, Lazaridis I, Skoglund P, Mallick S, Lamnidis TC, et al. Reconstructing the Deep Population History of Central and South America. Cell. 2018;175: 1185-1197.e22. doi:10.1016/j.cell.2018.10.027
48. Mendoza-Caamal EC, Barajas-Olmos F, Garciá-Ortiz H, Cicerón-Arellano I, Martínez-Hernández A, Córdova EJ, et al. Metabolic syndrome in indigenous communities in Mexico: A descriptive and cross-sectional study. BMC Public Health. 2020;20. doi:10.1186/s12889-020-8378-5
49. Cid-Soto MA, Martínez-Hernández A, García-Ortíz H, Córdova EJ, Barajas-Olmos F, Centeno-Cruz F, et al. Gene variants in AKT1, GCKR and SOCS3 are differentially associated with metabolic traits in Mexican Amerindians and Mestizos. Gene. 2018;679: 160–171. doi:10.1016/j.gene.2018.08.076
50. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526: 68–74. doi:10.1038/nature15393
51. Gunnarsdóttir ED, Li M, Bauchet M, Finstermeier K, Stoneking M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. Genome Res. 2011;21: 1–11. doi:10.1101/gr.107615.110
52. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760. doi:10.1093/bioinformatics/btp324
53. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17: 122. doi:10.1186/s13059-016-0974-4

54. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33: 1870–1874. doi:10.1093/molbev/msw054
55. Katoh K, Frith MC. Adding unaligned sequences into an existing alignment using MAFFT and LAST. Bioinformatics. 2012;28: 3144–3146. doi:10.1093/bioinformatics/bts578
56. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol. 2016; msw260. doi:10.1093/molbev/msw260
57. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst Biol. 2010;59: 307–321. doi:10.1093/sysbio/syq010
58. Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput Biol. 2014;10. doi:10.1371/journal.pcbi.1003537
59. Peakall R, Smouse PE. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012;28: 2537–2539. doi:10.1093/bioinformatics/bts460
60. Drummond AJ, Bouckaert RR. Bayesian Evolutionary Analysis with BEAST. Cambridge: Cambridge University Press; 2015. doi:10.1017/CBO9781139095112
61. Drummond AJ. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. Mol Biol Evol. 2005;22: 1185–1192. doi:10.1093/molbev/msi103
62. Heled J, Drummond AJ. Bayesian inference of population size history from multiple loci. BMC Evol Biol. 2008;8. doi:10.1186/1471-2148-8-289
63. R Core Team (2020). R: A language and environment for statistical computing. Vienna, Austria; 2020. Available: https://www.R-project.org

Supporting information

S1 Fig. Admixture inference assuming K = 3 clusters. The inference of ancestry proportions in the MAIS dataset was done using a reference panel of African, European and Native-American populations. Purple colored bars represent the proportion of inferred Native American ancestry in each sample.

675 **S2 Fig. Bayesian phylogenetic analysis of A2+64.** Calibrated tree from the whole
676 mitogenome sequences of modern and ancient A2+64 haplogroups mitogenomes

677 **S3 Fig. Mantel test.** Mantel test with 999 permutations based on Nei's genetic distances
678 using the variable sites of the mitochondrial genome and geographic distance (GGD).

679 **S1 Table. Populations size included in the present study.**

680 **S2 Table. Non-Native American Haplogroups identified in the studied populations.**

681 **S3 Table. Haplogroup count and frequency by indigenous population (n and %).**

682 **S4 Table. A2+64 Haplogroup frequency by linguistic Family and ethnic group.**

