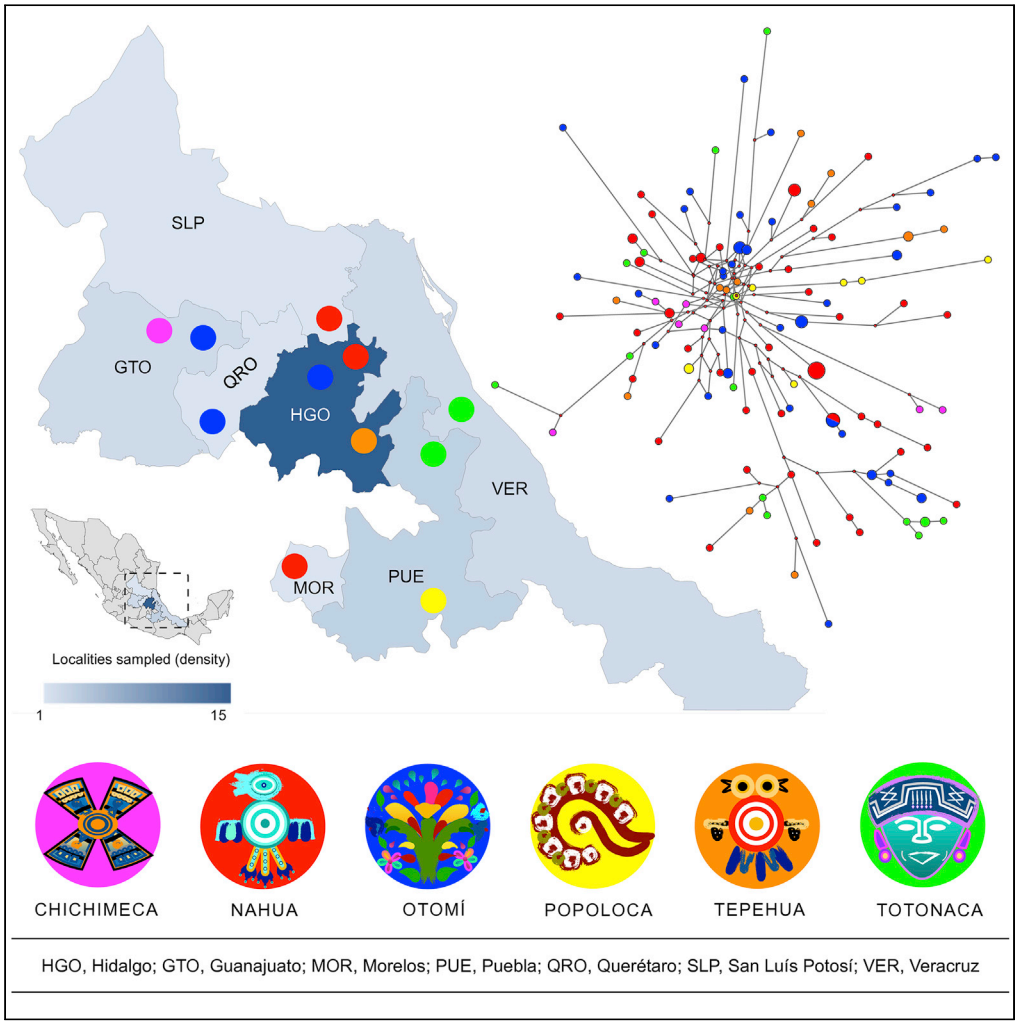
iScience

CellPress
OPEN ACCESS

Article

Y chromosome diversity in *Aztlan* descendants and its implications for the history of Central Mexico



Rocío Gómez, Miguel G. Vilar, Marco Antonio Meraz-Ríos, ..., Amanda C. Owings, Jill B. Gaieski, Theodore G. Schurr

tgschurr@sas.upenn.edu

Highlights

Enormous Y chromosome diversity observed in Native Mexican populations.

Haplogroups Q-MEH2, Q-M3, Q-Z768, Q-L663, Q-Z780, and Q-PV3 were identified.

Patterns of Y chromosome diversity not shaped by ethnicity, geography, or language.

Multiple population dispersals contributed to Y chromosome diversity in Mexico.

Gómez et al., iScience *24*, 102487
May 21, 2021
https://doi.org/10.1016/j.isci.2021.102487

iScience

CellPress
OPEN ACCESS

Article

Y chromosome diversity in *Aztlan* descendants and its implications for the history of Central Mexico

Rocío Gómez,[1] Miguel G. Vilar,[2,3] Marco Antonio Meraz-Ríos,[4] David Véliz,[5,6] Gerardo Zúñiga,[7] Esther Alhelí Hernández-Tobías,[8] Maria del Pilar Figueroa-Corona,[4] Amanda C. Owings,[2] Jill B. Gaieski,[2] Theodore G. Schurr,[2,10,*] and The Genographic Consortium[9]

SUMMARY

Native Mexican populations are crucial for understanding the genetic ancestry of Aztec descendants and coexisting ethnolinguistic groups in the Valley of Mexico and elucidating the population dynamics of the prehistoric colonization of the Americas. Mesoamerican societies were multicultural in nature and also experienced significant admixture during Spanish colonization of the region. Despite these facts, Native Mexican Y chromosome diversity has been greatly understudied. To further elucidate their genetic history, we conducted a high-resolution Y chromosome analysis with Chichimecas, Nahuas, Otomies, Popolocas, Tepehuas, and Totonacas using 19 Y-short tandem repeat and 21 single nucleotide polymorphism loci. We detected enormous paternal genetic diversity in these groups, with haplogroups Q-MEH2, Q-M3, Q-Z768, Q-L663, Q-Z780, and Q-PV3 being identified. These data affirmed the southward colonization of the Americas via Beringia and connected Native Mexicans with indigenous populations from South-Central Siberia and Canada. They also suggested that multiple population dispersals gave rise to Y chromosome diversity in these populations.

INTRODUCTION

A geographic area known for its abundant resources, the Valley of Mexico (VM) was the homeland for great ancient civilizations such as *Teotihuacan*, *Tula*, and the Aztec Empire. Before the arrival of the Aztecs, the VM was occupied by Otomies (*Hñähñús,* as they self-identify), Toltecs (the first *Nahuatl* speakers to reside there), Chichimecas (*Éza'r,* as they self-identify, singular *Úza'*), Totonacas, Tepehuas, and Popolocas, among other ethnic groups (Lastra, 2006; Matos-Moctezuma, 2012; Nichols and Rodríguez-Alegría, 2017; Soustelle, 1993). These populations experienced different migratory displacements before settling in the VM, and intermarrying with their neighbors, thereby diversifying their gene pools (Instituto Nacional de los Pueblos Indígenas, 2019; Matos-Moctezuma, 2012). The Aztecs, an amalgam of different peoples including the *Mexicas* (seven *Nahuatl*-speaking migrant tribes from *Aztlan*), *Tepanecas* (from Tlacopan and Tacuba), and *Alcohuas* (from Texcoco), also intermarried with resident populations to establish political and military alliances (Berdan et al., 1996). Through this strategy, the Aztecs obtained supreme power in Mesoamerica and, in the process, created a culturally, linguistically, and ethnically diverse empire (Nichols and Rodríguez-Alegría, 2017).

Considered to be contemporary descendants of the Aztecs, Nahuas are the largest indigenous group of Mexico. Since the 13th century CE, they have occupied the VM, a region where more than 30% of Amerindian populations currently live (Instituto Nacional de los Pueblos Indígenas, 2019). These populations are comprised by more than 7 million people who speak languages from different linguistic families, with many of them remaining geographically isolated and deeply rooted to their pre-Hispanic past (Instituto Nacional de Estadística, Geografía e Informática, 2019). Despite their present-day population sizes and sustained cultural practices, these populations lost a significant proportion of their indigenous genetic lineages during the last five centuries of European colonization and Mexican state rule (Sandoval et al., 2012). Thus high-resolution genetic analysis with single nucleotide polymorphism (SNP) and short tandem repeat (STR) loci located in the non-recombining region of the Y chromosome (NRY) is crucial for reconstructing the genetic history of Aztec descendant populations.

[1]Departamento de Toxicología, CINVESTAV-IPN, Mexico City 07360, Mexico

[2]Department of Anthropology, University of Pennsylvania, Philadelphia, PA 19104-6398, USA

[3]National Geographic Society, Washington, DC 20005, USA

[4]Departmento de Biomedicina Molecular, CINVESTAV-IPN, Mexico City 07360, Mexico

[5]Departamento de Ciencias Ecológicas, Instituto de Ecología y Biodiversidad, Facultad de Ciencias, Universidad de Chile, Santiago 7800003, Chile

[6]Núcleo Milenio de Ecología y Manejo Sustentable de Islas Oceánicas, Departamento de Biología Marina, Universidad Católica del Norte, Coquimbo 1781421, Chile

[7]Departamento de Zoología, Laboratorio de Variación Biológica y Evolución, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Mexico City 11340, Mexico

[8]Universidad Autónoma de Nuevo León, Facultad de Salud Pública y Nutrición, Monterrey, Nuevo León 64460, Mexico

[9]A complete list of Genographic Consortium members can be found in the supplemental information text

[10]Lead contact

*Correspondence: tgschurr@sas.upenn.edu

https://doi.org/10.1016/j.isci.2021.102487

Haplogroups C and Q are the main paternal lineages found in Native Americans, being derived from ancestral lineages in Native Siberian populations (Schurr and Sherry, 2004). Haplogroup Q is the most prevalent and ancient founding lineage in the New World, having diverged into at least four different sublineages over the past 20,000 years, including Q-MEH2 (Q1a), Q-M346 (Q1b), Q-L54 (Q1b1a), and Q-M3 (Q1b1a1a) (Dulik et al., 2012a, 2012b; Malyarchuk et al., 2011). Although all four sublineages have been observed in the Americas, the latter two are more common in Central and South America (Battaglia et al., 2013; Dulik et al., 2012a, 2012b; Malyarchuk et al., 2011; Regueiro et al., 2013; Schurr and Sherry, 2004). By contrast, haplogroup C3 occurs at low frequencies in indigenous North American populations (Malyarchuk et al., 2011; Pinotti et al., 2019; Roewer et al., 2013), excepting a different C3 sublineage observed in South American populations (Jota et al., 2016; Roewer et al., 2013). It derives from an ancient East Asian lineage that has expanded throughout northeast Siberia over the past several millennia (Dulik et al., 2012a, 2012b; Jota et al., 2016; Malyarchuk et al., 2010, 2012; Wei et al., 2017).

The multi-ethnic nature of contemporary Native Mexicans provides a unique opportunity to explore their genetic history, which is crucial for expanding our knowledge about the initial colonization of the New World. The genetic features of these groups have been briefly explored and consequently remain unclear (Battaglia et al., 2013; Dulik et al., 2012a, 2012b; Mata-Miguez et al., 2012; Regueiro et al., 2013; Roewer et al., 2013; Sandoval et al., 2012; Vargas-Alarcon et al., 2007). Hence, in this study, we characterized Y chromosome diversity in populations speaking languages belonging to three linguistic families (Oto-Manguean, Totonaca-Tepehua, and Uto-Aztecan) that coexisted during the Aztec empire and presently occupy the VM in an effort to better understand the formation of these ethnolinguistic groups.

Our results reveal a complex array of paternal lineages that appear to have arrived or expanded in the VM at different times. Comparison of Native Mexican data with those from populations of Asia and North, Central, and South America further reveals connections between *Aztlan* descendants and South-Central Siberians and indigenous populations from both Canada and South America. These findings affirm the southward colonization route for ancestral Native Americans and support the view that Mesoamerica played a crucial role in the diversification of indigenous Y chromosomes in the Americas.

RESULTS

Frequency and diversity of Native Mexican paternal lineages

We traced the patrilineal ancestry of 231 unrelated men belonging to Oto-Manguean-, Totonaca-Tepehua-, and Uto-Aztecan-speaking populations (Table S1). Based on the analysis of 19 STR and 21 SNP loci, 175 (75.8%) male participants exhibited Amerindian Y chromosomes and 56 (24.2%) had non-native Y chromosomes (Table S2). Our study populations also exhibited a remarkably high proportion of Native American paternal lineages relative to Mestizo populations subjected to similar analyses (Gonzalez-Sobrino et al., 2016; Lopez-Ramirez et al., 2020; Santana et al., 2014). All chromosomes were assigned to paternal lineages or haplogroups based on variants at these loci.

Of the indigenous haplogroups seen in Native Mexicans, Q-M3 was the most prominent (79.4%), followed by Q-L54 (18.3%) and then Q-MEH2 (2.3%) (Figure 1). SNP genotyping further revealed Q-M3 to contain subhaplogroup Q-Z768 (12.9%), which was present in *Éza'r*, *Hñähñús*, Nahuas, and Tepehuas, and subhaplogroup Q-L663 (0.8%), which appeared in only *Hñähñús* from Hidalgo (HGO), with the remainder (86.3%) being defined by only the Q-M3 SNP. Haplogroup Q-L54 was primarily composed by subhaplogroup Q-Z780 (92.8%), which appeared in all populations except for the Popolocas, and subhaplogroup Q-PV3 (7.1%), which occurred only in the Nahuas from San Luis Potosi (Nahua-SLP) and Tepehuas (Figure 1; Tables S1 and S2). No individuals had Y chromosomes belonging to haplogroup C3.

Overall, Y-STR haplotype diversity in Native Mexican populations was enormous, with 146 different indigenous haplotypes being observed among 175 unrelated men. Oto-Manguean- (h = 0.995), and Totonaca-Tepehua- (h = 0.994) speaking populations exhibited the greatest diversity of Y-STR haplotypes, whereas Uto-Aztecan speakers (Aztec descendants) were slightly less diverse (h = 0.989) (Table 1). Most interestingly, only 1.37% of the 146 Y-STR haplotypes were shared between populations. This sharing occurred mainly within Oto-Manguean (*Éza'r* and *Hñähñús*) groups and between Oto-Manguean- (*Hñähñús*) and Uto-Aztecan- (Nahuas) speaking groups (Table S3).

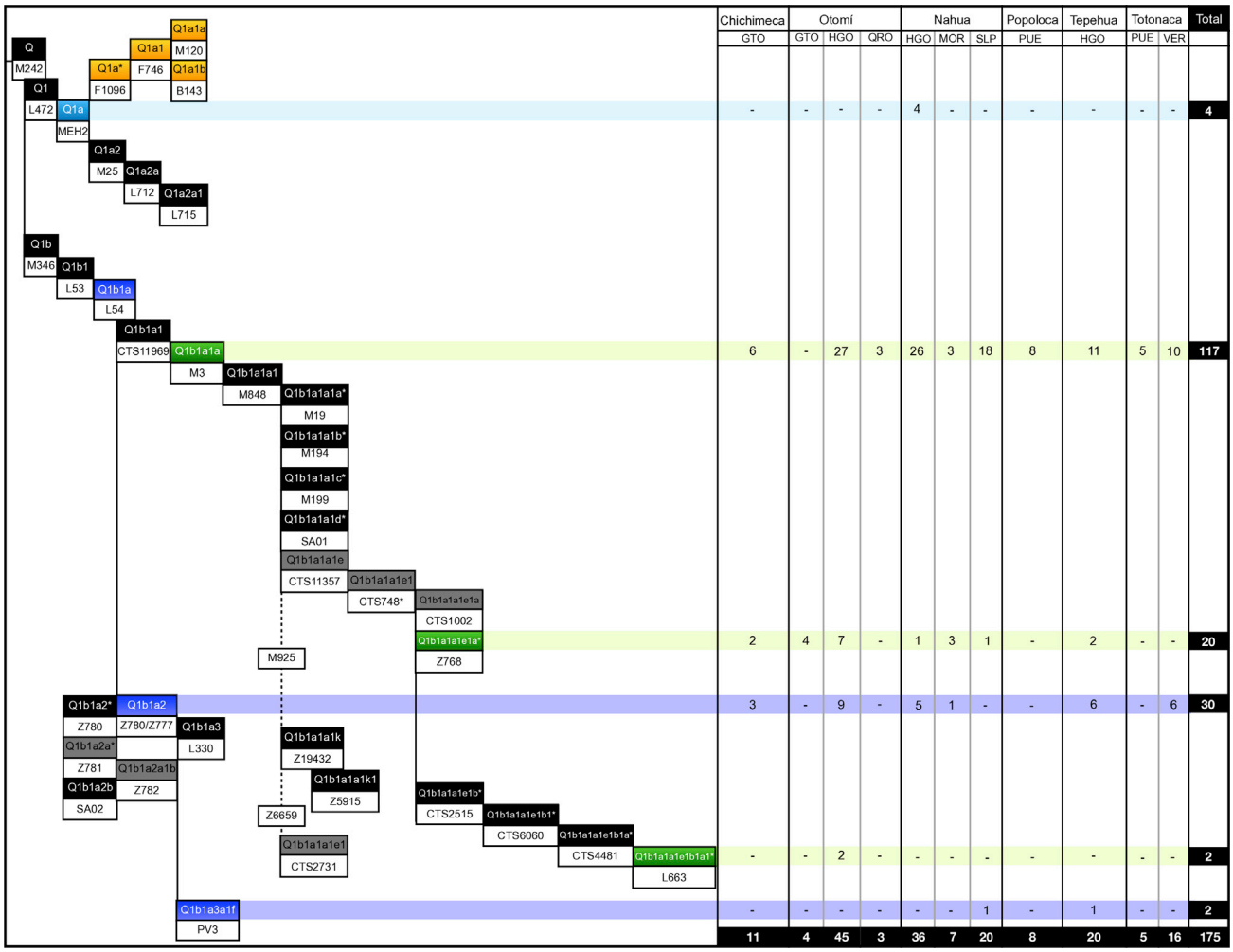


Figure 1. A phylogenetic tree of haplogroup Q

The haplogroup nomenclature follows the Y-chromosome Consortium (2002), Karafet et al., (2008) and International Society of Genetic Genealogy (2019–2020) recommendations. **Note:** The number of individuals belonging to each haplogroup in Native Mexican populations is shown on the right side in the table. The blue, green, and purple colors show the haplogroups found in the present study; the yellow color indicates the Q-F1096 sublineage within Q-MEH2. The dotted line indicates continuity of haplotypes within a specific lineage that cannot be fully represented here due to space constraints.

Statistical analysis of NRY diversity in Native Mexicans

Analysis of molecular variance

To delineate the structure of the study populations, we conducted an analysis of molecular variance (AMOVA) using different criteria (Table 2). Overall, the differences within groups regardless of the categories into which they were placed accounted for the vast majority of the diversity observed in them. Based on geographic location, these ethnic groups were categorized using two criteria. In the geography-1 criterion, variation among groups was insignificant, whereas a marginally significant difference among groups ($p = 0.085$) was found using the geography-2 criterion. On the basis of their linguistic affiliation, the AMOVA did not reveal any significant differences between populations.

To determine the influence of ethnicity on the pattern of genetic variation, an AMOVA was conducted with data from the *Hñähñús*, Nahuas, and Totonacas (Table S4). *Hñähñú* and Totonaca populations did not show any significant genetic differences among themselves. By contrast, Nahua populations exhibited small but statistically significant differences among themselves ($p = 0.009$).

Table 1. Diversity parameters for NRY haplogroups in Native Mexican populations

Population	*N*	*H*	Haplotype diversity	MPD	ρ ± σ	*Vp*
Q1a (Q-MEH2)						
Nahuas	4	2	1 ± 0.500	0	9.060 ± 1.45	0
Q1b1a1a (Q-M3)						
Éza'r	6	6	1.000 ± 0.096	8.600 ± 3.633	12.667 ± 2.444	0.533
Hñähñús	30	26	0.988 ± 0.013	8.110 ± 3.872	12.069 ± 1.915	0.561
Nahuas	47	38	0.982 ± 0.012	7.273 ± 3.467	9.953 ± 1.678	0.464
Popolocas	8	7	0.964 ± 0.077	6.250 ± 3.322	9.334 ± 2.149	0.504
Tepehuas	11	10	0.982 ± 0.046	7.891 ± 3.987	17.154 ± 3.018	0.548
Totonacas	15	14	0.990 ± 0.028	7.590 ± 3.753	12.500 ± 2.019	0.581
Overall	117	101	0.996 ± 0.002	7.850 ± 3.679	21.025 ± 3.501	0.550
Q1b1a1a1e1a* (Q-Z768)						
Éza'r	1	1	1.000	0	8.000 ± 0.500	0
Hñähñús	12	8	0.924 ± 0.057	5.015 ± 2.621	15.769 ± 3.395	0.306
Nahuas	5	4	0.900 ± 0.161	5.900 ± 3.391	12.833 ± 2.769	0.304
Tepehuas	2	2	1.000 ± 0.500	1.000 ± 1.000	7.667 ± 2.236	0.029
Overall	20	15	0.968 ± 0.025	5.326 ± 2.683	15.286 ± 3.357	0.640
Q1b1a1a1e1b1a* (Q-L633)						
Hñähñús	2	2	1	0	5.333 ± 0.943	0
Q1b1a2 (Q-Z777/Z780)						
Éza'r	3	3	1.000 ± 0.272	8.000 ± 5.127	13.250 ± 1.300	0.471
Hñähñús	9	8	0.972 ± 0.064	9.028 ± 4.593	12.400 ± 2.069	0.716
Nahuas	6	6	1.000 ± 0.096	10.267 ± 5.472	12.143 ± 1.895	0.671
Tepehuas	6	4	0.800 ± 0.172	5.534 ± 3.097	19.571 ± 3.717	0.358
Totonacas	6	5	0.934 ± 0.122	2.534 ± 1.578	17.286 ± 3.761	0.212
Overall	30	26	0.988 ± 0.013	8.689 ± 4.127	26.000 ± 4.220	0.635
Q1b1a3a1f (Q-PV3)						
Nahuas	1	1	1.000 ± 0.272	0	12.000 ± 1.225	0
Tepehuas	1	1	1.000 ± 0.272	0	9.500 ± 0.500	0
Overall	2	2	1.000 ± 0.500	6.000 ± 4.583	14.300 ± 2.963	0

N, number of samples; *h*, number of different haplotypes found; MPD, mean pairwise differences; ρ, average distance root haplotypes; σ, standard deviation for ρ; *Vp*, intrapopulation genetic variance.