683 **S5 Table. Pairwise Nei's genetic Distance Matrix.**

684 **S6 Table. Eigen Vectors inferred from Nei's genetic distance Matrix.**

685 **S7 Table. Pairwise geographic distance Matrix.**

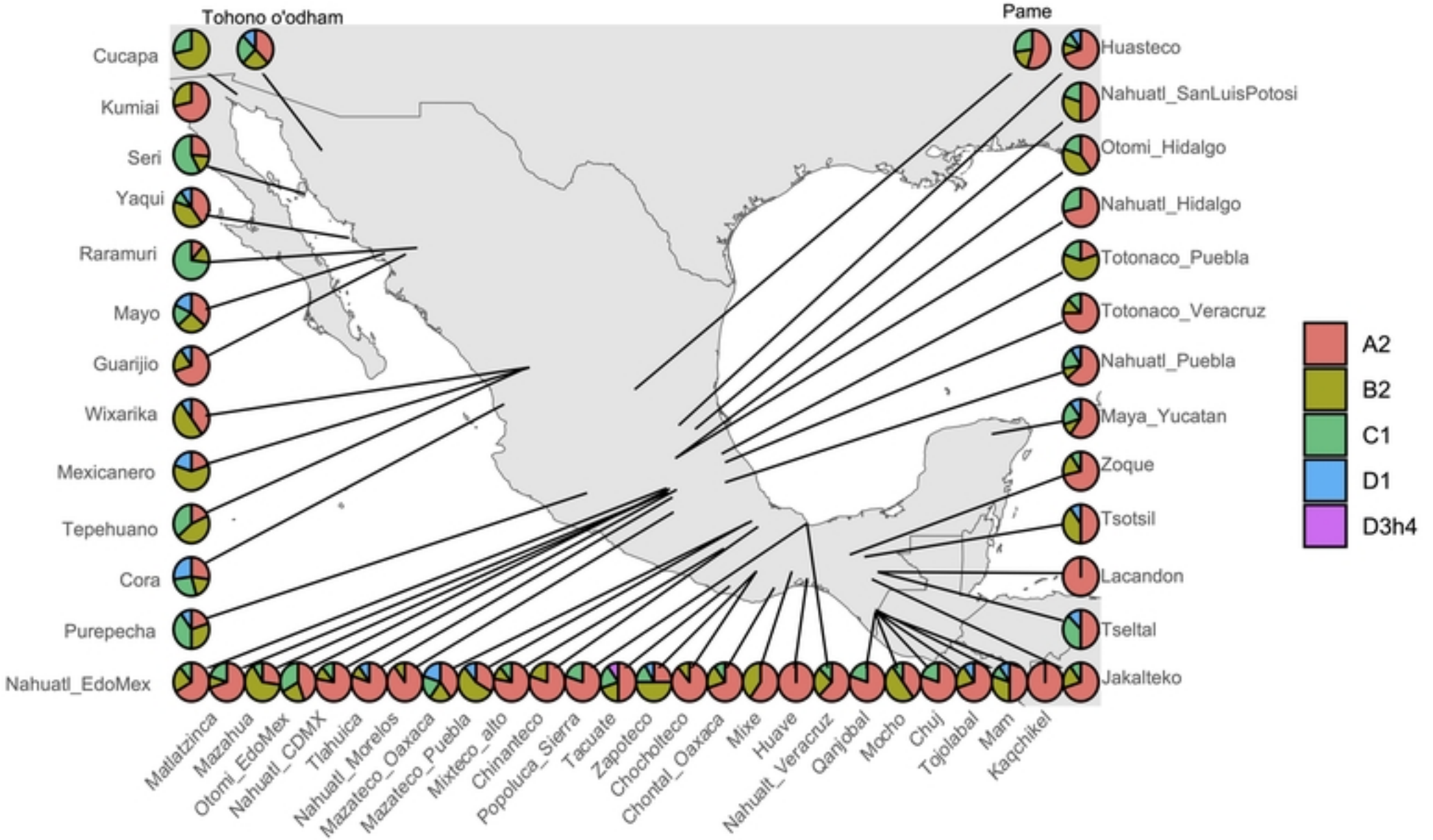

Tohono o'odham
Pame
Cucapa
Kumiai
Seri
Yaqui
Raramuri
Mayo
Guarijio
Wixarika
Mexicanero
Tepehuano
Cora
Purepecha
Nahuatl_EdoMex
Huasteco
Nahuatl_SanLuisPotosi
Otomi_Hidalgo
Nahuatl_Hidalgo
Totonaco_Puebla
Totonaco_Veracruz
Nahuatl_Puebla
Maya_Yucatan
Zoque
Tsotsil
Lacandon
Tseltal
Jakalteko
A2
B2
C1
D1
D3h4
Matlatzinca
Mazahua
Otomi_EdoMex
Nahuatl_CDMX
Tlahuica
Nahuatl_Morelos
Mazateco_Oaxaca
Mazateco_Puebla
Mixteco_alto
Chinanteco
Popoluca_Sierra
Tacuate
Zapoteco
Chocholteco
Chontal_Oaxaca
Mixe
Huave
Nahuatl_Veracruz
Qanjobal
Mocho
Chuj
Tojolabal
Mam
Kaqchikel

Figure 1

Figure 2

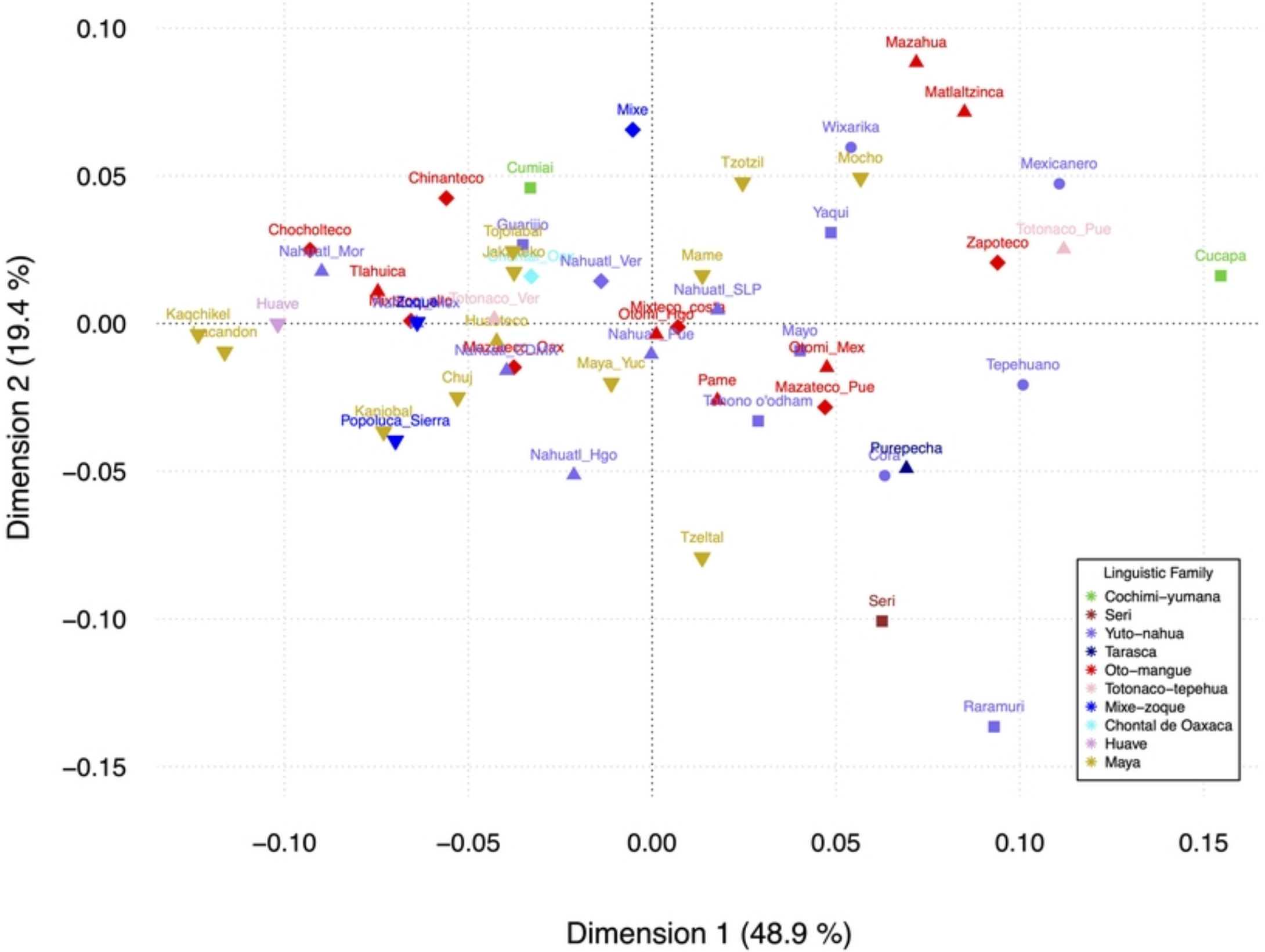

0.10
0.05
0.00
−0.05
−0.10
−0.15
Dimension 2 (19.4 %)
−0.10
−0.05
0.00
0.05
0.10
0.15
Dimension 1 (48.9 %)
Mazahua
Matlaltzinca
Mixe
Wixarika
Tzotzil
Mocho
Mexicanero
Cumiai
Chinanteco
Yaqui
Chocholteco
Totonaco_Pue
Zapoteco
Nahuatl_Mor
Tlahuica
Nahuatl_Ver
Mame
Cucapa
Totonaco_Ver
Nahuatl_SLP
Huave
Kaqchikel
Mixteco_costa
Otomi_Hgo
Mayo
Nahuatl_Pue
Otomi_Mex
Maya_Yuc
Tepehuano
Chuj
Pame
Mazateco_Pue
Kanjobal
Popoluca_Sierra
Purepecha
Nahuatl_Hgo
Tzeltal
Seri
Raramuri
Linguistic Family
Cochimi−yumana
Seri
Yuto−nahua
Tarasca
Oto−mangue
Totonaco−tepehua
Mixe−zoque
Chontal de Oaxaca
Huave
Maya

Figure 3

A)

B)

C)

D)