Population differentiation

To further explore the pattern of patrilineal diversity in Native Mexican populations, we analyzed the Y chromosome data with an exact test of population differentiation. This test was conducting using R_{st} values estimated from 17 loci Y-STR haplotypes, with p values being adjusted for false discovery rates (Benjamin and Hochberg, 1995). The populations were categorized by geography, language family, or ethnicity for these tests, as done for the AMOVA. The resulting R_{st} estimates for Native Mexican populations were then visualized in a multidimensional scaling (MDS) plot.

Using the geography-1 criterion, we observed a clear separation between the South-East populations (Totonacas-Veracruz) and those from the other geographic regions (Figure S1A). However, using the geography-2 criterion, there was no clear genetic separation between the South-Central region with the North-Central and Eastern populations (Figure S1B). When grouping Native Mexican populations by the language criterion, the Totonaca-Tepehua populations were clearly distinguished from both the Oto-Manguean and Uto-Aztecan populations, which themselves were genetically similar (Figure 2A; Table S5).

Table 2. Analysis of molecular variance using Y-STR haplotypes from Native Mexicans

Groups	Percentage of variation	p value
Geography-1		
Among groups	0.41	0.291
Between populations within group	5.85	≤0.0001
Within group	93.74	≤0.0001
Geography-2		
Among groups	2	0.085
Between populations within group	5.08	≤0.0001
Within group	92.91	≤0.0001
Language		
Among groups	−1.36	0.403
Between populations within group	7.16	≤0.0001
Within group	94.20	≤0.0001

Criteria definitions: *Geography-1*: Center (Hidalgo: Hñähñús, Nahuas, Tepehuas; Morelos: Nahuas; Puebla: Popolocas, Totonacas); West (Guanajuato: Hñähñús, Éza'r; Querétaro: Hñähñús); South-East (Veracruz, Totonacas); North-East (San Luis Potosí, Nahuas). *Geography-2:* North-Central (Guanajuato: Hñähñús, Éza'r; Querétaro: Hñähñús; San Luis Potosí: Nahuas); Eastern (Hidalgo: Hñähñús, Nahuas, Tepehuas; Puebla: Popolocas, Totonacas; Veracruz: Totonacas); South-Central (Morelos: Nahuas). Language: Oto-Manguean: Éza'r, Hñähñús, and Popolocas; Toto-Zoquean: Tepehuas and Totonacas; Uto-Aztecan: Nahuas.

Using the ethnicity criterion, we observed a set of overlapping relationships among the Native Mexican populations. There were differences between Nahuas-HGO and Nahua-SLP versus Nahuas-MOR, whereas the *Hñähñú* showed affinities with all groups except the *Éza'r* and Popolocas. Interestingly, the Tepehuas were genetically similar to all other groups. By contrast, the *Éza'r* showed genetic affinities with only Nahuas-MOR, Nahuas-SLP, and Tepehuas. The Popoloca were the most distant group from the rest of the populations, although showing affinities with Nahuas-SLP and Nahuas-MOR (Figure 2B; Table S6).

Non-native haplogroups in Native Mexicans

Non-indigenous Y chromosomes were observed within Native Mexican populations, and primarily those living in Hidalgo state, reflecting differing degrees of admixture in these ethnic groups. The extent of non-native admixture varied in the HGO populations: *Hñähñús* (62.3%), Nahuas (15.9%), and Tepehuas (11.6%) (Tables S2 and S7). The non-native lineages belonged mainly to West Eurasian haplogroups deriving from North-Central Europe, the Mediterranean region, and North Africa, with R1b representing over half of them (50.8%) (Table S8). These lineages also showed a high diversity of Y-STR haplotypes, with relatively few being shared in Native Mexicans populations.

In general, these results revealed the impact of the influx of European men into indigenous communities. The geographically diverse origins of the Spanish conquerors were mirrored in the diversity of the most frequent haplogroups found within Native Mexicans.

The detailed analysis of R1b haplotypes showed them to have the closest genetic relationship to those from the regions of Andalucía, Castilla La Mancha, Cataluña, and Lower Navarre (France) (Figure 3A; Tables S2, S7, and S8) (Martinez-Cadenas et al., 2016; Martinez-Gonzalez et al., 2012). Notably, the R1b haplotypes in the Nahua populations were distinct from those in the central cluster, which included the Mestizo, *Hñähñús*, and Tepehua populations, suggesting a specific source for them. Inside the *Hñähñús* from Hidalgo, several R1b haplotypes were present (Table S1; Figure 3B). Thus the *Hñähñús* from Ixmiquilpan were more related with Spaniards from Madrid and Ciudad Real, whereas *Hñähñús* from Pañhé and Tutotepec were closely related to French individuals (Martinez-Cadenas et al., 2016; Martinez-Gonzalez et al., 2012; Myres et al., 2011). By contrast, the *Hñähñús* from Zimapán showed similarities to the Jaén from the south-central region of Spain.

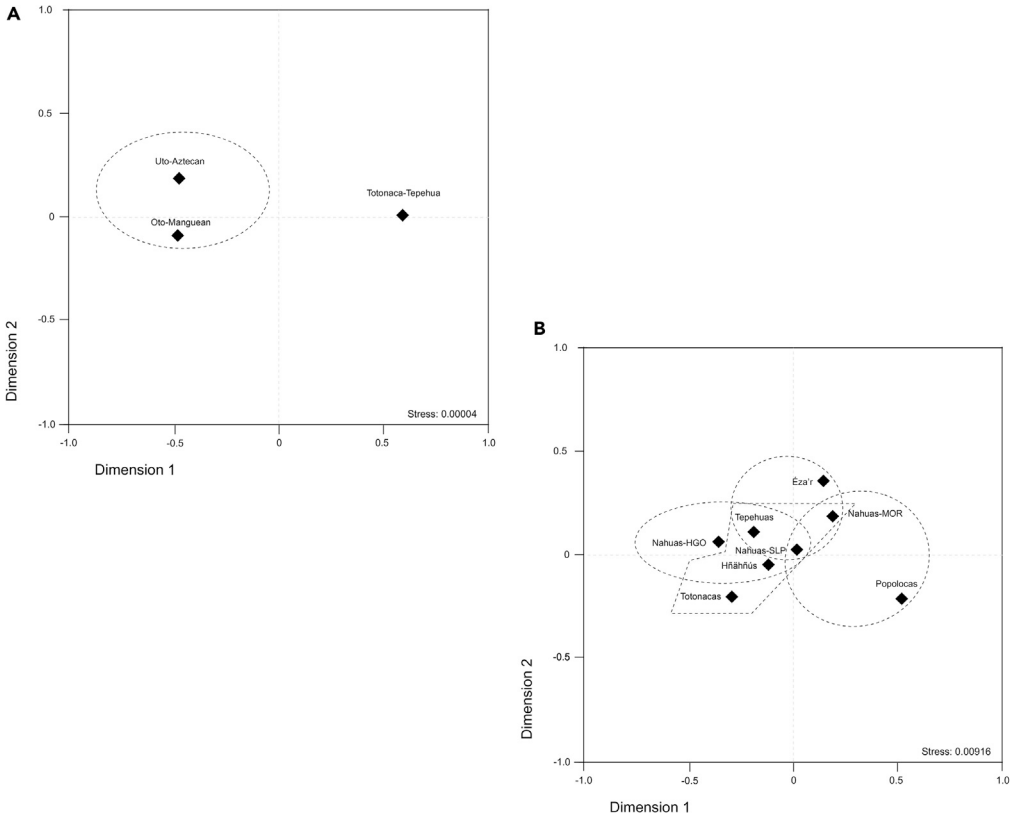


Figure 2. Multidimensional scale plot of R_{st} values estimated from 17 Y-STRs haplotypes among Central Valley of Mexico Native Americans belonging to haplogroup Q

(A) With linguistic affiliation as criterion.

(B) With ethnicity as criterion. The dotted circles enclose populations that share no significant genetic distances.

Note: HGO, Hidalgo; GTO, Guanajuato; MOR, Morelos; PUE, Puebla; QRO, Querétaro; SLP, San Luis Potosí; VER, Veracruz. Oto-Manguean: *Éza'r* (Chichimecas); *Hñähñús* (from HGO, GTO, and QRO) and Popolocas; Totonacas-Tepehuas: Tepehuas and Tononacas from PUE and VER; Uto-Aztecans: Nahuas from HGO, MOR, and SLP. All *p* values were adjusted with the method of false discovery rates in R-software.

Regarding other paternal lineages, the haplotypes from haplogroups E, I, and J in Native Mexicans showed similarities to those from Aragon, Castilla y Leon, Extremadura, Madrid, and the Valencian communities of Spain (Tables S2, S7, and S8; Figure S2) (Martinez-Cadenas et al., 2016; Martinez-Gonzalez et al., 2012). It is noteworthy that the Native Mexican populations living fairly close to urban settlements exhibited the highest frequency and widest diversity of non-native haplogroups (i.e., E1b1b, G, I, J, L, R1b, and T). In this regard, some of the sampling locations for the *Hñähñú* were located 11 to 20 km from urban settlements, a location that facilitates intermarriage with Mestizos. In fact, *Hñähñú* communities close to Ixmiquilpan-HGO had close to 30% non-native lineages, whereas the *Hñähñú* community located closest to the Queretaro border had a moderate frequency of haplogroup R1a haplotypes (18%)

Ancestral connections with Asian and other Native American populations

The remarkable Y-STR diversity described above was mirrored in the networks for haplogroups Q-MEH2, Q-L54, and Q-M3. The data from these three lineages supported a connection between present-day Native Mexicans and ancient founder populations that gave rise to modern Native Americans. However, each haplogroup showed subtle differences in its phylogenetic structure.

The Q-MEH2 network contained haplotypes from Northeast Asians, Native North Americans, and the Nahuas-HGO (Figure 4). At least 16 mutations separated the haplotypes in Nahuas-HGO from those in the other populations, suggesting a significant divergence between them. Interestingly, the Tlingit

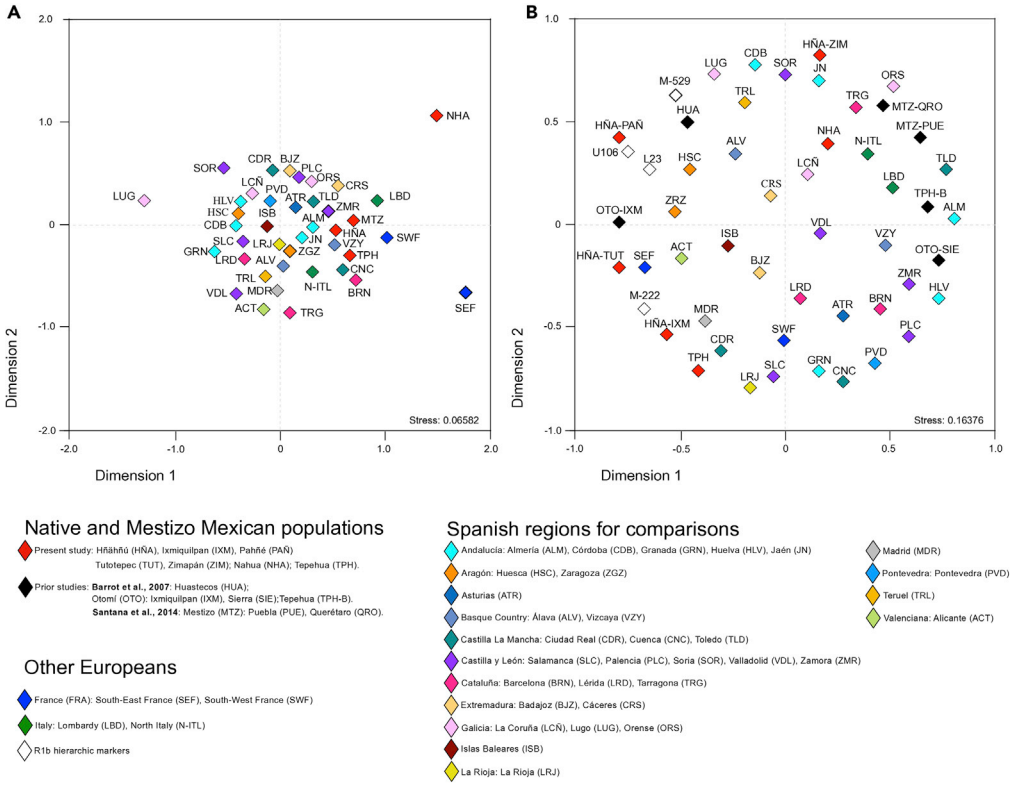


Figure 3. Multidimensional scale plot of R_{st} values

(A) Estimated from 14 Y-STRs haplotypes belonging to haplogroup R1b.

(B) Estimated from 7 Y-STRs and comparing the *Hñähñú* populations from Hidalgo state by sampling zone.

Note: The comparative data were taken from Barrot et al. (2007); Martinez-Cadenas et al. (2016); Martinez-Gonzalez et al. (2012); Myres et al. (2011); and Santana et al. (2014). Although *Hñähñú* is the name by which the Otomi now self-identify, we use the older name provided by the original authors (Barrot et al., 2007). The red diamonds show the data from the present study.

haplotype was closer to those of Northeast Asians than to that of the Nahua-HGO. This finding suggests that the ancestral Q-MEH2 haplotype in the Nahua-HGO arrived much earlier in the Americas than the haplotypes present in more northerly populations. However, in being based on only seven Y-STR loci, this network likely does not fully represent the pattern of diversity in this haplogroup.

By contrast, the Q-M3 and Q-L54 networks exhibited long branches, multiple reticulations, and missing intermediate nodes, with these features hinting at the existence of as yet undefined sublineages (Figures 5A, 5B and 6). Each network also lacked a clear central node, suggesting considerable mutational divergence from the founding haplotype for each of them. Furthermore, each network had a high number of singletons and the presence of isolated groups and distant haplotype clusters within it. The overall lack of haplotype sharing among these Native Mexican populations was again quite remarkable. The haplotype diversity within the Q-M3 network was reflected by the intrapopulation variance (*Vp*) obtained from 17 Y-STRs, which was high among the different Native Mexican populations, suggesting an ancient diversification process (Tables 1 and S2). A tight group with several long, separated branches suggested the coexistence of different sublineages within it (Figure 5A). When defining SNPs were added to the network analysis, the Q-Z768 haplotypes were mainly distributed inside of the central network, whereas Q-L663 and some Q-M3 haplotypes formed separated branches (Figure 5B). VM populations also presented nuanced connections with other Nahua populations from the south of Mexico (i.e., Santo Domingo Ocotitlan in Morelos, and Zitlala in Guerrero) and the southeast of Mexico City (i.e., San Pedro Actopan and Xochimilco) (Sandoval et al., 2012) (Figure S3; Table S9). Remarkably, these different Nahua groups did not share any haplotypes, a finding reflecting the great genetic diversity within this ethnic group.

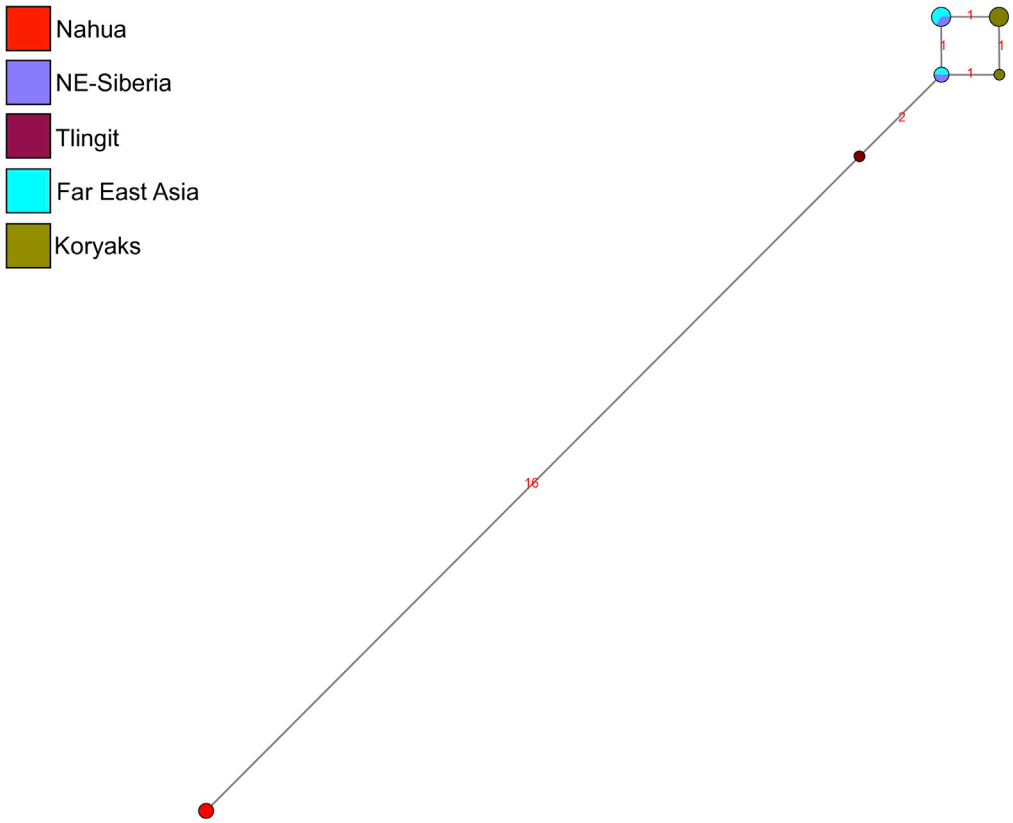


Figure 4. A median-joining network of Q1a (Q-MEH2) lineage using 7 Y-STRs

Comparative data from Regueiro et al. (2013) and Schurr et al. (2012). **Note:** The smallest circles represent one individual. NE-Siberia, Northeast Siberia.

Q-M3 haplotypes in Native Mexicans using 14 Y-STRs showed connections with those of different indigenous North American groups (Figure S4; Table S9). These included the Inuvialuit (*Hñähñús*, Nahuas-SLP, and Totonacas), Tlingit (*Hñähñús*, Nahuas-HGO, Nahuas-SLP, and Totonacas), and Gwich'in (*Hñähñús*, Nahuas-SLP, and Popolocas). A number of these North American haplotypes appeared on a single branch extending from the left side of the network. On the other side, a separate cluster (shown in inset) included Tlicho, Tlingit, Gwich'in, and one *Hñähñús* haplotype and branch extending from it included Gwich'in, Tlingit, Totonaca, and Nahua haplotypes.

We also compared Native Mexican haplotypes with those from the Greenland Inuits (Olofsson et al., 2015) and Inuvialuit of the Northwest Territories (Dulik et al., 2012a, 2012b) using 17 Y-STR haplotypes. Most Inuit and Inuvialuit haplotypes clustered together with a single Nahua haplotype (Figure S5; Table S9). One other Inuit Q-M3 haplotype appeared on a branch with Native Mexican haplotypes. The position of the Inuit haplotypes suggested a possible ancestor-descendant relationship with those of Native Mexicans, although the former may possess SNPs not present in the latter.

We further analyzed the genetic relationships between Native Mexican populations and indigenous groups from Central America (Guatemala, El Salvador, Nicaragua, and Panama) using R_{ST} values from 14 Y-STR haplotypes belonging to Q-M3. Through this analysis, we observed some similarities among these populations (Figure S6; Table S10). Populations in regions close to Mexico (Guatemala, El Salvador, Nicaragua) were more genetically similar, whereas those from Panama differed from nearly all Native Mexican groups. This finding contrasts with previous studies, where certain Chibchan-Panamanian haplotypes were similar to ones appearing in the Nahuas and *Hñähñús*, suggesting possible Uto-Aztecan and Oto-Manguean gene flow to the Panama region (Nunez et al., 2010).

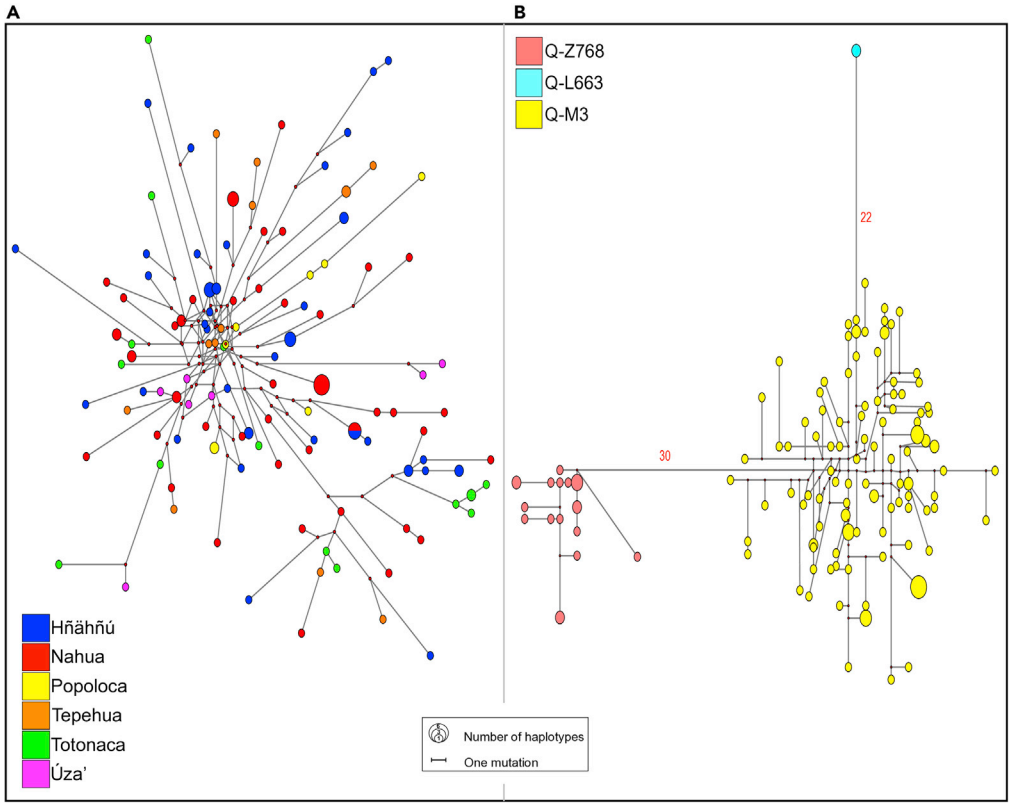


Figure 5. Median-joining network of haplogroup Q1b1a1a (Q-M3) based on 17 Y-STRs

(A) With the ethnicity as criterion.