Figure 4

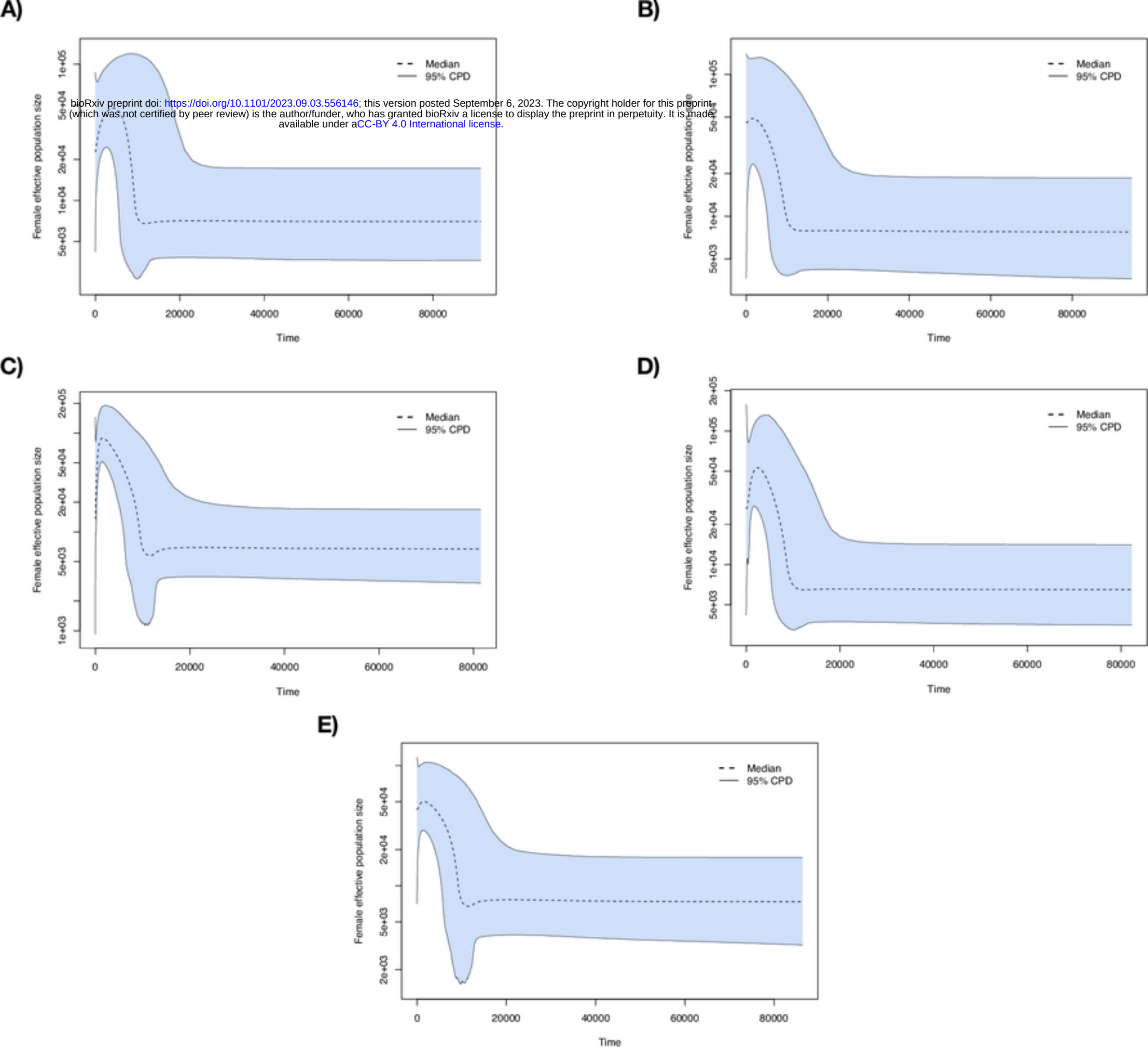
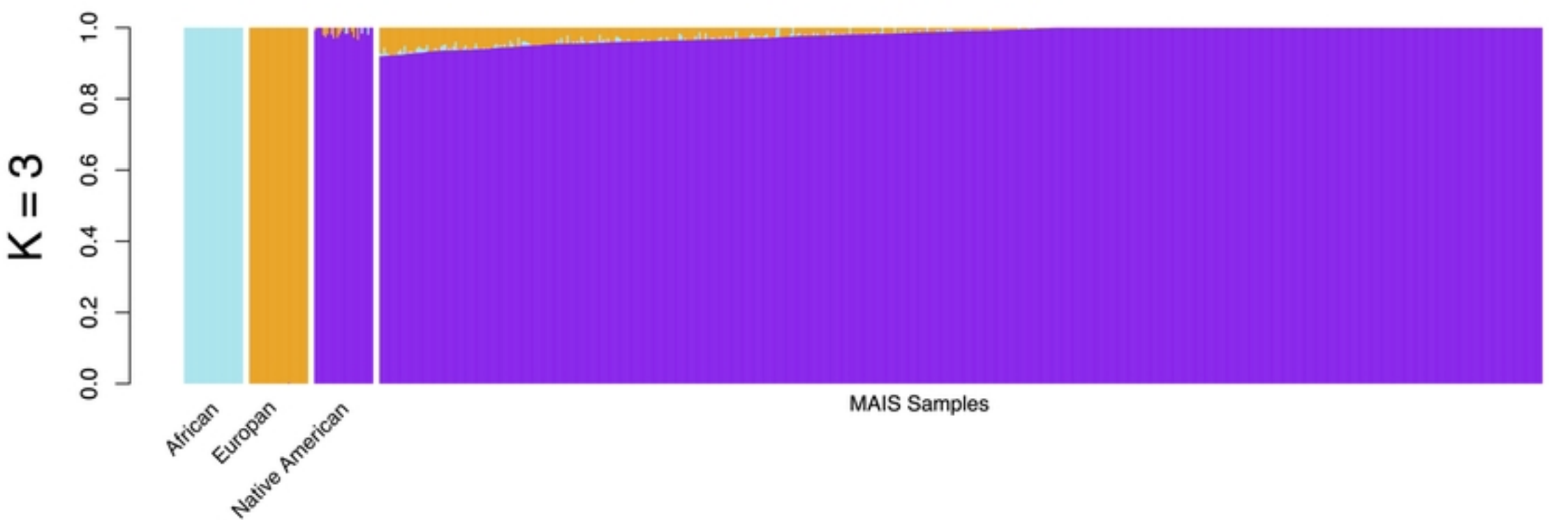