(B) With the distribution of sublineages Q-Z768 and Q-L663.

Note: The circle size indicates the number of individuals with shared haplotypes; the smallest circles represent one individual.

We also evaluated the genetic relationships between indigenous South American and Native Mexican populations using 14 Y-STRs from Q-M3 (Figure S7; Table S11). In the resulting MDS plot, the South American populations were generally similar to Native Mexicans, although appearing around the periphery of the main cluster of Mexican populations. When mapping known SNPs in South American Q-M3 haplotypes onto these populations, we observed that the Q-SA04, Q-SA05, and Q-Z5915 sublineages were confined to South American populations, as previously seen (Jota et al., 2016). The Q-SA04 sublineage seems to have a recent origin, whereas the Z5915 SNP along with other markers may help to differentiate Y-chromosomes in South American populations (Jota et al., 2016).

The apparent connection between Bolivian and Peruvian populations may be due to a shared apomorphy, which has diverged recently in indigenous South Americans. However, Bolivians with Q-CTS11357 haplotypes clustered among Native Mexicans. This was not wholly surprising, as Q-CTS11357 haplotypes have recently been documented in Mexican populations (International Society of Genetic Genealogy [ISOGG], 2019-20). Thus, it is likely that some Native Mexican men having Q-M3 Y chromosomes belong to this sublineage.

Unlike that for Q-M3, the Q-L54 network exhibited a slightly simpler topology, with a number of long branches (Figure 6). *Hñähñú* and Nahua haplotypes were scattered across the network reflecting their great diversity, an assessment also supported by the highest *Vp* and the mean pairwise differences values in these groups (Table 1). A similar distribution was previously noted in Mexican and some Central American populations (Battaglia et al., 2013) (Figure S8; Table S9). Totonacan haplotypes grouped into a single subbranch and also showed the lowest *Vp* values (Table 1). Again, remarkably, almost none of these Y-STR haplotypes were shared between the study and comparative populations. Furthermore, the

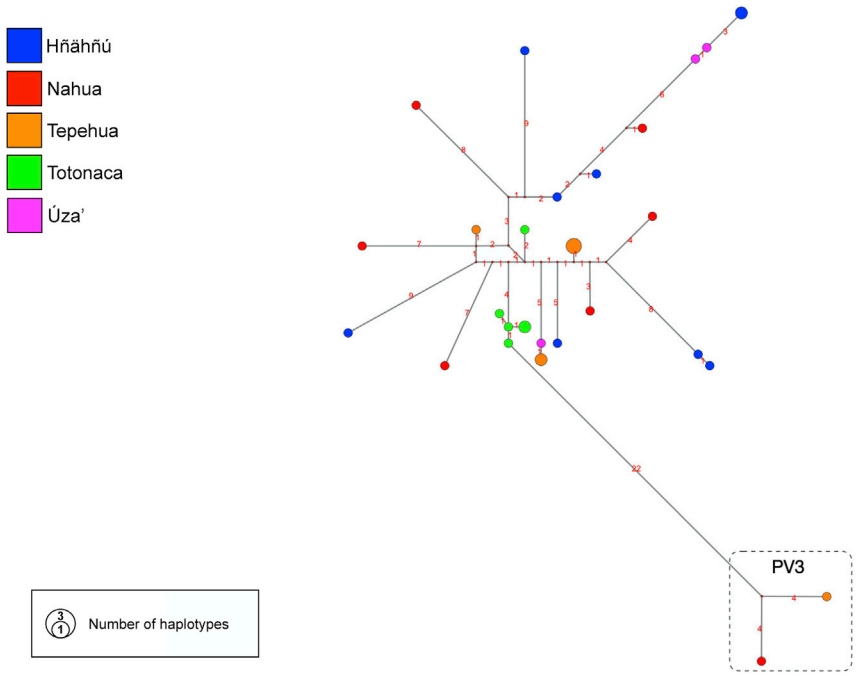


Figure 6. A median-joining network of Q1b1a2 (Q-Z780) and Q1b1a3a1f (Q-PV3) lineages using 17 Y-STRs
Note: The dotted line shows the carriers of PV3 sublineage.

Q-PV3 sublineage within Q-L54 was quite divergent from the other haplotypes in the network, suggesting it might represent another founding lineage in the Americas (Figure 6).

We evaluated the genetic relationship of Q-L54 haplotypes present in Native Mexican and other comparative populations using 14 Y-STR haplotypes. We noted that Native Mexicans showed considerable similarity to other Mesoamerican and South American populations (Battaglia et al., 2013) (Figure S9; Tables S9 and S12). By contrast, the Gwich'in, Altai-kizhi, and Tuvans were distinct from all New World populations.

A network analysis of Q-L54 haplotypes defined by 15 Y-STRs in Native Mexican and comparative populations revealed substructure within this haplogroup (Figure S10A; Table S9). Altai-kizhi and Tuvan haplotypes formed a central cluster in the network that contained one Tepehua sample, whereas all indigenous Canadian haplotypes appeared in a separate branch containing one Nahua haplotype. The remaining Native Mexican haplotypes were widely dispersed along other branches.

We further explored the relationship between Native Mexicans and South American populations using 15 Y-STR haplotype data from Q-L54. The resulting MDS plot showed genetic affinities between these populations through the Q-CTS1780 sublineage (Figure S10B; Tables S9 and S13), which is closely related to the Q-Z780 sublineage. Both Q-CTS1780 and Q-Z780 define the sub-haplogroup Q1b1a2 (ISOGG, 2019-20). Their dissimilarity with Colombian populations was possibly related to the ancient differentiation of Q-L54 haplotypes in the Isthmo-Colombian region, as reported previously (Grugni et al., 2015) and evident by the presence of the Q-SA02 and Q-SA03 sublineages (Jota et al., 2016). In both the Q-M3 and in Q-L54 networks, the comparative populations shared few haplotypes, a finding mirroring the great Y chromosome diversity of the Native Mexican populations.

Timing of the arrival of indigenous haplogroups in Mesoamerica

Coalescent times (time to more recent common ancestor [in years] [TMRCAs]) were estimated for haplogroups Q-MEH2, Q-M3, and Q-L54 using ρ-statistics and Bayesian analysis. The results presented in Table 3 correspond to the average of three generation times (18, 20, and 25 years). In general, the TMRCA estimates for Q-M3 and Q-L54 suggested their simultaneous arrival in Mesoamerica, consistent with previous coalescent estimates for these lineages (Grugni et al., 2019). Nonetheless, subtle differences in the

Table 3. Coalescent time estimates for NRY haplogroups in Native Mexican populations

Population	*N*	ρ-Statistic TMRCA (YBP)	ρ-Statistic SE	Batwing TMRCA (YBP)	Batwing 95% CI	Beast TMRCA (YBP)	95% HPD
Q1a (Q-MEH2)							
Nahuas	4	3,624	±1,001	9,699	8,067–11,331	4,400	0–12,600
Q1b1a1a (Q-M3)							
Éza'r	7	4,700	±875	24,310	21,218–27,402	8,810	3,880–14,400
Hñähñús	44	6,254	±805	21,041	19,974–22,109	14,990	9,460–20,420
Nahuas	52	5,436	±694	17,823	16,991–18,654	15,170	12,520–24,270
Popolocas	8	4,926	±981	22,921	20,194–25,648	9,870	3,720–16.500
Tepehuas	13	5,506	±968	24,880	22,557–27,202	10,470	4,510–16,890
Totonacas	15	6,946	±971	20,043	18,320–21,785	13,470	6,860–20,480
Overall	139	4,771	±583	18,589	18,058–19,119	17,151	15,970 - 20,250
Q1b1a2 (Q-L54)							
Éza'r	3	5,662	±555	22,775	18,350–27,200	12,000	2,000–29,000
Hñähñús	9	6,946	±1,087	22,702	20,155–25,249	13,000	3,000–35,000
Nahuas	7	5,371	±782	23,788	20,763–26,814	13,890	2,000–24,000
Tepehuas	7	2,537	±496	20,462	17,859–23,065	7,000	1,000–18,000
Totonacas	6	2,265	±830	17,505	15,100–19,910	5,000	1,000–13,000
Overall	32	5,053	±664	27,479	25,844–29,113	17,155	14,730 - 24,150

N, number of samples; ρ, average distance from root haplotype (in years); YBP, years before present; SE, standard error; TMRCA, time to more recent common ancestor (in years); CI, confident interval; HPD, highest posterior density.

TMRCAs were found among the populations. Using Bayesian methods, the TMRCA for Q-M3 ranged between 17,823 years before present (YBP) (Nahuas) and more than 24,000 YBP (*Éza'r* and Tepehuas), whereas the TMRCA for Q-L54 ranged from 17,505 YBP (Totonacas) to 23,788 YBP (Nahuas). These data support the diversification of these lineages soon after their arrival in Mexico and Mesoamerica. The TMRCA for Q-MEH2 was 9,699 YBP, suggesting a more recent arrival in a separate migration, although the sample size for this lineage was quite small.

Overall, the TMRCAs determined using Bayesian methods were 2–3 times the values generated with ρ-statistics (Bouckaert et al., 2014; Wilson et al., 2003). Of note, the overall TMRCA values obtained for Q-M3 by the two Bayesian methods were consistent, whereas those estimated for Q-L54 were discrepant with each other, possibly due to the modest sample size for this paternal lineage.

DISCUSSION

Spanish chroniclers described the Aztec Empire as being culturally and ethnically diverse, where populations from different language groups, such as *Hñähñú*, *Nahuatl*, Totonac, Popoloca, and Huasteco, coexisted in the tropical lowlands during the Olmec period (Sahagun, 2011; Soustelle, 1993; Torquemada, 2018). These features provide a unique opportunity to reconstruct the diversity of the Aztec descendants and the coexisting ethnolinguistic groups, which has been briefly explored here. In addition, more fully characterizing the multi-ethnic nature of contemporary Mexican Native Americans is crucial for expanding our knowledge about the initial colonization of the New World. The key to answering these two questions is elucidating the kinds of ancient lineages present in the contemporary Native Mexicans rather than Mestizo populations, which have experienced higher levels of European and African admixture (Battaglia et al., 2013; Sandoval et al., 2012).

Y chromosome diversity in Native Mexican populations

Contemporary Native Mexicans exhibit remarkable Y chromosome diversity within the three main haplogroups found within their populations. This diversity is not strongly structured by ethnicity and

geography. Comparable findings have been reported previously in indigenous populations from Mexico (Sandoval et al., 2012) and South America (Roewer et al., 2013).

The notable diversity in the *Hñähñús* could reflect their antiquity in the region, being contemporaries of the Cholula, *Pinotl*, Teotihuacan, Tlaxcalteca, and Toltec peoples (Sahagun, 2011; Soustelle, 1993; Torquemada, 2018). The significant differences found in the Nahuas-MOR could reflect their historic link with Toltec descendants (i.e., *Tlahuicas*, *Xochimilcas*), Aztec sister cultures (i.e., *Tepanecas* and *Alcohuas*), and the spread of Oto-Manguean languages (i.e., Matlatzincas, Mazahuas, *Hñähñús*), which has been supported in previous reports (Instituto Nacional de los Pueblos Indígenas, 2019; Nichols and Rodriguez-Alegria, 2017; Sandoval et al., 2012).

The continuous genetic exchange among ancestral Native Mexican populations likely enriched their diversity over time, leading to the relative genetic similarities among the study populations (Perego et al., 2010). These practices contributed to acculturation and allowed Aztecs to gain supremacy in Mesoamerica, incorporating new cultural, linguistic, and social variation into the communities of themselves and subject populations (Nichols and Rodriguez-Alegria, 2017). Aztec populations, diverse *per se*, intermarried with Toltecs, Totonacas, Tepehuas, and *Hñähñús* (Berdan et al., 1996; Duverger, 2007). The *Hñähñús* continuously interacted with other Oto-Manguean groups, such as the *Éza'r*, Mazatecas, Nonoalcas, and Popolocas (Josserand et al., 1984; Lastra, 2006). Furthermore, the *Hñähñús* also mixed with groups from Oaxaca (Chinantecos, Mixes, Mixtecos, and Zapotecos), Veracruz (Popolucas, Totonacas), Hidalgo (Tepehuas), and Puebla (Totonacas-PUE) (Lastra, 2006; Soustelle, 1993). Thus, continual gene flow between different peoples who apparently lacked cultural boundaries can partly explain the remarkable diversity found in the study populations.

Consequently, the Aztec Empire could have been a conglomerate of different populations that traced their origins to *Aztlan*. Aztec cities were founded through fissiparity, leading to the homogenization and agglutination of different peoples (Berdan et al., 1996; Duverger, 2007). The lack of genetic stratification by ethnicity and geographic criteria would be supported by these practices, besides the continuous population growth of the constituent ethnic groups (Roewer et al., 2013). Notwithstanding these population dynamics, the gene pool of pre-Columbian Mesoamericans has been diversifying since the founding populations first arrived there (de Azevedo et al., 2015). Thus the notable diversity observed in this study could also reflect admixture events before Mesoamerican populations began to ethnically diversify, consistent with previous reports (de Azevedo et al., 2015; Regueiro et al., 2013; Ruiz-Linares, 2014). Founder populations had recurrent streams both between them and with recently emerging Native Mexican groups, who had probably become locally differentiated (Kumar et al., 2011).

The significant differences between the Popolocas and the other four ethnic groups was intriguing. These differences could possibly be related to the Popoocas' recent separation from the *Ngiwas* (Chocho-Popolocan speakers) around 1000 YBP (Instituto Nacional de los Pueblos Indígenas, 2019). Popolocas have also maintained genetic and cultural exchange with Nahuas-MOR and Totonacas-PUE, whereas their interactions with the Nahuas-SLP are marginal. Thuis, further studies are needed to elucidate their remarkable differences, perhaps by increasing the sample size.

The amalgam between Oto-Manguean and Uto-Aztecan linguistic families could reflect the biocultural makeup of the ancient cosmopolitan societies. A tight relationship between the Nahua ancestors and the first settlers of the VM was described (Sahagun, 2011; Soustelle, 1993; Torquemada, 2018; Vargas-Alarcon et al., 2007). Before the Aztecs began their pilgrimage to the VM, *Hñähñús* and other ethnic groups already resided in this geographic area, establishing cultural and military alliances with them (Berdan et al., 1996; Duverger, 2007; Nichols and Rodriguez-Alegria, 2017; Soustelle, 1993). Overall, this pattern of settlement is consistent with the TMRCAs for the Q haplogroups, which support the antiquity of these indigenous groups in the region (Delgado-Burbano et al., 2010; Sandoval et al., 2012). Thus, the connection between these two linguistic families could reflect at least 370 years of gene flow between them (Beezley, 2011). The *Nahuatl* monopolized Mesoamerica, at least for three centuries, homogenizing the different peoples and serving as a *lingua franca* during the Aztec Empire (Duverger, 2007; Hill, 2001).

Alternately, the remarkable diversity within the Nahua groups could reflect the consequences of the "*Nahualization*" process experienced by the neo-Aztec civilizations during the diffusion of maize

domestication with the subsequent imperial expansion (Vargas-Alarcon et al., 2007). However, considerable work will be needed to determine whether the ancestors of the *Hñähñús* or the Nahuas contributed to the diversity of the others (aside from the military supremacy). Given the lack of continuity between the Uto-Aztecan and Aztec populations, some results should be interpreted in light of this limitation.

The differences between Oto-Manguean and Uto-Aztecan regarding the Totonaca-Tepehua linguistic affiliation could be a consequence of its geographic localization (Puebla's *Sierra Norte*). While the Tepehua and the Totonac were related for at least 26 centuries (Hernandez-Montes and Heiras-Rodríguez, 2004), the Totonac have also been associated with Mayan, Mixe, Popoluca, *P'urhépecha*, and Zoque languages (Hernandez-Montes and Heiras-Rodríguez, 2004; Moreno-Estrada et al., 2014). Likewise, populations from Guatemala, El Salvador, and Panama have historically interacted with Totonacas, suggesting genetic interaction between them (Witschey and Brown, 2012). The separation between the Tepehuas and the Totonacas could also be a consequence of cultural and demographic events (with sex bias), as well as by the genetic drift to which uniparental markers are more susceptible (Raghavan et al., 2015).

The absence of population stratification observed in our analysis contrasts with the results of previous reports analyzing autosomal diversity (Ávila-Arcos et al., 2020; Moreno-Estrada et al., 2014). In this case, although genetic substructure was reported, it was detected only when long geographic distances were compared (i.e., Mesoamerican versus Northern populations, or among Central, Northern, and Southern regions) (Ávila-Arcos et al., 2020; Gonzalez-Sobrino et al., 2016; Sandoval et al., 2012). Alternatively, the lack of genetic structure could be explained by the high prevalence of Q-M3 haplotypes in Native Mexicans (>79%). The fact that the majority of the populations studied were geographically adjacent could also possibly have influenced our analysis. It may further be the case that the lack of genetic stratification found in our study could focus on the haploid portion of the genome, not autosomal markers as in other studies (Ávila-Arcos et al., 2020; Moreno-Estrada et al., 2014).

Impact of European admixture in Native Mexican population from Hidalgo

The most significant contribution of non-native paternal lineages is within ethnic groups from Hidalgo state (*Hñähñús*, Nahuas, and Tepehuas), some of them living in the *Sierra de Pachuca*, a small extension of the *Sierra Madre Oriental*. This geographic region is rich in metallic minerals, with the silver mines discovered before 1552 being exploited since the colonial period (Saavedra-Silva and Sanchez-Salazar, 2008). Spanish immigrants moved to this geographic region to establish the first mining towns (Beezley, 2011) and, in the process, contributed their paternal lineages to local Native Mexican communities. The scarcity of female migrants (6%) (Ongaro et al., 2019) further stimulated gene flow between Spanish men and Native Mexican women, reducing the representation of indigenous paternal lineages in these communities but allowing them to maintain their indigenous languages. This admixture process could have occurred primarily in those communities located near the Pachuca-Real del Monte corridor (Saavedra-Silva and Sanchez-Salazar, 2008), and then extended to nearby towns (i.e., Ixmiquilpan). Gene flow between populations from urban and rural regions of Central Mexican states could thus have affected the frequency of Y chromosome lineages in indigenous populations, including haplogroup Q and its major sub-branches (Gonzalez-Sobrino et al., 2016).

Given this history, it was not unexpected to find that the European paternal lineages in Native Mexicans resembled those from several regions of the Iberian Peninsula. Similar findings using eight Y-STR haplotypes were previously reported in Huastecos and *Hhähñús* from Ixmiquilpan, where 26% non-indigenous lineages were found (Barrot et al., 2007). In addition, our findings were in accordance with those reported in the *Hñähñú* population from Mezquital Valley (located at 27 km to Ixmiquilpan in Hidalgo) based on Y-SNPs (Gonzalez-Sobrino et al., 2016). On the other hand, some Nahua R1b haplotypes showed some genetic dissimilarity from others coming from Spain. These haplotypes could have been brought more recently (during the 19th century), when mining economic activity was opened to other foreign investment from England, France, Germany, Italy, and North America (Ongaro et al., 2019; Saavedra-Silva and Sanchez-Salazar, 2008).

Another important contribution of non-indigenous haplogroups was found on the border between Hidalgo and Queretaro (18%), where individuals with R1a Y chromosomes were found. A previous analysis of paternal diversity in populations from several states from the VM (including Queretaro) also detected R1a in them (Santana et al., 2014). Queretaro is a state with a remarkable diversity of paternal lineages

from Spain, the Basque Country, and even Jewish populations (Santana et al., 2014), due its importance in the mining production (i.e., gold, silver, copper, among others) (Servicio Geológico Mexicano, 2018).

Native American origins from a Y chromosome perspective

In the context of the peopling of the Americas, the striking diversity found in Native Mexican populations, as well as the genetic connections with Central and South American and Northeast Siberia populations, supports a Beringian entrance with subsequent genetic differentiation, as also seen with mtDNA data (Dryomov et al., 2015; Llamas et al., 2016; Schurr and Sherry, 2004; Skoglund and Reich, 2016). The presence of haplogroups Q-L54 and Q-M3 reflects the emergence and diversification of these lineages in Asia before the Mesoamerican settlement (de Azevedo et al., 2015; Llamas et al., 2016; Reich et al., 2012; Ruiz-Linares, 2014), as well as their subsequent diversification during the long Beringian standstill (Grugni et al., 2019; Pinotti et al., 2019) and the subsequent pre-Columbian period (Sandoval et al., 2012).

Haplogroup Q-L54, which connects modern Nahua with indigenous Altaians and Tuvinians of Siberia (Dulik et al., 2012b), seems to have undergone *in situ* differentiation during which a number of derivative branches arose (Grugni et al., 2019; Jota et al., 2016; Pinotti et al., 2019). Within Q-Z780, the most prominent lineage (~93%) found in our Q-L54 samples, three main sublineages have been defined, namely, Q-Z781, Q-SA02, and the Q-Z780 paragroup (Q-Z780)* (Grugni et al., 2019). Based on the frequency and the TMRCA for these sublineages (Q-Z780* + Q-SA02; 26,900 YBP) (Grugni et al., 2019), our Mexican Q-L54 samples likely belong to the Q-Z780* sublineage.