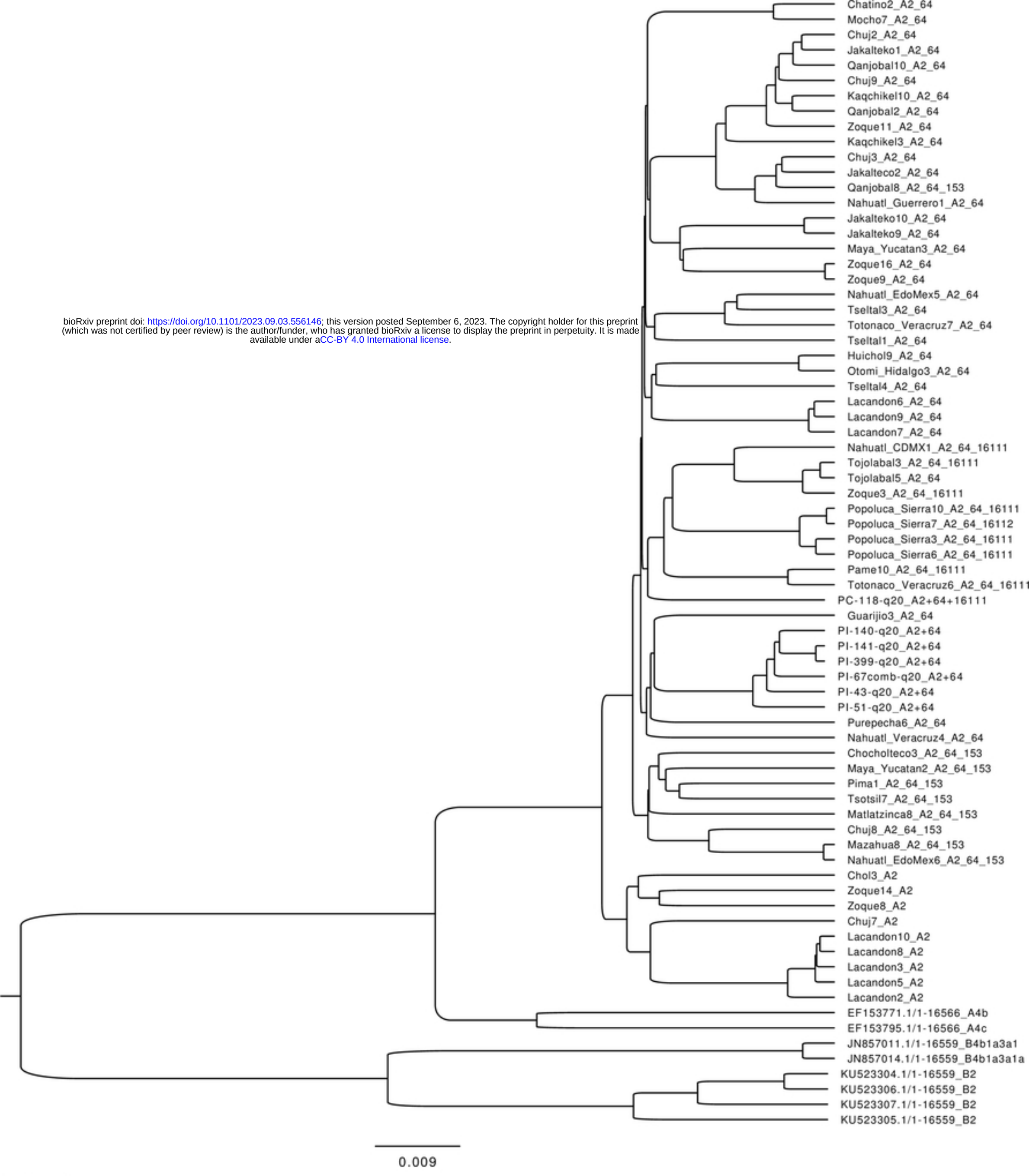


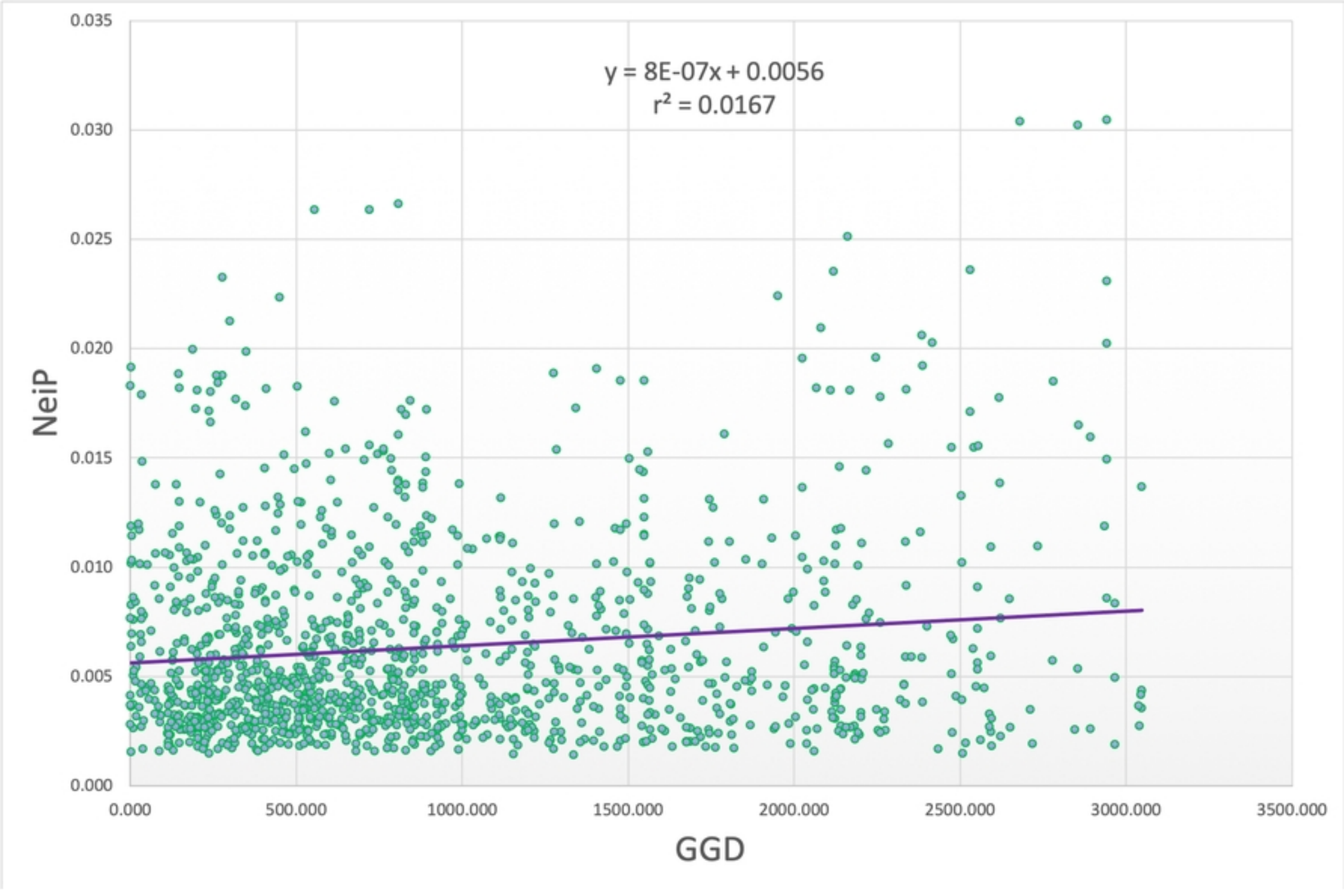
Figure 5


K = 3
0.0
0.2
0.4
0.6
0.8
1.0
African
Europan
Native American
MAIS Samples

S1 Figure



S2 Figure


y = 8E-07x + 0.0056
r² = 0.0167
NeiP
GGD
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0.000
500.000
1000.000
1500.000
2000.000
2500.000
3000.000
3500.000

S3 Figure