Nonetheless, we did not dismiss the possible presence of other sublineages within Q-Z780. In light of other published evidence, samples from Nahuas, Pimas, and others obtained from Mexican ancestry in Los Angeles (MXL) have been shown to have Q-Z781 Y chromosomes, while the Q-Z782 has also been found in Mexican samples (ISOGG, 2019-20; Pinotti et al., 2019; The International Genome Sample Resource, 2020). In addition, the Z780 and CTS1780 SNPs defining Q1b1a2 have been observed in 6% of MXL samples (ISOGG, 2019-20; The International Genome Sample Resource, 2020). Thus, we may detect these sublineages in our populations through more detailed SNP genotyping.

The main sublineage of Q-L54 (Q-Z-780/Z-777) also connects *Aztlan* descendants to the Clovis individual found in Montana (Anzick-1) (Rasmussen et al., 2014). This connection suggests that individuals carrying Q-Z780 Y chromosomes followed the Pacific coast route southward, first peopling Mexico and then moving into South America (Grugni et al., 2019; Sandoval et al., 2012). Although the Q-PV4 and Q-L191 sublineages have been previously described within this haplogroup in Mexican populations (Battaglia et al., 2013), they were not found in the present study.

The other major Y chromosome lineage in the Americas, Q-M3, exhibits extraordinary diversity (with sublineages Q-Z768 and Q-L663) in Native Mexican populations. Our comparative analysis of this paternal lineage also reveals close connections between populations from North, Central, and South America, affirming a southward peopling process (Jota et al., 2016; Olofsson et al., 2015). Although most Y chromosome haplotypes belonging to this haplogroup were not fully defined by the panel of SNPs and STRs used in this analysis, our data suggest that additional undefined sublineages are present among them. To further assess this possibility, we compared our data with those from recently published studies to make inferences about the substructure of haplogroup Q-M3 in Native Mexican populations.

Recent reports have identified two major sub-branches within Q-M3, these being Q-M848 (Q1b1a1a) and Q-Y4276 (Q1b1a1a2) (Grugni et al., 2019). Several different sublineages within Q-M848 have also been defined, including Q-M925 (Q1b1a1a1e) (Grugni et al., 2019; ISOGG, 2019-20). As both the Q-L663 and Q-Z768 sublineages have the M925 SNP, they must belong to this particular sublineage, as well as to Q-M848. The rest of our Q-M3 samples could belong to other sublineages within Q-M925, such as Q-CTS748 (Q1b1a1a1e1) or Q-CTS1002 (Q1b1a1a1e1a), which have also been identified in Native Mexicans (Grugni et al., 2019; The International Genome Sample Resource, 2020).

Along these same lines, the MDS plot depicted in Figure S9 showed a close connection between Bolivian carriers of Q-CTS11357 Y chromosomes and our Nahua and *Hñähñú* samples. CTS11357 is essentially equivalent to the M925 SNP, phylogenetically speaking (ISOGG, 2019-20), and thus identifies the same sublineage. Previous studies have shown the Nahua and Pima to have the Q-CTS11357 sublineage

(Pinotti et al., 2019), and the same is true for MXL, Colombians in Medellin, Karitianas from Brazil, and indigenous populations from Guatemala, El Salvador, and Nicaragua (ISOGG, 2019-20; The International Genome Sample Resource, 2020). Thus, the Q-CTS11357/Q-M925 sublineage (Q1b1a1a1e) is ubiquitous among indigenous populations of the Americas.

Yet other sublineages with haplogroup Q-M3 could be present in our Native Mexican populations. One of these might be Q-Z5915 (Q1b1a1a1k1) (ISOGG, 2019-20). This is suggested by Bolivian and Peruvian populations having Q-Z5915 Y chromosomes (The International Genome Sample Resource, 2020) showing some connection with the Mexican *Éza'r*, Nahuas, Popolocas, and Tepehuas (Figure S9) (Jota et al., 2016). Others may be part of the clade Q1b1a1a1m (Q-CTS2731), because samples belonging to this sublineage have been detected in MXL, Mixtec, and Zapotec populations (ISOGG, 2019-20; Pinotti et al., 2019; The International Genome Sample Resource, 2020). However, because Mixtec and Zapotec populations live in southern Mexico, it is possible they will not be present in Native Mexican groups from the VM.

As noted above, a second major branch of Q-M3 is Q-Y4276 (Q1b1a1a2), which appears in populations extending from Siberia to South America (Grugni et al., 2019). Its distribution across this region suggests that this sublineage could be present in our Native Mexican populations. This possibility is suggested by the general similarity of our Native Mexican populations and Athapaskan-speaking populations from North America. If true, then we might find Q-Y4276 Y chromosomes in Mexico, perhaps as a result of North American populations migrating the southern part of California and northern Mexico (Grugni et al., 2019).

By contrast, the genetic relationship with the Greenland Inuit and Aboriginal populations from Canada seen in the Q-M3 and Q-L54 data for our study populations suggests the possible influence of post-glacial migrations during which ancestors of these populations crossed eastern North America to reach Greenland. To explore this possibility, the Mexican Q-M3 samples might be surveyed for markers such as Q-B143 that have been related to the Arctic route (Grugni et al., 2019), even if this SNP has not yet been observed in Native Mexican samples. In any case, this connection should be interpreted with caution given the small sample size of Greenland Inuit and Mi'kmaq haplotypes used in the analysis.

Despite Q-MEH2, Q-L54, and Q-M3 being related to the initial migration (Battaglia et al., 2013), the belated arrival of Q-MEH2 (3624–4400 YBP) in Mesoamerica could be associated with a more recent migration (close to 5,500 YBP, via Beringia) (Rasmussen et al., 2010). Owing to the low-resolution (i.e., seven Y-STRs) data used to construct the median network and the reduced sample size, however, the TMRCA for this haplogroup should be taken with caution.

Given the relationship between Nahuas and Athapaskan-speaking populations (present study), Q-MEH2 could have arrived in Mexico using the Pacific coastal route, a view that is also supported by mtDNA evidence (i.e., A2, D2a, and D4h3) (Dulik et al., 2012a; Llamas et al., 2016; Raghavan et al., 2015). Under this scenario, Q-MEH2 would have been brought by founder populations dispersing along the western coast of North America (Figure 7) (Raff et al., 2015). Of the four Nahua carriers of Q-MEH2, three of them also had mtDNA haplogroup A2, a maternal lineage that is particularly frequent within Nahua populations (Mata-Miguez et al., 2012; Penaloza-Espinosa et al., 2007). These postglacial migrations appear to have involved recurrent gene flow between Beringia to the Arctic Circle (Dulik et al., 2012b; Helgason et al., 2006; Rasmussen et al., 2010; Ruiz-Linares, 2014). Alternatively, the apparent connections with circumpolar populations could be due to recurrent mutations (Olofsson et al., 2015) or perhaps a recent Y chromosome diversification process (Dulik et al., 2012b).

However, Q-F1096, a sublineage within Q-MEH2, has been reported in Alaska and Peru (Figure 7) (Grugni et al., 2019; Karmin et al., 2015; The International Genome Sample Resource, 2020). Thus, although the frequency of this haplogroup is limited *per se*, its rarity in Native Mexicans could be explained by demographic events occurring before the entry into Mesoamerica. Founder effects (possibly occurred in North America), along with isolation (geographic and population), may have made populations more susceptible to strong genetic drift, which would have impacted the diversity and frequency of Q-MEH2 haplotypes (Bortolini et al., 2003; Regueiro et al., 2013).

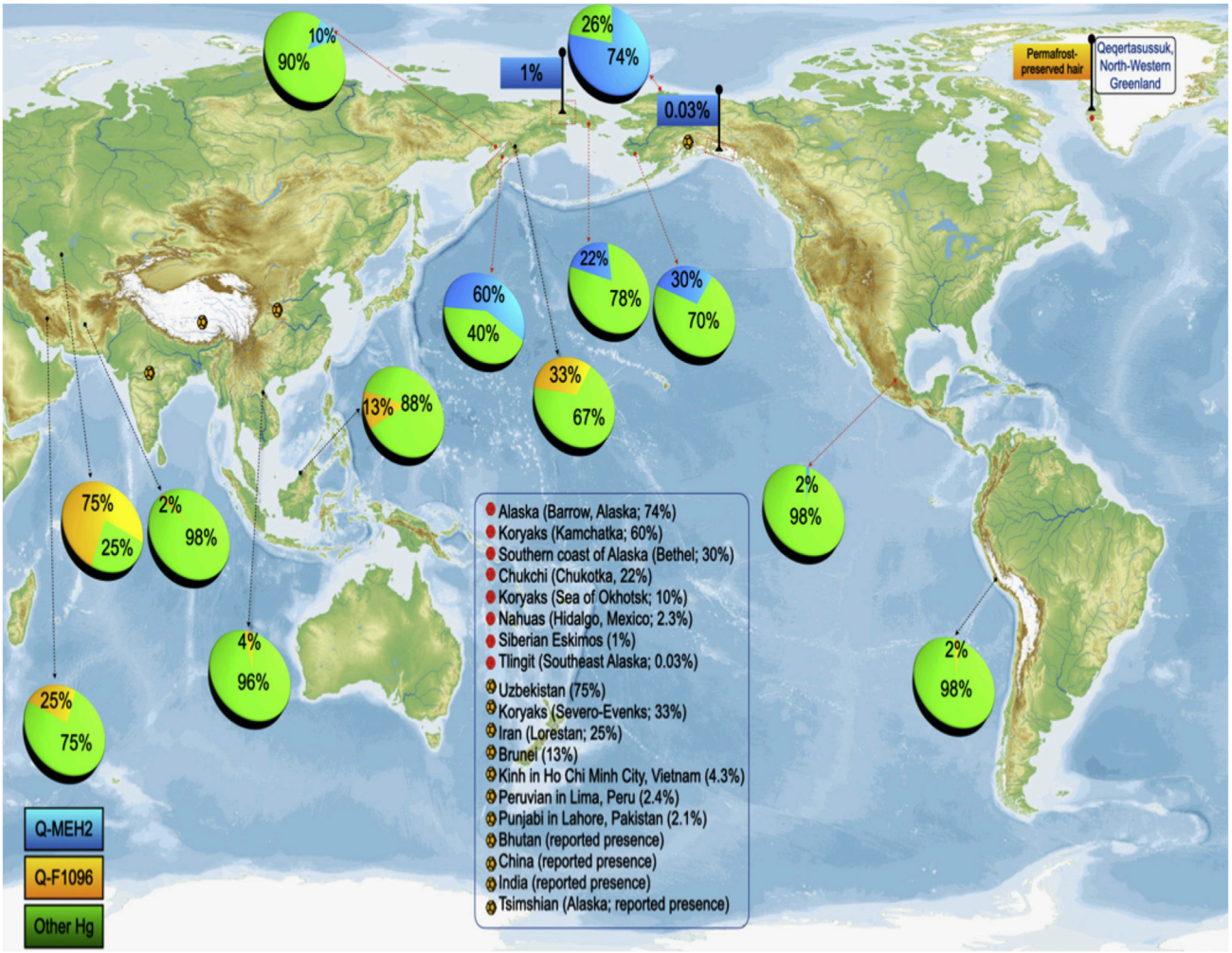


Figure 7. Frequency of Q-MEH2 lineage and its sublineage Q-F1096 in Asian and Native American populations

Note: Data from Battaglia et al. (2013), Bisso-Machado et al. (2011), Karmin et al. (2015) Malyarchuk et al. (2011), Rasmussen et al. (2010), Regueiro et al. (2013), and Schurr et al. (2012). The map was obtained from http://mapswire.com.

To summarize, although the genetic signatures of contemporary Amerindians do not totally reflect the diversity of Late Pleistocene expansions, our data demonstrate a clear connection between Native Mexican and other indigenous Native Americans. It is important to note that these insights emerged by studying isolated populations, which allowed us to make inferences about the genetic dispersal processes with greater precision than possibly with admixed or Mestizo populations.

The genetic diversity of *Aztlan* descendants indicates that the colonization of Mesoamerica was the result of complex population dynamics beginning from the first migratory wave in the Late Pleistocene to more recent Holocene migrations. The genetic lineages brought by these different migrations were then exchanged among emerging ethnic communities in subsequent millennia, giving rise to pattern observed today. Thus, Native Mexicans, like other indigenous American peoples, do not descend from a single, unique founding population, but instead from those arriving in a series of migratory waves.

Overall, our data contribute to the understanding of the initial peopling of the Americas and the settlement of Mesoamerica and provide further impetus to new studies that will clarify the population history of Mexico. Moreover, we have revealed the extraordinary diversity of haplogroup Q-M3 in Native Mexicans, with the further phylogenetic dissection of this lineage requiring additional Y chromosome sequencing and SNP genotyping.

Limitations of the study

Despite the 21 SNPs characterized in the samples from our study populations, these were insufficient to fully elaborate the diversity within haplogroup Q-M3. In addition, owing to the modest number of Q-MEH2 Y chromosomes found, the TMRCA for this lineage may have not been accurately estimated.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Theodore G. Schurr (tgschurr@sas.upenn.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The published article includes all datasets generated or analyzed during this study in the supplemental tables.

METHODS

All methods can be found in the accompanying transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102487.

ACKNOWLEDGMENTS

We thank the enthusiastic participation of the Native American populations represented herein: Chichimecas (*Éza'r*), Nahuas, Otomies (*Hñähñús*), Popolocas, Tepehuas, and Totonacas; their collaboration made this study possible. We greatly appreciate the substantial support of *Comisión para el Desarrollo de los Pueblos Indígenas* (CDI), currently *Instituto Nacional de los Pueblos Indígenas* (INPI), especially Lic. Janitzio Durán, Lic. María Giner de los Ríos, Lic. José Luis de Jesús García Ruíz, and Lic. Olaf Ivan Corro Labra. The authors also thank CDI Delegates Eduardo Cabrera (HGO), Juan Escamilla Santiago (HGO), Federico Vázquez (Ixmiquilpan, HGO), Alonso Zurita (GTO), Verónica Jiménez (MOR), Alfredo López (QRO), Vicenta Cruz (Valle del Mezquital, HGO), Flora Elvira Cadena (VER), and Guillermo Hernández Domínguez (VER). In addition, we appreciated the enthusiastic participation of the Nahua community from Xoloco (Axtla de Terrazas, SLP), the Popoloca communities from Santa Inés Ahuatempan and San Felipe Otlaltepec (Tepexi de Rodríguez, PUE), and the Totocana community from Huehuetla (Francisco I. Madero, PUE). We especially thank the Council of Elders of the Totonaca community from El Tajin, Veracruz. We also thank the *Laboratorio de Genómica, Protéomica y Metabolómica*, and *LaNSE*-Cinvestav for assisting with the genotyping work. Finally, the authors thank Carla Santana Torres, Ph.D.; Gino F. Noris García, Ph.D.; Emma S. Calderón Aranda, Ph.D.; Indira E. Cruz Plancarte, M.Sc.; Francisco Castaneda Montes, M.Sc.; and Rafael Camacho-Mejorado, M.Sc., for their assistance with the logistics of sample collection and processing. We further thank the anonymous reviewers for their constructive comments on this manuscript. Funding for this study was provided by CONACYT grant #78239 (R.G.), the National Geographic Society, IBM, the Waitt Family Foundation, and the University of Pennsylvania (T.G.S.).

AUTHOR CONTRIBUTIONS

R.G., T.G.S., M.G.V., and M.A.M.-R. designed the study. R.G., M.G.V., T.G.S., A.O., M.A.M.-R., and J.G. coordinated and conducted fieldwork and sample collection. R.G. and P.F. performed DNA isolation. R.G. performed the Y-STR and Y-SNP genotyping. R.G., M.G.V., D.V., G.Z., and E.A.H.-T., performed the statistical analyses. R.G., T.G.S., and M.G.V. wrote the manuscript. All authors reviewed and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 16, 2020
Revised: December 8, 2020
Accepted: April 27, 2021
Published: May 21, 2021

REFERENCES

Ávila-Arcos, M.C., McManus, K.F., Sandoval, K., Rodríguez-Rodríguez, J.E., Villa-Islas, V., Martin, A.R., Luisi, P., Peñaloza-Espinosa, R.I., Eng, C., Huntsman, S., et al. (2020). Population history and gene divergence in Native Mexicans inferred from 76 human exomes. Mol. Biol Evol. *37*, 994–1006.

Barrot, C., Sanchez, C., Ortega, M., Gonzalez-Martin, A., Brand-Casadevall, C., Gorostiza, A., Huguet, E., Corbella, J., and Gene, M. (2007). Characterization of three Amerindian populations from Hidalgo state (Mexico) by 15 STR-PCR polymorphisms. Int. J. Legal. Med. *119*, 111–115. https://doi.org/10.1007/s00414-004-0488-6.

Battaglia, V., Grugni, V., Perego, U.A., Angerhofer, N., Gomez-Palmieri, J.E., Woodward, S.R., Achilli, A., Myres, N., Torroni, A., and Semino, O. (2013). The first peopling of South America: new evidence from Y-chromosome haplogroup Q. PLoS One *8*, e71390. https://doi.org/10.1371/journal.pone.0071390.

Beezley, W.H. (2011). Mexico in World History (Oxford University Press).

Benjamin, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B *57*, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031x.

Berdan, F.F., Blanton, R.E., Boone, E.H., Hodge, M.H., Smith, M.E., and Umberger, E. (1996). Aztec Imperial Strategies (Dumbarton Oaks Research Library and Collection).

Bisso-Machado, R., Jota, M.S., Ramallo, V., Paixão-Côrtes, V.R., Lacerda, D.R., Salzano, F.M., Bonatto, S.L., Santos, F.R., and Bortolini, M.C. (2011). Distribution of Y-chromosome Q lineages in Native Americans. Am. J. Hum. Biol. *23*, 563–566.

Bortolini, M.-C., Salzano, F.M., Thomas, M.G., Stuart, S., Nasanen, S.P.K., Bau, C.H.D., Hutz, M.H., Layrisse, Z., Petzl-Erler, M.L., Tsuneto, L.T., et al. (2003). Y-chromosome evidence for differing ancient demographic histories in the Americas. Am. J. Hum. Genet. *73*, 524–539. https://doi.org/10.1086/377588.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *10* (4), e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

de Azevedo, S., Bortolini, M.C., Bonatto, S.L., Hunemeier, T., Santos, F.R., and Gonzalez-Jose, R. (2015). Ancient remains and the first peopling of the Americas: reassessing the Hoyo Negro skull. Am. J. Phys. Anthropol. *158*, 514–521. https://doi.org/10.1002/ajpa.22801.

Delgado-Burbano, M.E., Scott, G.R., and Turner, C.G. (2010). The Uto-Aztecan premolar among North and south Amerindians: geographic variation and genetics. Am. J. Phys. Anthropol. *143*, 570–578. https://doi.org/10.1002/ajpa.21351.

Dryomov, S.V., Nazhmidenova, A.M., Shalaurova, S.A., Morozov, I.V., Tabarev, A.V., Starikovskaya, E.B., and Sukernik, R.I. (2015). Mitochondrial genome diversity at the Bering Strait area highlights prehistoric human migrations from Siberia to northern North America. Eur. J. Hum. Genet. *23*, 1399–1404. https://doi.org/10.1038/ejhg.2014.286.

Dulik, M.C., Owings, A.C., Gaieski, J.B., Vilar, M.G., Andre, A., Lennie, C., Mackenzie, M.A., Kritsch, I., Snowshoe, S., Wright, R., et al. (2012a). Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. Proc. Natl. Acad. Sci. U S A *109*, 8471–8476. https://doi.org/10.1073/pnas.1118760109.

Dulik, M.C., Zhadanov, S.I., Osipova, L.P., Askapuli, A., Gau, L., Gokcumen, O., Rubistein, S., and Schurr, T.G. (2012b). Mitochondrial DNA and Y-chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. Am. J. Hum. Genet. *90*, 229–246. https://doi.org/10.1016/j.ajhg.2011.12.014.

Duverger, C. (2007). El primer mestizaje: la clave para entender el pasado mesoamericano (Universidad Nacional Autonoma de Mexico, Instituto Nacional de Antropologia e Historia, Consejo Nacional para la Cultura y las Artes).

Gonzalez-Sobrino, B.Z., Pintado-Cortina, A.P., Sebastian-Medina, L., Morales-Mandujano, F., Contreras, A.V., Aguilar, Y.E., Chavez-Benavides, J., Carrillo-Rodriguez, A., Silva-Zolezzi, I., and Medrano-Gonzalez, L. (2016). Genetic diversity and differentiation in urban and indigenous populations of Mexico: Patterns of mitochondrial DNA and Y-chromosome lineages. Biodemography Soc. Biol. *62*, 53–72. https://doi.org/10.1080/19485565.2015.1117938.

Grugni, V., Battaglia, V., Perego, U.A., Raveane, A., Lancioni, H., Olivieri, A., Ferretti, L., Woodward, S.R., Pascale, J.M., Cooke, R., et al. (2015). Exploring the Y chromosomal ancestry of modern Panamanians. PLoS One *10*, e0144223. https://doi.org/10.1371/journal.pcbi.0144223.

Grugni, V., Raveane, A., Ongaro, L., Battaglia, V., Trombetta, B., Colombo, G., Capodiferro, M.R., Olivieri, A., Achilli, A., Perego, U.A., et al. (2019). Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. BMC Biol. *17*, 3. https://doi.org/10.1186/s12915-018-0622-4.

Helgason, A., Palsson, G., Pedersen, H.S., Angulalik, E., Gunnarsdottir, E.D., Yngvadottir, B., and Stefansson, K. (2006). mtDNA variation in Inuit populations of Greenland and Canada: migration history and population structure. Am. J. Phys. Anthropol. *130*, 123–134. https://doi.org/10.1002/ajpa.20313.

Hernandez-Montes, M., and Heiras-Rodríguez, C.G. (2004). Tepehuas (Comisión Nacional para el Desarrollo de los Pueblos Indígenas).

Hill, J.H. (2001). Proto-Uto-Aztecan: a community of cultivators in Central Mexico? Am. Anthropol. *103*, 913–934. https://doi.org/10.1525/aa.2001.103.4.913.

International Society of Genetic Genealogy. (2019-2020). ISOGG Y-DNA haplogroup tree 2019-2020, version 7.05.00. http://isogg.org/tree/Q/.

Instituto Nacional de Estadística, Geografía e Informática. (2019). Data from "Hablantes de Lengua Indígena en México.". https://cuentame.inegi.org.mx/poblacion.

Instituto Nacional de los Pueblos Indígenas. (2019). Data from "Atlas de los pueblos Indígenas de México.". https://atlas.cdi.gob.mx.

Josserand, J.K., Winter, M., and Hopkins, N.A. (1984). Essays in Otomanguean Culture History (Vanderbilt University, Publications in Anthropology, no. 31).

Jota, M.S., Lacerda, D.R., Sandoval, J.R., Vieira, P.P., Ohasi, D., Santos-Júnior, J.E., Acosta, O., Cuellar, C., Revollo, S., Paz-Y-Mino, C., et al. (2016). New native South American Y chromosome lineages. J. Hum. Genet. *61*, 593–603. https://doi.org/10.1038/jhg.2016.26.

Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., and Hammer, M.F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Gen. Res. *18*, 830–838. https://doi.org/10.1101/gr.7172008.

Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M.A., Jarve, M., Gerst Talas, U., Rootsi, S., Ilumae, A.-M., Magi, R., Mitt, M., et al. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res. *25*, 459–466. https://doi.org/10.1101/gr.186684.114.

Kumar, S., Bellis, C., Zlojutro, M., Melton, P.E., Blangero, J., and Curran, J.E. (2011). Large scale mitochondrial sequencing in Mexican Americans suggests a reappraisal of Native American origins. BMC Evol. Biol. *11*, 293. https://doi.org/10.1186/1471-2148-11-293.

Lastra, Y. (2006). Los otomies: su lengua y su historia (Universidad Nacional Autonoma de Mexico, Instituto de Investigaciones Antropológicas).

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S.M., Rohrlach, A., et al. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. Sci. Adv. *2*, e1501385. https://doi.org/10.1126/sciadv.1501385.

Lopez-Ramirez, Y.L., Aguilar-Velazquez, J.A., Lopez-Armenta, M., Ruiz-Hernandez, M., and Rangel-Villalobos, H. (2020). Paternal lineages and forensic parameters based on 23 Y-STRs (Powerplex® Y 23) in Mestizo males from Mexico City. Int. J. Legal Med. *134*, 199–202, 10.1007.s00414-019-02183-1.

Malyarchuk, B., Derenko, M., Denisova, G., Wozniak, M., Grzybowski, T., Dambueva, I., and Zakharov, I. (2010). Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. Ann. Hum. Genet. *74*, 539–546. https://doi.org/10.1111/j.1469-1809.2010.00601.x.

Malyarchuk, B., Derenko, M., Denisova, G., Maksimov, A., Wozniak, M., Grzybowski, T., Dambueva, I., and Zakharov, I. (2011). Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a. J. Hum. Genet. *56*, 583–588. https://doi.org/10.1038/jhg.2011.64.

Malyarchuk, B.A., Derenko, M., and Denisova, G. (2012). On the Y-chromosome haplogroup C3c classification. J. Hum. Genet. *57*, 685–686. https://doi.org/10.1038/jhg.2012.93.

Martinez-Cadenas, C., Blanco-Verea, A., Hernando, B., Busby, G.B., Brion, M., Carracedo, A., Salas, A., and Capelli, C. (2016). The relationship between surname frequency and Y chromosome variation in Spain. Eur. J. Hum. Genet. *24*, 120–128. https://doi.org/10.1038/ejhg.2015.75.

Martinez-Gonzalez, L.J., Martinez-Espin, E., Alvarez, J.C., Albardaner, F., Rickards, O., Martinez-Labarga, C., Calafell, F., and Lorente, J.A. (2012). Surname and Y chromosome in southern Europe: a case study with Colom/Colombo. Eur. J. Hum. Genet. *20*, 211–216. https://doi.org/10.1038/ejhg.2011.162.

Mata-Miguez, J., Overholtzer, L., Rodriguez-Alegria, E., Kemp, B.-M., and Bolnick, D.A. (2012). The genetic impact of Aztec imperialism: ancient mitochondrial DNA evidence from Xaltocan, Mexico. Am. J. Phys. Anthropol. *149*, 504–516. https://doi.org/10.1002/ajpa.22152.

Matos-Moctezuma, E. (2012). Escultura Monumental Mexica (Fondo de Cultura Económica).

Moreno-Estrada, A., Gignoux, C.R., Fernandez-Lopez, J.C., Zakharia, F., Sikora, M., Contreras, A.V., Acuna-Alonso, V., Sandoval, K., Eng, C., Romero-Hidalgo, S., et al. (2014). The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science *344*, 1280–1285. https://doi.org/10.1126/science.1251688.

Myres, N.M., Rootsi, S., Lin, A.A., Jarve, M., King, R.J., Kutuev, I., Cabrera, V.M., Khusnutdinova, E.K., Pshenichnov, A., Yunusbayev, B., et al. (2011). A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. Eur. J. Hum. Genet. *19*, 95–101. https://doi.org/10.1038/ejhg.2010.146.

Nichols, D.L., and Rodriguez-Alegria, E. (2017). The Oxford Handbook of the Aztecs (Oxford University Press).

Nunez, C., Baeta, M., Sosa, C., Casalod, Y., Ge, J., Budowle, B., and Martinez-Jarreta, B. (2010). Reconstructing the population history of Nicaragua by means of mtDNA, Y-chromosome STRs, and autosomal STR markers. Am. J. Phys. Anthropol. *143*, 591–600. https://doi.org/10.1002/ajpa.21355.

Olofsson, J.K., Pereira, V., Borsting, C., and Morling, N. (2015). Peopling of the North Circumpolar region-insights from Y chromosome STR and SNP typing of Greenlanders. PLoS One *10*, e0116573. https://doi.org/10.1371/journal.pcbi.0116573.

Ongaro, L., Scliar, M.O., Flores, R., Raveane, A., Marnetto, D., Sarno, S., Gnecchi-Ruscone, G.A., Alarcon-Riquelme, M.E., Patin, Etienne, Wangkumhang, P., et al. (2019). The genomic impact of European colonization of the Americas. Curr. Biol. *29*, 1–13. https://doi.org/10.1016/j.cub.2019.09.076.

Perego, U.A., Angerhofer, N., Pala, M., Olivieri, A., Lancioni, H., Hooshiar Kashani, B., Carossa, V., Ekins, J.E., Gomez-Carballa, A., Huber, G., et al. (2010). The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. Genome Res. *20*, 1174–1179. https://doi.org/10.1101/gr.109231.110.

Penaloza-Espinosa, R.I., Arenas-Aranda, D., Cerda-Flores, R.M., Buentello-Malo, L., Gonzalez-Valencia, G., Torres, J., Alvarez, B., Mendoza, I., Flores, M., Sandoval, L., et al. (2007). Characterization of mtDNA haplogroups in 14 Mexican indigenous populations. Hum. Biol. *79*, 313–320. https://doi.org/10.1353/hub.2007.0042.

Pinotti, T., Bergstrom, A., Geppert, M., Bawn, M., Ohasi, D., Shi, W., Lacerda, D.R., Solli, A., Norstedt, J., Reed, K., et al. (2019). Y chromosome sequences reveal a short Beringian standstill, rapid expansion, and early population structure of Native American founders. Curr. Biol. *29*, 149–157. https://doi.org/10.1016/j.cub.2018.11.029.

Raff, J.A., Rzhetskaya, M., Tackney, J., and Geoffrey Hayes, M. (2015). Mitochondrial diversity of Inupiat from the Alaskan North Slope provides evidence for the origins of the Paleo- and Neo-Eskimo peoples. Am. J. Phys. Anthropol. *157*, 603–614. https://doi.org/10.1002/ajpa.22750.

Raghavan, M., Steinrucken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Avila-Arcos, M.C., Malaspinas, A.S., et al. (2015). Genomic evidence for the Pleistocene and recent population history of Native Americans. Science *349*, aab3884-1-aab3884-10. https://doi.org/10.1126/science.aab3884.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, M., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extint Palaeo-Eskimo. Nature *463*, 757–762. https://doi.org/10.1038/nature08835.

Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Jr., Rasmusse, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature *506*, 225–229. https://doi.org/10.1038/nature13025.

Regueiro, M., Alvarez, J., Rowold, D., and Herrera, R.J. (2013). On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists. Am. J. Phys. Anthropol. *150*, 333–348. https://doi.org/10.1002/ajpa.22207.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., et al. (2012). Reconstructing Native American population history. Nature *488*, 370–374. https://doi.org/10.1038/nature11258.

Roewer, L., Nothnagel, M., Gusmao, L., Gomes, V., Gozalez, M., Corach, D., Sala, A., Alechine, E., Palha, T., Santos, N., et al. (2013). Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. PLoS Genet. *9* (4), e1003460. https://doi.org/10.1371/journal.pgen.1003460.

Ruiz-Linares, A. (2014). How genes have illuminated the history of early Americans and Latino Americans. Cold Spring Harb. Perspect. Biol. *7*, a008557. https://doi.org/10.1101/cshperspect.a008557.

Saavedra-Silva, E.E., and Sanchez-Salazar, M.T. (2008). Minería y espacio en el distrito minero Pachuca-Real del Monte en el siglo XIX. Investigaciones Geográficas, Boletín del Instituto de Geografía. UNAM *65*, 82–101. https://doi.org/10.14350/rig.17976.

Sahagun, Bd (2011). Historia general de las cosas de Nueva España. Códice Florentino (Cambridge University Press), pp. 1540–1585.

Sandoval, K., Moreno-Estrada, A., Mendizabal, I., Underhill, P.A., Lopez-Valenzuela, M., Penaloza-Espinosa, R., Lopez-Lopez, M., Buentello-Malo, L., Avelino, H., Calafell, F., et al. (2012). Y-chromosome diversity in Native Mexicans reveals continental transition of genetic structure in the Americas. Am. J. Phys. Anthropol. *148*, 395–405. https://doi.org/10.1002/ajpa.22062.

Santana, C., Noris, G., Meraz-Rios, M.A., Magana, J.J., Calderon-Aranda, E.-S., Munoz, M.L., and Gomez, R. (2014). Genetic analysis of 17 Y-STRs in a Mestizo population from the Central Valley of Mexico. Hum. Biol. *86*, 289–312. https://doi.org/10.13110/humanbiology.86.4.0289.

Schurr, T.G., and Sherry, S.T. (2004). Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am. J. Hum. Biol. *16*, 420–439. https://doi.org/10.1002/ajhb.20041.

Schurr, T.G., Dulik, M.C., Owings, A.C., Zhadanov, S.I., Gaieski, J.B., Vilar, M.G., Ramos, J., Moss, M.B., and Natkong, F.; Genographic Consortium (2012). Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska. Am. J. Phys. Anthropol. *148*, 422–435. https://doi.org/10.1002/ajpa.22068.

Servicio Geólogico Mexicano. (2018). Panorama Minero del Estado de Querétaro. Secretaría de Economía. http://sgm.gob.mx.

Skoglund, P., and Reich, D. (2016). A genomic view of the peopling of the Americas. Curr. Opin. Genet. Dev. *41*, 27–35. https://doi.org/10.1016/j.gde.2016.06.016.

Soustelle, J. (1993). La familia Otomi-Pame del Centro de Mexico (Fondo de Cultura Económica).

The International Genome Sample Resource. (2020). Data from "1000 Genomes Project,". https://internationalgenome.org.

Torquemada, F.Jd (2018). Los Veinte I Un Libros Rituales I Monarchia Indiana (Wentworth Press).

Vargas-Alarcon, G., Moscoso, J., Martinez-Laso, J., Rodriguez-Perez, J.-M., Flores-Dominguez, C., Serrano-Vela, J.I., Moreno, A., Granados, J., and Arnaiz-Villena, A. (2007). Origin of Mexican Nahuas (Aztecs) according to HLA genes and their relationships with worldwide populations. Mol. Immunol. *44*, 747–755. https://doi.org/10.1016/j.molimm.2006.04.014.

Wei, L.H., Huang, Y.Z., Yan, S., Wen, S.Q., Wang, L.X., Du, P.X., Yao, D.L., Li, S.L., Yang, Y.J., Jin, L., et al. (2017). Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations. J. Hum. Genet. *62*, 915–918. https://doi.org/10.1038/jhg.2017.60.

Wilson, I.J., Weale, M.E., and Balding, D.J. (2003). Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. J. R. Stat. Soc. Stat. *166*, 155–188. https://doi.org/10.1111/1467-985X.00264.

Witschey, W.R.T., and Brown, C.T. (2012). Historical Dictionary of Mesoamerica (Scarecrow Press, Inc.).

Y-chromosome Consortium (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res. *12*, 339–348.

iScience, Volume *24*

Supplemental information

Y chromosome diversity in *Aztlan* descendants and its implications for the history of Central Mexico

Rocío Gómez, Miguel G. Vilar, Marco Antonio Meraz-Ríos, David Véliz, Gerardo Zúñiga, Esther Alhelí Hernández-Tobías, Maria del Pilar Figueroa-Corona, Amanda C. Owings, Jill B. Gaieski, Theodore G. Schurr, and The Genographic Consortium

TRANSPARENT METHODS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects and Methods

We conducted field research in seven different states of Mexico during 2011-12. They include Guanajuato (GTO), Hidalgo (HGO), Morelos (MOR), Puebla (PUE), Querétaro (QRO), San Luis Potosí (SLP), and Veracruz (VER). We enrolled adult males between the ages of 18 and 65 years old who identified themselves as Native Mexicans and belonging to populations speaking languages belonging to one of three different linguistic families. They include the Nahuatl (Uto-Aztecan); Chichimeca (*Úza'*, plural *Éza'r*), Otomí (*Hñähñú*), and Popoloca (Oto-Manguean); and Tepehua and Totonaca (Totonacan), with the language family being given in parentheses. *Úza'* and *Hñähñú* are the names by which the Chichimeca and the Otomí self-identify, respectively. Thus, these names have been used in the main text in order to respect these identities.

During the field expeditions, blood and saliva samples and genealogical data were collected from 289 men following informed consent. These men belonged to six Native Mexican ethnic groups that resided in 25 different localities (Table S1). A total of 236 men were unrelated participants through three generations, including 13 *Éza'r*, 71 Nahuas (44 from HGO, 7 from MOR, and 20 from SLP), 90 *Hñähñús* (4 from GTO, 81 from HGO, and 5 from QRO), 8 Popolocas, 28 Tepehuas, and 21 Totonacas (5 from PUE and 16 from VER). Five individuals who identified themselves in the genealogical survey as a Mestizos were excluded from the analyses. As a result, a total of 231 males participated in this study (Table S2).

The University of Pennsylvania Institutional Review Board #8, the Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav-IPN), and

the Comisión Nacional para el Desarrollo de los Pueblos Indígenas, now named Instituto Nacional de los Pueblos Indígenas, approved the data and sample collection protocol for this project.

METHOD DETAILS

Molecular Genetic Analysis

DNA isolation

DNA from peripheral blood leukocytes and collected in EDTA-Vaccutainer® tubes was isolated using Gentra PureGene Blood Kit (Qiagen) following the recommendations of the manufacturer. Saliva cells were collected in 15mL polypropylene tubes with commercial mouthwash (Scope®) from which buccal cells were pelleted by centrifugation with an Eppendorf centrifuge 5408R using a swing-bucket A-4-44 rotor (Eppendorf AG) at 2,935*g* for 5 min. The buccal cells pellets were incubated with Cell Lysis Solution and Proteinase K (Qiagen) at 56ºC overnight followed by the DNA isolation as previously described.

Genotyping

All samples were genetically characterized using nineteen Y-STR (Y-chromosome short tandem repeats) using the AmpFℓSTR Y-filer PCR amplification kit (Applied Biosystems) and two additional Y-STRs (DYS388 and DYS426) in a Custom STR-SNP assay (Applied Biosystems). In addition, six fragment length polymorphisms (M17, M60, M91, M139, M175, and M186) were amplified using a custom STR-SNP assay (Applied Biosystems). The samples were amplified using a Verity 96-Well Fast Thermal Cycler (Applied Biosystems) The resulting amplicons were run on a ABI Prism 3130XL and fragment sizes mapped with the GeneMapper

ID v.3.2 software (Applied Biosystems). Each sample was further characterized using 21 SNPs (Table S3) with custom TaqMan genotyping assays (Applied Biosystems,) which were amplified in a C1000 Touch Thermal Cycler (Bio-Rad). The Y Chromosome Consortium-2008 (YCC) and International Society of Genetic Genealogy 2019-20 recommendations were used to determine the haplogroup nomenclature (Karafet et al., 2008; International Society of Genetic Genealogy, 2020). Non-Amerindian haplogroups were assigned to haplogroups using Y-STR haplotypes with a Bayesian predictor software; assignments ~ 80% were included in the analyses (Table S5) (Athey, 2005).

Comparative Data

Y-SNP and Y-STR data from Native Mexican populations were compared to data available from published reports (Table S7). A total of 158 populations (n = 1811) from Asia, North, Central, and South America were included in the database and used for further analysis. Non-Amerindian data from Native Mexican populations were similarly compared to published data from indigenous and non-native populations (Table S6). They included 2798 individuals from Mexico, Europe (Spain, France, and Italy), the Kingdom of Bahrain, Tunisia, and Morocco.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data Analysis

A number of statistical methods were used to analyze the Y-STR data. They included haplotype diversity (*h*), pairwise differences, and analysis of molecular variance (AMOVA) which were made with Arlequin v.3.5 (Excoffier and Lischer, 2010) using 1000 permutations. Intrapopulation variances (*Vp*) were estimated as described in Kayser et al., (2001). Genetic

distances (R_{ST} values) were estimated with Arlequin v.3.5 (Excoffier and Lischer, 2010) using an exact test of population differentiation with 10,000 steps in Markov chain and 1000 demorization steps and visualized in a multidimensional scaling plot (MDS) using SPSS v.11 (IBM SPSS, 2001).

NRY Phylogeography

The phylogenetic relationships among Y-STR haplotypes were constructed using median-joining (MJ) and reduced-median options in Network v.5.0 (Bandelt et al., 1999) and visualized with Network Publisher. The network analyses were generated using 17-YSTR; DYS385a/b was excluded from all the analyses performed given that it represents a duplicate STR locus. DYS839II value was obtained using the subtraction between DYS389II and DYS389I. The Y-STR loci were weighted using the inverse of variance for each one of them.

The demographic history of the populations, coalescence times, and mutation rates were inferred using Beast2 v.2.4.3, and Batwing (Bouckaert et al., 2014; Wilson et al., 2003). The coalescent times for founding lineages were assessed using Network v.5.0 software (Bandelt et al., 1999) with ρ-statistics. A rate of one mutation every 453 years was used; this estimate was generated by taking the inverse per generation mutation rate of each locus multiplied by the number of loci and by generation time, or 25 years (Chandler, 2006). Each haplotype was connected to all other haplotypes from which it differed by one repeat unit step at a single microsatellite locus.

Markov Chain Monte Carlo (MCMC) methods were also used to infer the times of more recent common ancestor (TMRCA) and the Y-chromosomal lineage histories. For the inference using Batwing software (Bayesian Analysis of Trees With Internal Node Generation) (Wilson et

al., 2003) a constant population growth size model was used. Other settings for the Bayesian methods included the following: gamma distribution (1.47, 2130) for the microsatellite mutation rate (Zhivotovsky et al., 2004), gamma (1, 0.0001) for the initial effective population size, and gamma distribution (2,400) for the population growth rate alpha *per* generation as described by Contu et al. (2008). Results obtained with this software were post-processed using the Batwing library implemented by the authors in the R software (R Core Team, 2019). Despite other studies using 25 years as generation time (Thomson et al., 2000) we explored different times (i.e., 18, 20, and 25 years) in order to obtain more reliable values.

The Phylogenetic-Coalescent analysis multi-locus performed in Beast2 v.2.4.3 (Bouckaert et al., 2014) and the implementing phylogenetic analysis for microsatellites (BEASTvntr) (Wu and Drummond, 2011) allowed inference of the TMRCA based on the allelic frequencies of haplogroup-integrated microsatellites. Estimates for Q-M3 and Q-L54 haplogroups were determined in Arlequin v3.5 (Excoffier and Lischer, 2010). Mutation rates for the microsatellites used (DYS19 2.24, DYS389-I 2.93, DYS389-II 4.12, DYS390 2.11, DYS391 2.45 DYS392 5.19, DYS393 1.05, DYS437 1.22, DYS438 3.75 DYS439 5.45 DYS448 1.52, DYS456 4.29, DYS458 6.36, DYS635 6.36, YGATA-H4 3.03), relaxed clock model, and Yule tree model were the priors for these analyses.

Two reference points were used to tree calibration. The first one located in the node of haplogroup Q-M3 dated to 20,000 years BP, based on mitochondrial lineages reported by Schurr and Sherry (2004). The second, located in haplogroup Q-L54 was dated at 8720 years BP, according to the last connection with the Altai population from which this haplogroup is derived (Dulik et al., 2012b). Three independent runs of 10 million MCMCs were conducted, with a 10% burn-in, and resampling every 10,000 states. The coverage of each run was verified in Tracer

v.1.7.1 (Rambaut and Drummond, 2007) with a minimum ESS value of 200; the results of each run were combined in LogCombiner v.1.7 (Drummond and Rambaut, 2007). Finally, the maximum credibility tree was obtained with the program TreeAnnotator v.1.6.1 (Helfrich et al., 2017), and visualized in FigTree v.1.4.4 (Rambaut, 2018).

SUPPLEMENTAL REFERENCES

Athey, T.W. (2005). Haplogroup prediction from Y-STR values using an allele-frequency approach. J. Genet. Geneal. *1*, 1-7.

Bandelt, H.J., Forster, P., and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. *16*, 37-48. https://doi.org/10.1093.oxfordjournals.molbev.a026036.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., and Drummond, A.J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. *10*: e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

Chandler, J.F. (2006). Estimating Per-Locus Mutation Rates. J. Genet. Geneal. *2*, 27-33.

Contu, D., Morelli, L., Santoni, F., Foster, J. W., Francalacci, P., and Cucca, F. (2008). Y-chromosome based evidence for pre-Neolithic origin of the genetically homogeneous but diverse Sardinian population: Inference for association scans. PLoS ONE *3*, e1430. https://doi.org/10.1371/journal.pone.0001430.

Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. *7*, 214-221. https://doi.org/10.1186/1471-2148-7-214.

Dulik, M.C., Zhadanov, S.I., Osipova, L.P., Askapuli, A., Gau, L., Gokcumen, O., Rubistein, S., and Schurr, T.G (2012b). Mitochondrial DNA and Y-chromosome variation provides

evidence for a recent common ancestry between Native Americans and Indigenous Altaians. Am. J. Hum. Genet. *90*, 229-246. https://doi.org/10.1016/j.ajhg.2011.12.014.

Excoffier, L., and Lischer, H.E.L. (2010). Arlequin suite, ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Res. *10*, 564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x.

Helfrich, P., Rieb, E., Abrami, G., Lücking, A., and Mehler, A. (2018). TreeAnnotator: Versatile visual annotation of hierarchical text relations. In *Proceedings of the Eleventh Edition of the Language Resources and Evaluation Conference, May 7 - 12*, LREC 2018, Miyazaki, Japan.

IBM Statistical Package for the Social Sciences (SPSS) Statistics. (2001). SPSS for Windows, Version 11.0. Chicago, SPSS Inc.

International Society of Genetic Genealogy (2019-2020). ISOGG Y-DNA Haplogroup Tree 2019-2020, Version 7.05.00, http://isogg.org/tree/Q/.

Karafet, T.M., Mendez, F.L., Meilerman, M.B., Underhill, P.A., Zegura, S.L., and Hammer, M.F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Gen. Res. *18*, 830-838. https://doi.org/10.1101/gr.7172008.

Kayser, M., Krawczak, M., Excoffier, L., Dieltjes, P., Corach, D., Pascali, V., Gehrig, C., Bernini, L.F., Jespersen, J., Bakker, E., et al. (2001). An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. Am. J. Hum. Genet. *68*, 990-1018. https://doi.org/10.1086/319510.

R Core Team (2019). R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna).

Rambaut, A. Drummond, D.J. 2007. Tracer v1.5: MCMC Trace Analyses Tool. (http://beast.bio.ed.ac.uk/Tracer) [accessed 2019 Aug 20].

Rambaut, A. (2018) FigTree v1.4.4, A graphical viewer of phylogenetic trees. (http://tree.bio.ed.ac.uk/software/figtree/) [accessed 2019 Aug 20].

Schurr, T.G., and Sherry, S.T. (2004). Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. Am. J. Hum. Biol. *16*, 420-439. https://doi.org/10.1002/ajhb.20041.

Thomson, R., Pritchard, J.K., Shen, P., Oefner, P.J., and Feldman, M.W. (2000). Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc. Natl. Acad. Sci. U.S.A. *97*, 7360–7365. https://doi.org/10.1073/pnas.97.13.7360.

Wilson, I.J., Weale, M.E., and Balding, D.J. (2003). Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. J**.** R. Stat. Soc. Stat. *166*, 155-188. https://doi.org/10.1111/1467-985X.00264.

Wu, C.H., and Drummond, A.J. (2011). Joint inference of microsatellite mutation models, population history and genealogies using transdimensional Markov Chain Monte Carlo. Genetics. *188*, 151-164. https://doi.org/10.1534/genetics.110.125260.

Zhivotovsky, L.A., Underhill, P.A., Cinnioglu, C., Kayser, M., Morar, B., Kivisild, T., Scozzari, R., Cruciani, F., Destro-Bisol, G., Spedini, G., et al. (2004). The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am. J. Hum. Genet. *74*, 50–61. https://doi.org/10.1086/380911.

Rambaut, A. Drummond, D.J. 2007. Tracer v1.5: MCMC Trace Analyses Tool. (http://beast.bio.ed.ac.uk/Tracer) [accessed 2019 Aug 20].

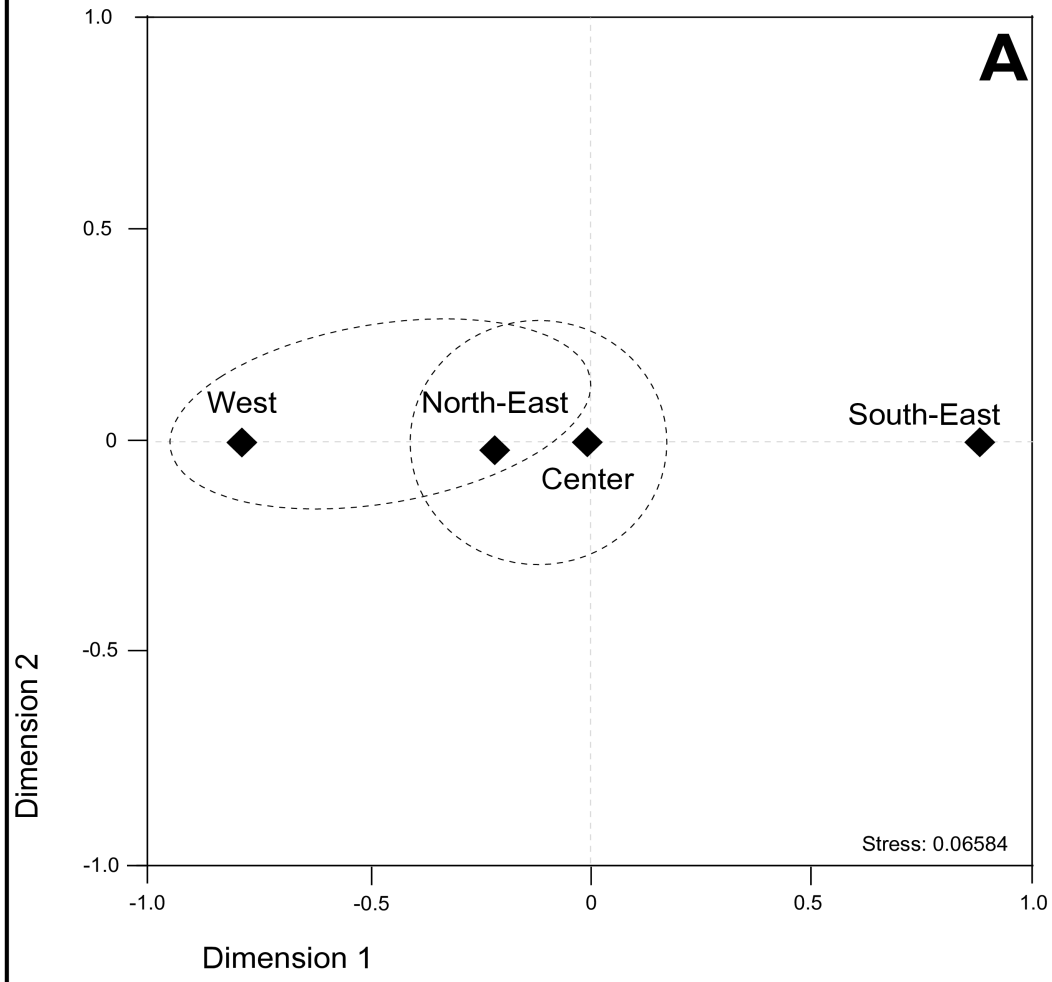
SUPPLEMENTAL INFORMATION TEXT

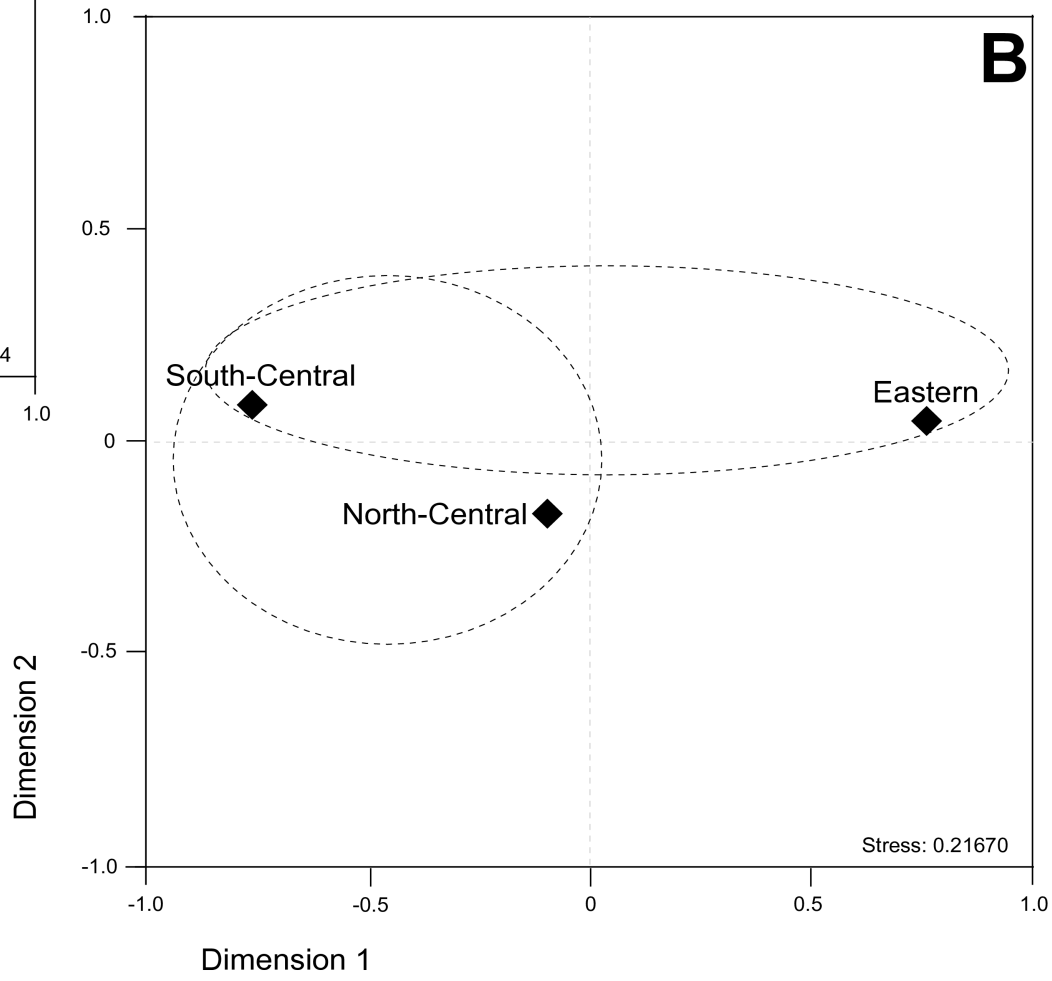
The Genographic Consortium. Syama Adhikarla[a], Christina J. Adler[b], Elena Balanovska[c], Oleg Balanovsky[c] , Jaume Bertranpetit[d] , Andrew C. Clarke[e], David Comas[d], Alan Cooper[b], Clio S. I. Der Sarkissian[b], ArunKumar GaneshPrasad[a,f], Wolfgang Haak[b], Marc Haber[d,g], Angela Hobbs[h], Asif Javed[i], Li Jin[j], Matthew E. Kaplan[k], Shilin Li[j], Begoña Martínez-Cruz[d], Elizabeth A. MatisooSmith[e], Marta Melé[d], Nirav C. Merchant[k], R. John Mitchell[l], Laxmi Parida[i], Ramasamy Pitchappan[a], Daniel E. Platt[i], Luis Quintana-Murci[m], Colin Renfrew[n], Daniela R. Lacerda[o], Ajay K. Royyuru[i], Fabrício R. Santos[o], Himla Soodyall[h], David F. Soria Hernanz[p], Pandikumar Swamikrishnan[q], Chris Tyler-Smith[r], Arun Varatharajan Santhakumari[a], Pedro Paulo Vieira[s], R. Spencer Wells[p] , Pierre A. Zalloua[g], Janet S. Ziegle[t], and Madurai Kamara[k].

[a] University, Madurai, Tamil Nadu, 625021 India; [b] University of Adelaide, South Australia, 5005 Australia; [c] Research Centre for Medical Genetics, Russian Academy of Medical Sciences, Moscow 115478, Russia; [d] Universitat Pompeu Fabra, E-08003 Barcelona, Spain; [e] University of Otago, Dunedin, 9054 New Zealand; [f] School of Chemical and Biotechnology, SASTRA University, Thanjavur 613401, India; [g] Lebanese American University, Chouran, Beirut, 1102 2801, Lebanon; [h] National Health Laboratory Service, Johannesburg 2000, South Africa; [i] IBM, Yorktown Heights, NY 10598; [j] Fudan University, Shanghai 200433, China; [k] University of Arizona, Tucson, AZ 85719; [l] La Trobe University, Melbourne, Victoria 3086, Australia; [m] Institut Pasteur, 75724 Paris Cedex 15, France; [n] University of Cambridge, Cambridge CB2 1TN, United Kingdom; [o] Universidade Federal de Minas Gerais, Belo Horizonte, 31270-010, Minas Gerais, Brazil; [p] National Geographic Society, Washington, DC 20036; [q] IBM, Somers, NY 10589; [r] The Wellcome Trust Sanger Institute, Hinxton CB10 1SA, United Kingdom; [s]

Universidade Federal do Rio de Janeiro, 21941-901, Rio de Janeiro, Brazil; and [t]Applied Biosystems, Foster City, CA 94404.

Figure S1


A
1.0
0.5
0
-0.5
-1.0
-1.0
-0.5
0
0.5
1.0
Dimension 2
Dimension 1
West
North-East
Center
South-East
Stress: 0.06584

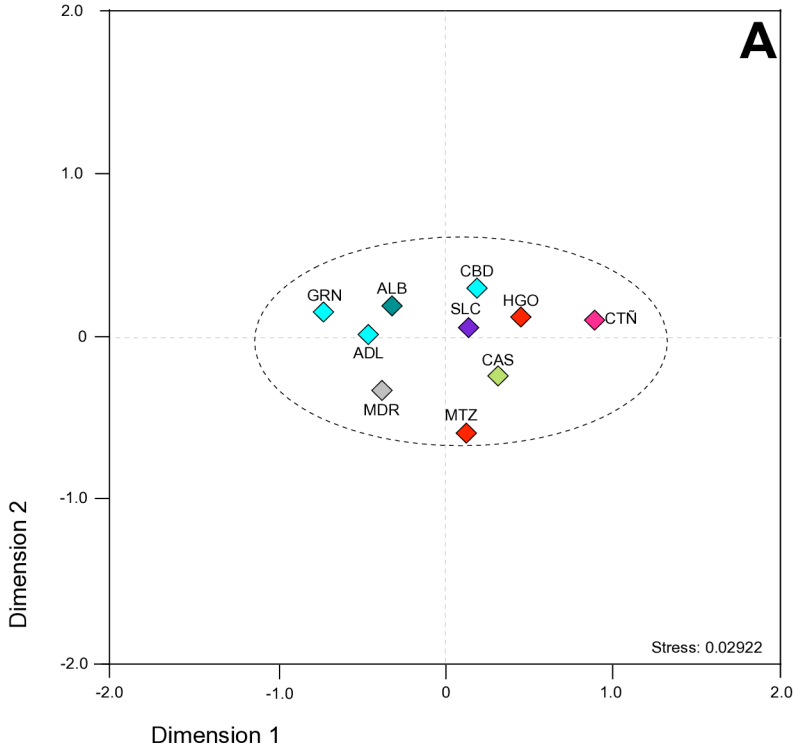

B
1.0
0.5
0
-0.5
-1.0
-1.0
-0.5
0
0.5
1.0
Dimension 2
Dimension 1
South-Central
Eastern
North-Central
Stress: 0.21670

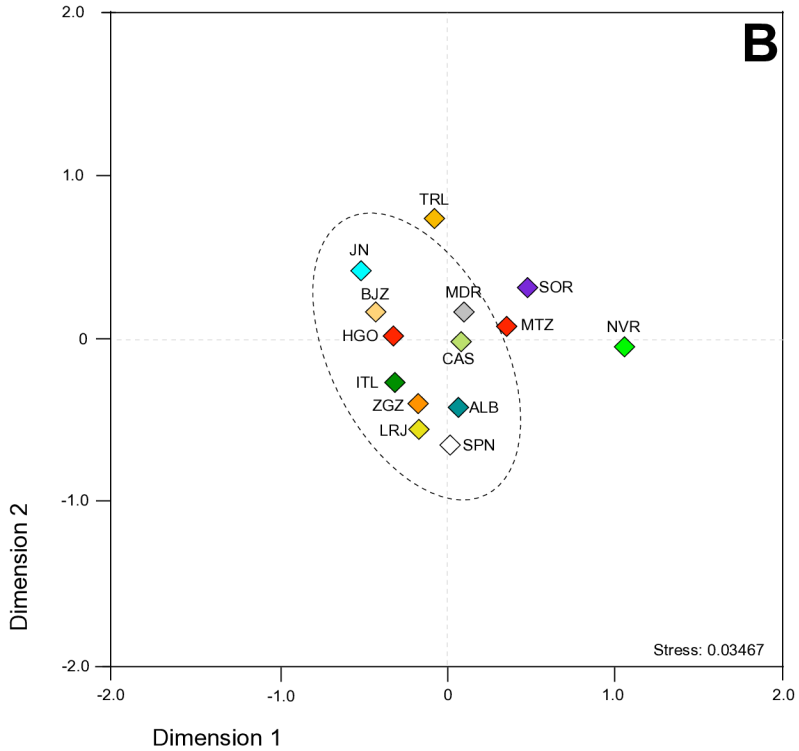
SUPPLEMENTAL FIGURE 1 LEGEND

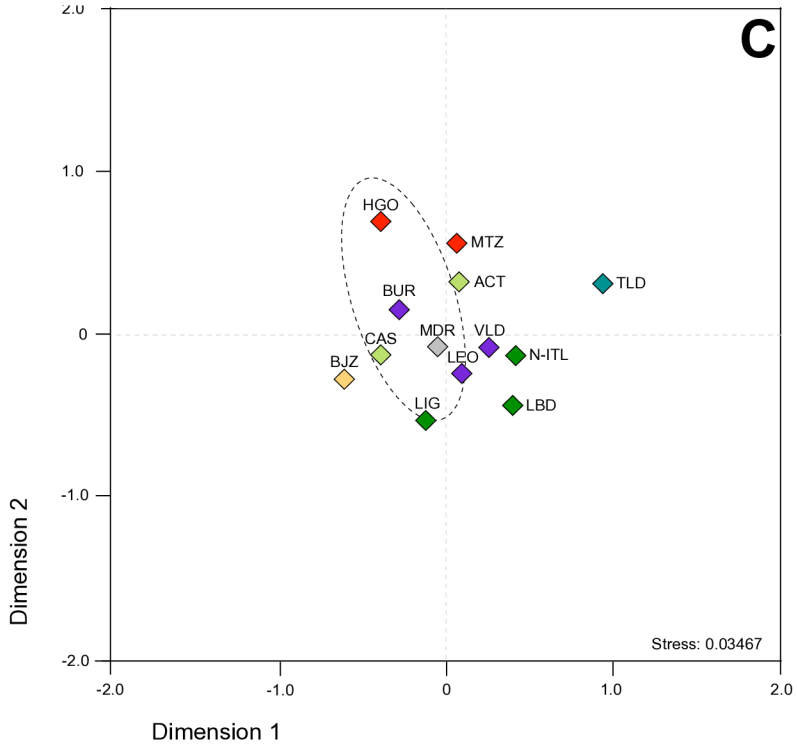
Title: A MDS plot of R_{ST} values estimated using 17 Y-STRs haplotypes from haplogroup Q in Native Mexicans populations from the Central Valley of Mexico; related to Figures 1 and 2 and Table S3.

Note: (A) With the geography-1 criteria, Center (Hidalgo: *Hñähñús*, Nahuas, Tepehuas; Morelos: Nahuas; Puebla: Popolocas, Totonacas); West (Guanajuato: *Hñähñús*, *Éza'r*; Querétaro: *Hñähñús*); South-East (Veracruz, Totonacas); North-East (San Luis Potosí, Nahuas). (B) With the geography-2 criteria, North-Central (Guanajuato: *Hñähñús*, *Éza'r*; Querétaro: *Hñähñús*; San Luis Potosí: Nahuas); Eastern (Hidalgo: *Hñähñús*, Nahuas, Tepehuas; Puebla: Popolocas, Totonacas; Veracruz: Totonacas); South-Central (Morelos: Nahuas). The dotted circle encloses populations that share no significant genetic distances.

Figure S2







Native and Mestizo Mexican populations

- Hidalgo (HGO); Mestizo (MTZ).

Spanish regions for comparisons

- Andalucía: Andalucía (ADL), Córdoba (CDB), Granada (GRN), Jaén (JN)
- Aragón: Zaragoza (ZGZ)
- Castilla La Mancha: Albacete (ALB), Toledo (TLD)
- Castilla y León: Burgos (BUR), León (LEO), Salamanca (SLC), Soria (SOR), Valladolid (VDL)
- Cataluña (CTÑ)
- Extremadura: Badajoz (BJZ)
- La Rioja: La Rioja (LRJ)
- Madrid (MDR)
- Navarra (NVR)
- Spain (SPN)
- Teruel (TRL)
- Valenciana: Alicante (ACT), Castellón (CAS)

Other Europeans

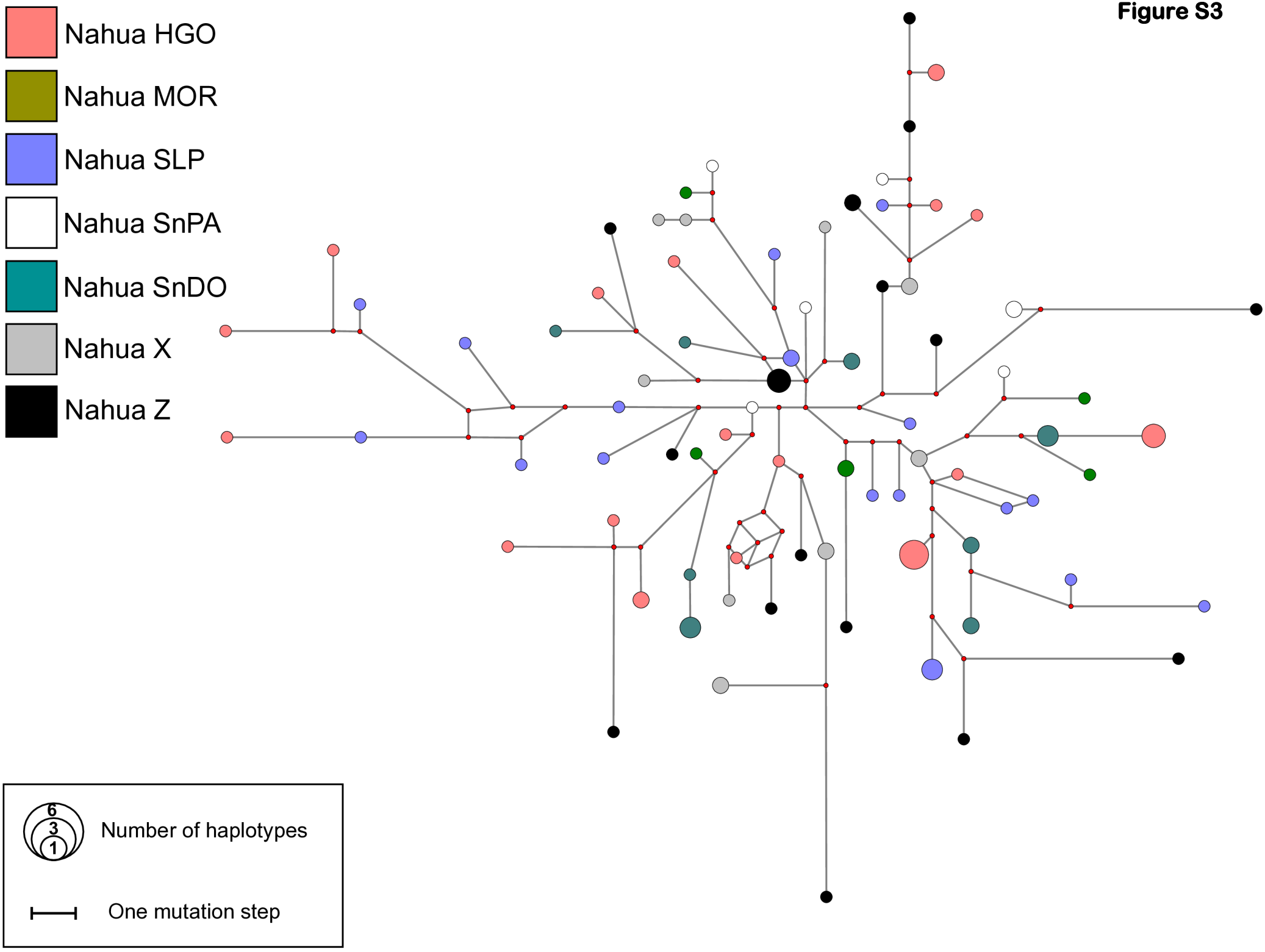
- Italy: Liguria (LIG), Lombardy (LBD), North Italy (N-ITL)

SUPPLEMENTAL FIGURE 2 LEGEND

Title: MDS plots of R_{ST} values estimated using 14 Y-STRs haplotypes from non-Native haplogroups in Native Mexican populations from the Central Valley of Mexico; related to Figure 3 and Table S7.

Note: (A) Haplogroup E; (B) Haplogroup I; and (C) Haplogroup J. The comparative data were taken from Martinez-Cadena et al., (2016), Martinez-Gonzalez et al., (2012), and Santana et al., (2014). The dotted circle encloses populations that share no significant genetic distances.

Figure S3

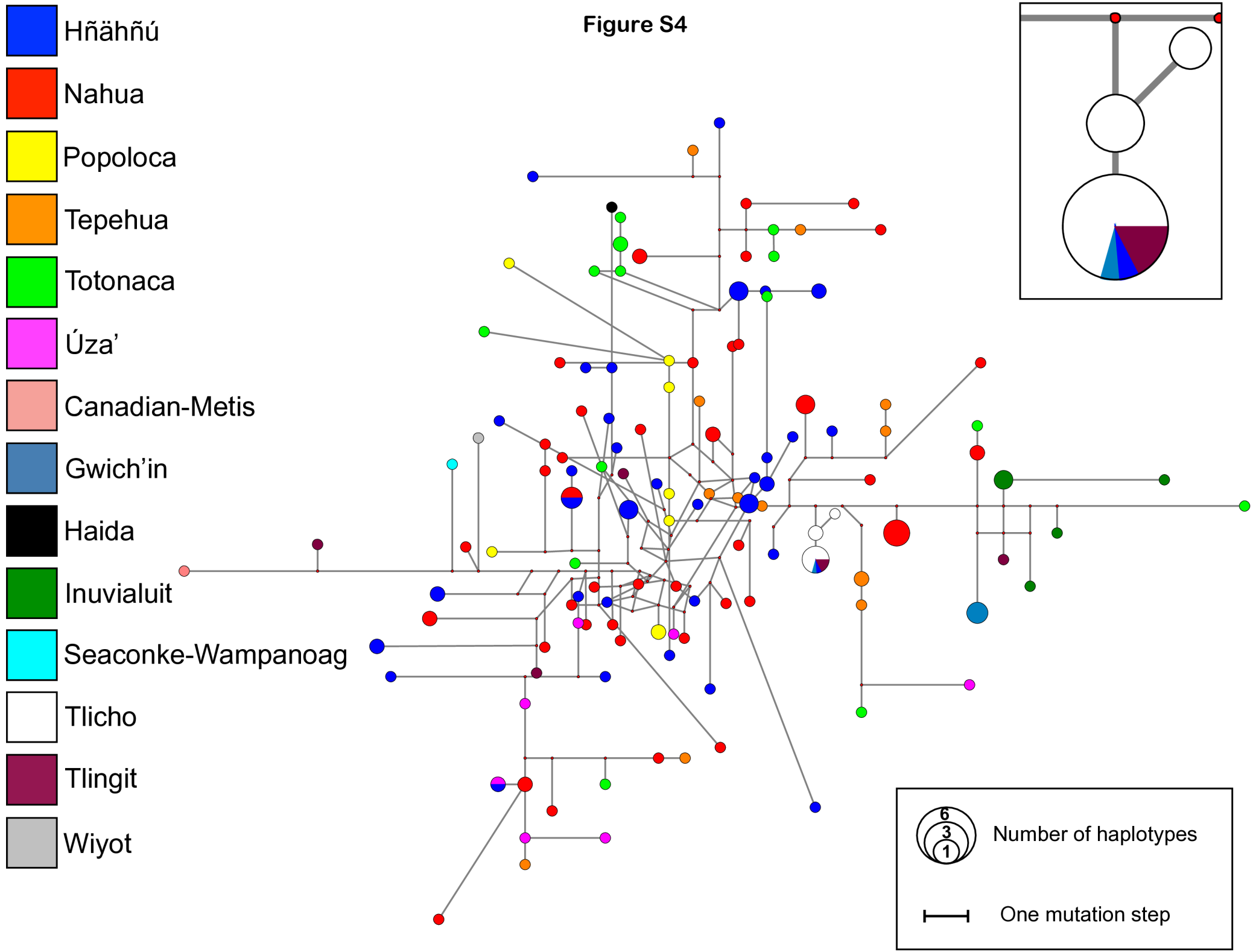

Nahua HGO
Nahua MOR
Nahua SLP
Nahua SnPA
Nahua SnDO
Nahua X
Nahua Z
6
3
1
Number of haplotypes
One mutation step

SUPPLEMENTAL FIGURE 3 LEGEND

Title: A median-joining network of Q1b1a1a (Q-M3) in Nahua populations using 17 Y-STR haplotypes. The phylogenetic analysis involved data from Nahua populations (separated by geographic region) from the present study and published data from Nahua populations living in other localities of Mexico; related to Figures 1 and 5 and Tables S3 and S9.

Note: The comparative data were obtained from Sandoval et al., (2012). HGO: Hidalgo state; SLP: San Luis Potosi state; MOR: Morelos state; SnPA: San Pedro Actopan (boundary between Mexico City and Morelos state); SnDO: Santo Domingo Ocotitlan (Morelos state); X: Xochimilco (southern Mexico City); Z: Zitlala (Guerrero state).

Figure S4


Hñähñú
Nahua
Popoloca
Tepehua
Totonaca
Úza'
Canadian-Metis
Gwich'in
Haida
Inuvialuit
Seaconke-Wampanoag
Tlicho
Tlingit
Wiyot
6
3
1
Number of haplotypes
One mutation step

SUPPLEMENTAL FIGURE 4 LEGEND

Title: A median-joining network of Q1b1a1a (Q-M3) using 14 Y-STR haplotypes from Native Mexican and indigenous North American populations; related to Figures 1 and 5 and Table S9.

Note: The box shows the shared haplotypes at the center of the network. The comparative data were obtained from Dulik et al., (2012), Regueiro et al., (2013), and Schurr et al., (2012).

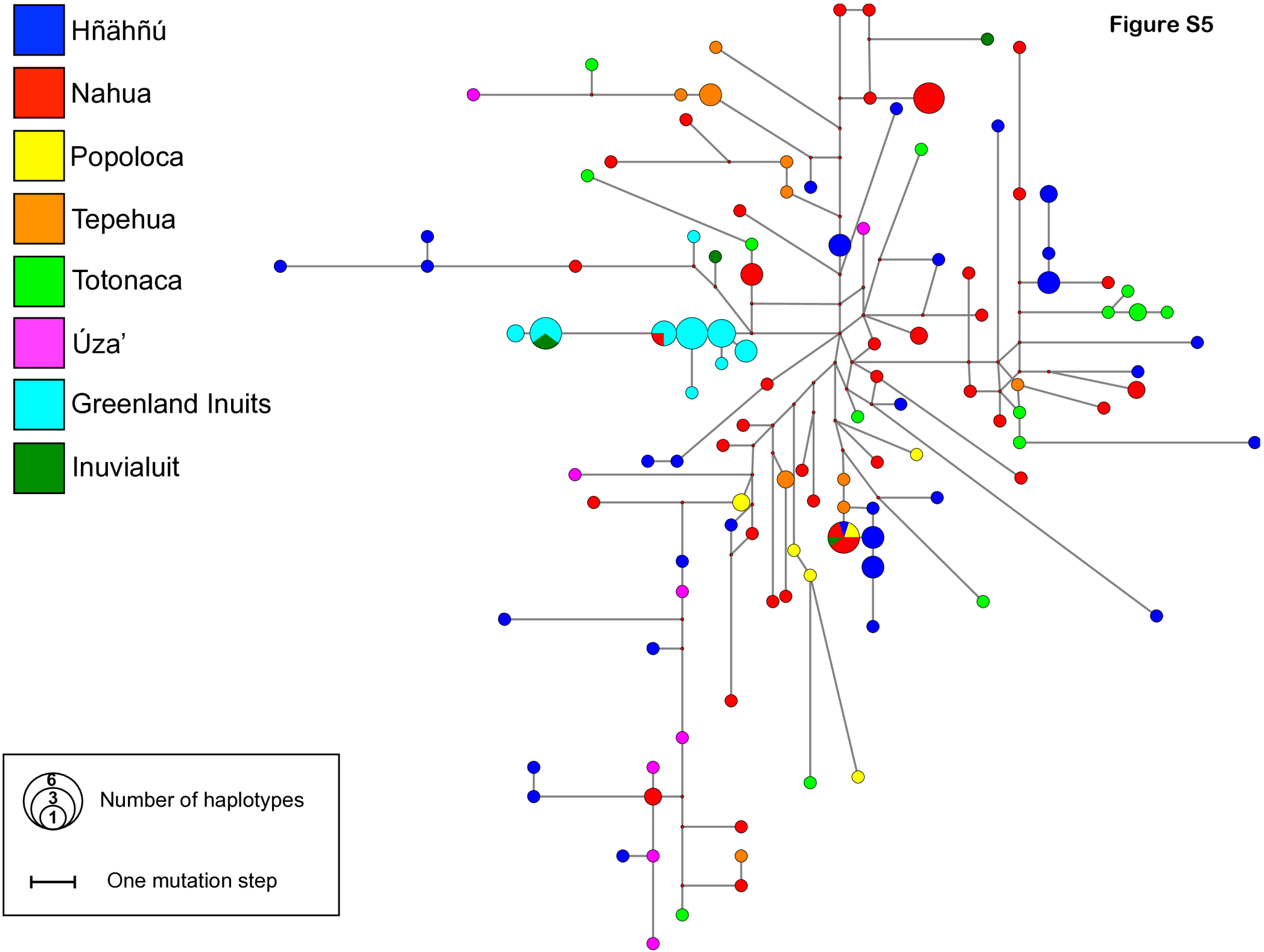


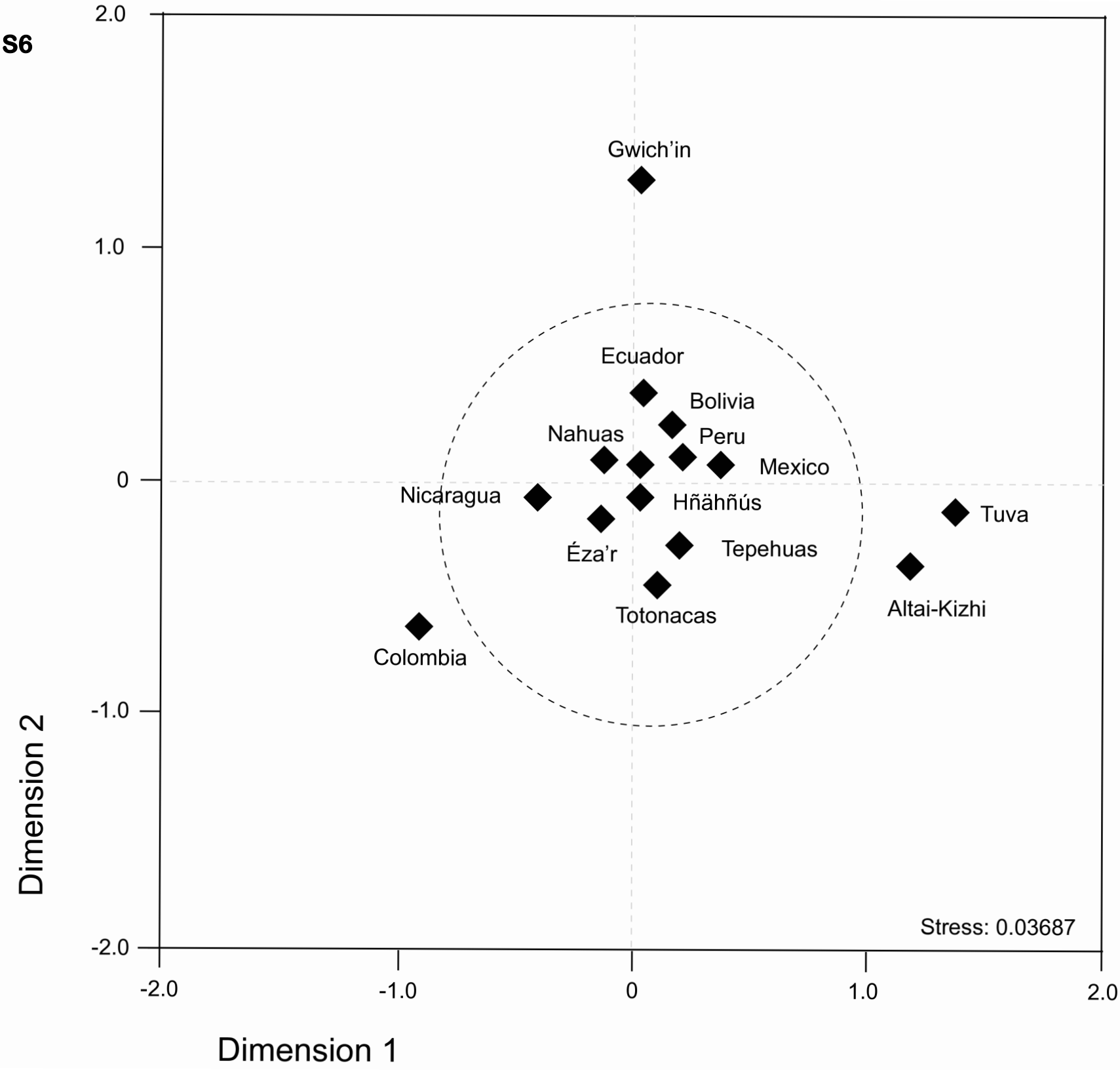
Figure S5
Hñähñú
Nahua
Popoloca
Tepehua
Totonaca
Úza'
Greenland Inuits
Inuvialuit
6
3
1
Number of haplotypes
One mutation step

SUPPLEMENTAL FIGURE 5 LEGEND

Title: A median-joining network of Q1b1a1a (Q-M3) based on 17 Y-STR haplotypes in Native Mexicans, Greenland Inuits and Canadian Inuvialuit. Greenland Inuits encompass the East and West Semersooq, Kujalleq, Qeqqata, and Qaasuitsup populations; related to Figures 1 and 5 and Table S9.

Note: The comparative data were obtained from Dulik et al., (2012) and Oloffson et al., (2015).

Figure S6


2.0
1.0
0
-1.0
-2.0
-2.0
-1.0
0
1.0
2.0
Dimension 2
Dimension 1
Gwich'in
Ecuador
Bolivia
Nahuas
Peru
Mexico
Nicaragua
Hñähñús
Éza'r
Tepehuas
Totonacas
Colombia
Tuva
Altai-Kizhi
Stress: 0.03687

SUPPLEMENTAL FIGURE 6 LEGEND

Title: A MDS plot of R_{ST} values estimated from 14 Y-STRs haplotypes in Native Mexican and other Mesoamerican populations belonging to haplogroup Q1b1a1a (Q-M3); related to Figure 5 and Tables S9 and S10.

Note: The comparative data were obtained from Battaglia et al., (2013) and Nunez et al., (2012). The dotted circle encloses populations that share no significant genetic distances.

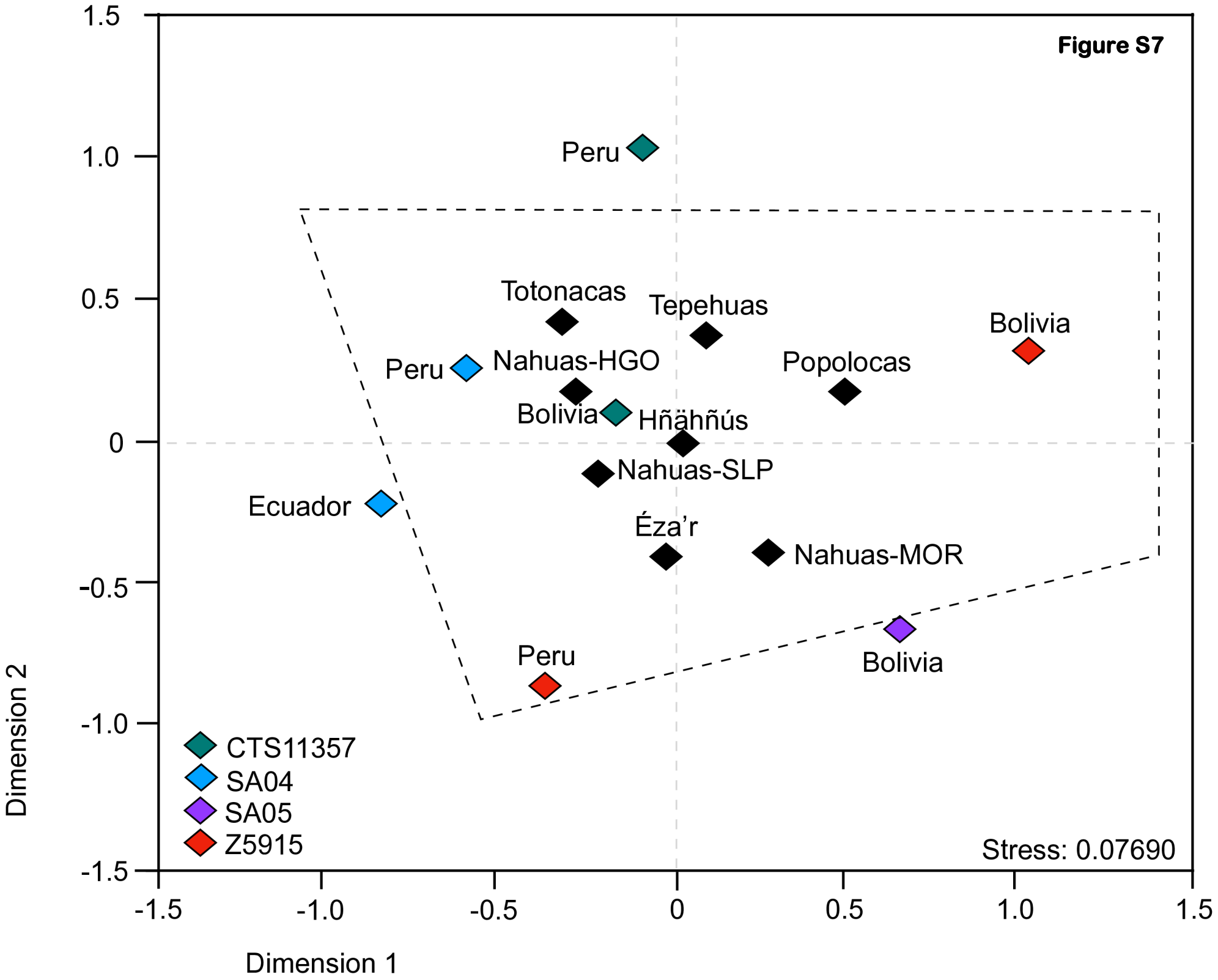


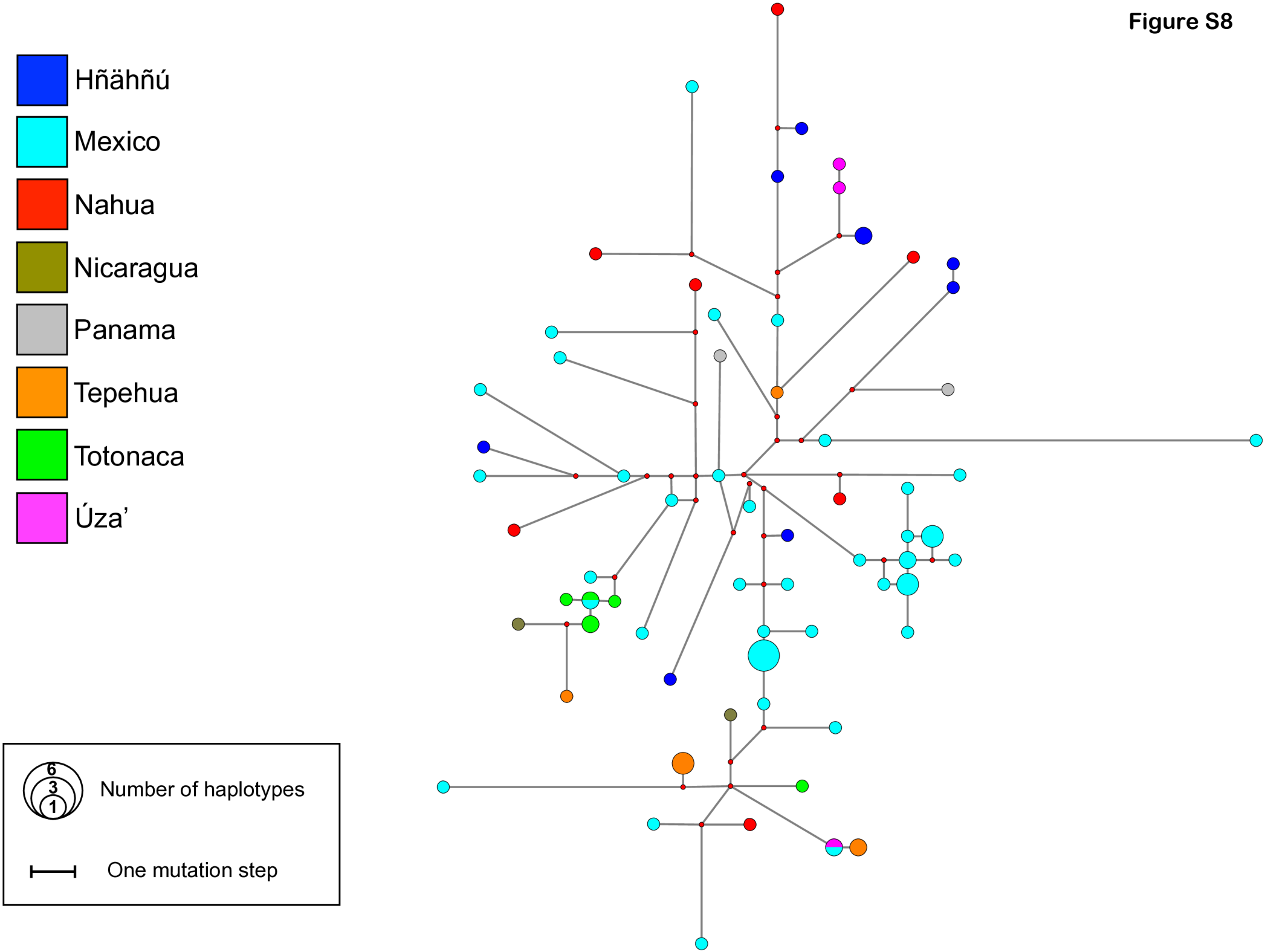
Figure S7
Peru
Totonacas
Tepehuas
Bolivia
Peru
Nahuas-HGO
Popolocas
Bolivia
Hñähñús
Nahuas-SLP
Ecuador
Éza'r
Nahuas-MOR
Peru
Bolivia
CTS11357
SA04
SA05
Z5915
Stress: 0.07690
Dimension 1
Dimension 2
1.5
1.0
0.5
0
-0.5
-1.0
-1.5

SUPPLEMENTAL FIGURE 7 LEGEND

Title: A MDS plot of R_{ST} values estimated from 14 Y-STRs haplotypes in Native Mexican and indigenous South American populations belonging to haplogroup Q1b1a1a (Q-M3); related to Figure 5 and Tables S9 and S11.

Note: The comparative data were obtained from Jota et al., (2016). The dotted circle encloses populations that share no significant genetic distances.

Figure S8

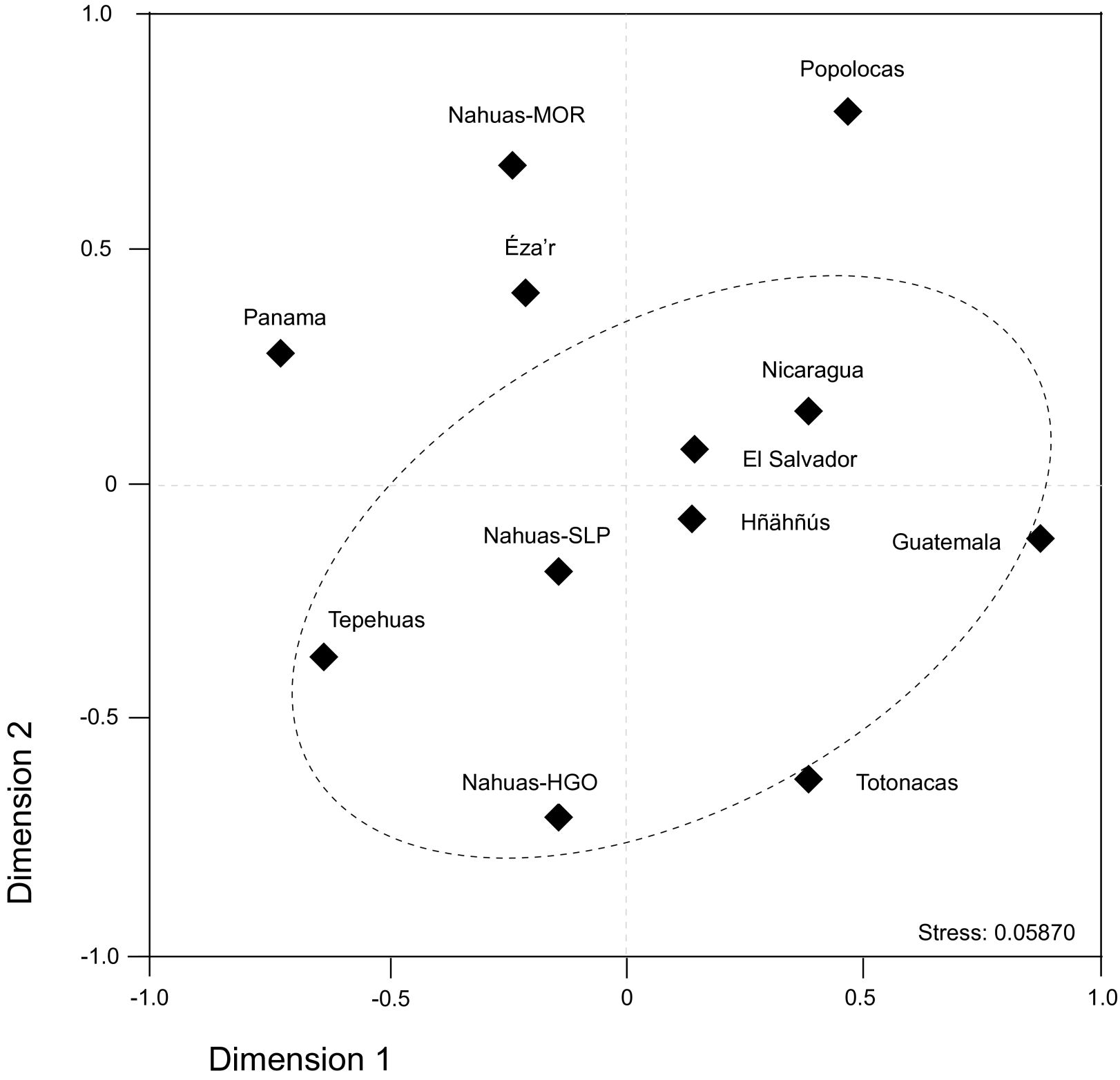

Hñähñú
Mexico
Nahua
Nicaragua
Panama
Tepehua
Totonaca
Úza'
6
3
1
Number of haplotypes
One mutation step

SUPPLEMENTAL FIGURE 8 LEGEND

Title: A median-joining network of Q1b1a2 (Q-Z777) based on 14 Y-STR haplotypes in Native Mexican and indigenous Mesoamerican populations; related to Figure 1 and Table S9.

Note: The comparative data were obtained from Battaglia et al., (2013).

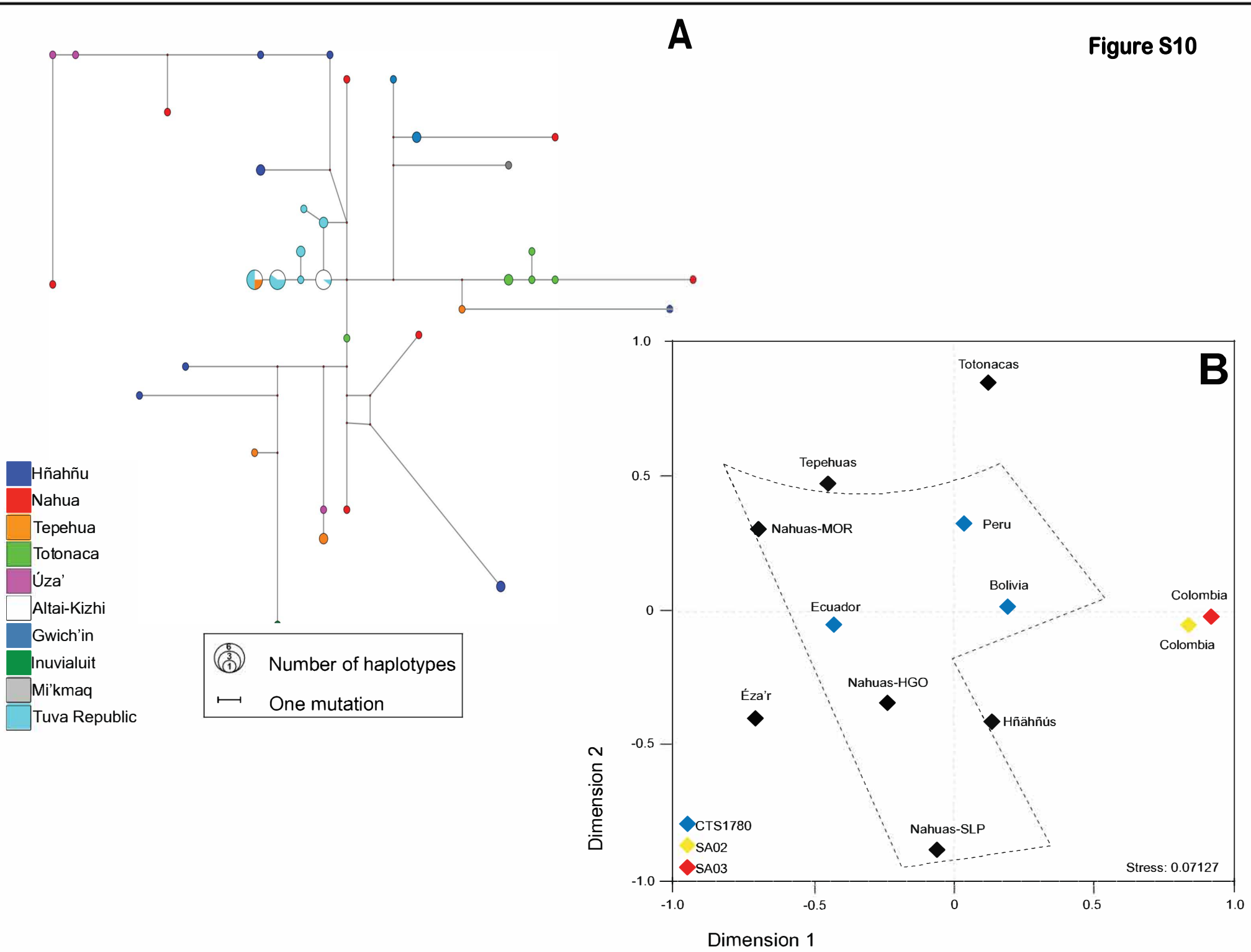
Figure S9


1.0
0.5
0
-0.5
-1.0
-1.0
-0.5
0
0.5
1.0
Dimension 2
Dimension 1
Popolocas
Nahuas-MOR
Éza'r
Panama
Nicaragua
El Salvador
Hñähñús
Nahuas-SLP
Guatemala
Tepehuas
Nahuas-HGO
Totonacas
Stress: 0.05870

SUPPLEMENTAL FIGURE 9 LEGEND

Title: A MDS plot of R_{ST} values estimated from 14 Y-STRs haplotypes belonging to haplogroup Q-Z777/Z780 in Native Mexicans, indigenous populations from North, Central and South America, and Altai-Kizhi and Tuvan populations from Siberia; related to Figure 1 and Tables S9 and S12.

Note: The comparative data were obtained from Battaglia et al., (2013), Dulik et al., (2012a), Dulik et al., (2012b), and Regueiro et al., (2013). The dotted circle encloses populations that share no significant genetic distances.


A
Figure S10
Hñahñu
Nahua
Tepehua
Totonaca
Úza'
Altai-Kizhi
Gwich'in
Inuvialuit
Mi'kmaq
Tuva Republic
Number of haplotypes
One mutation
B
Totonacas
Tepehuas
Nahuas-MOR
Peru
Bolivia
Ecuador
Colombia
Colombia
Nahuas-HGO
Éza'r
Hñähñús
Nahuas-SLP
CTS1780
SA02
SA03
Stress: 0.07127
Dimension 1
Dimension 2
1.0
0.5
0
-0.5
-1.0
-1.0
-0.5
0
0.5
1.0

SUPPLEMENTAL FIGURE 10 LEGEND

Title: Analysis of genetic diversity in haplogroup Q1b1a2 (Q-Z777/Z780); related to Figure 1 and Tables S9 and S13.

Note: (A). A median-joining network based on 15 Y-STR haplotypes in Native Mexican, indigenous North American and Asian populations; **(**B). A MDS plot of R_{ST} values estimated from 15 Y-STRs haplotypes from Q1b1a2 in Native Mexican and indigenous South American populations. The comparative data were obtained from Dulik et al., (2012a), Dulik et al., (2012b), Jota et al., (2016), and Regueiro et al., (2013). The dotted polygon encloses populations that show no significant genetic distances between them.

Table S1. Localities and ethnic groups sampled during the two expeditions.

Region	State	Localities	Geographic location	Ethnic groups	Linguistic family
NC	GTO	Rancho Uza, San Luis de la Paz	N 21º 17' 27.7" / W 100º 29' 34.5"	Chichimeca (Úza')	Oto-Manguean
CVM	HGO	Los Reyes, Acaxochitlan	N 20º 9' 19.8" / W 98º 9' 41"	Nahua	Uto-Aztecan
		La Mesa Limantitla, Huejutla	N 21º 4' 29.5" / W 98º 25' 16"		
		Tepetitla, Yahualica	N 20º 57' 38.1" / W 98º 23' 5.2"		
NC	SLP	Xoloco, Axtla de Terrazas	N 21º 28' 19.9" / W 98º 53' 45.2"		
SC	MOR	Cuentepec, Temixco	N 18º 52' 00" / W 99º 19' 38.3"		
CVM	HGO	Yonte Chico, Alfajayucan	N 20º 24' 30.2" / W 99º 19' 40.5"	Otomí (Hñähñú)	Oto-Manguean
		La Florida, Cardonal	N 20º 30' 46.9" / W 99º 00' 0.4"		
		El Alberto, Ixmiquilpan	N 20º 24' 41.9" / W 99º 12' 47.1"		
		La Lagunita, Ixmiquilpan	N 20º 39' 57.2" / W 99º 14' 27.2"		
		San Juan Tlatepexi, Mezquital	N 20º 34' 2.4" / W 98º 53' 53.1"		
		Huistícola, Metztitlán	N 20º 38' 0.4" / W 98º 49' 00"		
		Bocua, Nicolás Flores	N 20º 43' 31.3" / W 99º 10' 0.7"		
		San Miguel, San Bartolo Tutotepec	N 20º 26' 29.7" / W 98º 12' 6.9"		
		Portezuelo, Tasquillo	N 20º 29' 14.7" / W 99º 18' 25.2"		
		Pañhé, Tecozautla	N 20º 31' 2.9" / W 99º 41' 12.6"		
		Xajha, Zimapan	N 20º 45' 52.7" / W 99º 29' 38.3"		
NC	GTO	Cieneguilla, Tierra Blanca	N 21º 4' 51" / W 100º 10' 56.7"		
	QRO	Cuicillo, Amealco de Bonfil	N 20º 8' 15.8" / W 99º 56' 37"		
SC	PUE	Santa Inés Ahuatempan	N 18º 24' 53.5" / W 98º 1' 23.2"	Popoloca	
		San Felipe Otlaltepec, Tepexi de Rodríguez	N 18º 24' 20.5" / W 97º 54' 44.2"		
CVM	HGO	Huehuetla, Tenango de Doria	N 20º 27' 41.4" / W 98º 3' 59.1"	Tepehua	Totozoquean
SC	PUE	Huehuetla, Francisco I. Madero	N 18º 24' 20.5" / W 97º 54' 44.2"	Totonaca	
E	VER	Plan de Hidalgo, Papantla	N 20º 23' 58.8" / W 97º 26' 24.2"		
		Zozocolco de Hidalgo	N 20º 8' 1" / W 97º 35' 56.7"		

SUPPLEMENTAL TABLE 1 LEGEND

Title: Localities and ethnic groups sampled during fieldwork expeditions; related to Figure 1.

Note: CVM, Central Valley of Mexico; E, Eastern; NC, North-Central; SC, South-Central; GTO, Guanajuato; HGO, Hidalgo; MOR, Morelos; PUE, Puebla; QRO, Querétaro; SLP, San Luis Potosí; VER, Veracruz; N, North; W, West.

Table S2. Frequency of Y-STR haplotypes from Amerindian and non-Amerindian lineages found in the study populations.

	Ethnic Groups										
	Hñähñú	Hñähñú	Hñähñú	Nahua	Nahua	Nahua	Popoloca	Tepehua	Totonaca	Totonaca	Úza'
Hg	GTO	HGO	QRO	HGO	MOR	SLP	PUE	HGO	PUE	VER	GTO
n	4	81	5	44	7	20	8	28	5	16	13
E-M96	-	0.049	-	-	-	-	-	-	-	-	-
E1b1b	-	-	-	-	-	-	-	-	-	-	0.083
G-M201	-	0.049	-	-	-	-	-	-	-	-	-
G2a	-	0.012	-	-	-	-	-	-	-	-	-
I-M170	-	0.012	-	-	-	-	-	-	-	-	-
I2a1	-	0.012	-	-	-	-	-	-	-	-	-
I2b	-	0.012	-	-	-	-	-	0.036	-	-	-
J-M304	-	0.062	-	-	-	-	-	-	-	-	-
J1	-	0.025	-	-	-	-	-	-	-	-	-
J2a1b	-	-	-	-	-	-	-	0.036	-	-	-
L	-	-	-	-	-	-	-	0.036	-	-	-
Q-L663	-	0.025	-	-	-	-	-	-	-	-	-
Q-M3	-	0.333	0.600	0.591	0.429	0.900	1.000	0.393	1.000	0.625	0.500
Q-MEH2	-	-	-	0.091	-	-	-	-	-	-	-
Q-PV3	-	-	-	-	-	0.050	-	0.036	-	-	-
Q-Z768	1.000	0.074	0.200	0.023	0.429	0.050	-	0.071	-	-	0.083
Q-Z777	-	0.111	-	0.114	0.143	-	-	0.214	-	0.375	0.25
R1a	-	0.037	-	-	-	-	-	-	-	-	-
R1b	-	0.173	0.200	0.136	-	-	-	0.179	-	-	0.083
T	-	-	-	0.045	-	-	-	-	-	-	-
T-M70	-	0.012	-	-	-	-	-	-	-	-	-

SUPPLEMENTAL TABLE 2 LEGEND

Title: Frequency of Y-chromosome haplogroups in the study populations; related to Figure 1.

Note: GTO, Guanajuato; HGO, Hidalgo; MOR, Morelos; PUE, Puebla; QRO, Querétaro; SLP, San Luis Potosí; VER, Veracruz. “Hg” = haplogroup; “n” = sample size.

Table S4. AMOVA based on geographiy using 17 Y-STR haplotypes from haplogroup Q in Native American populations

	Percentage of variation	*p*- value
Hñähñús		
Among populations	0.92	0.378
Within populations	99.08	**< 0.0001**
Nahuas		
Among populations	4.29	**0.009**
Within populations	95.71	**< 0.0001**
Totonacas		
Among populations	0.21	0.377
Within populations	99.79	**< 0.0001**

SUPPLEMENTAL TABLE 4 LEGEND

Title: AMOVA estimates based on geography using haplogroup Q Y-STR haplotypes in Native Mexican populations; related to Table 2.

Note: Y-STR haplotypes were based on 17 STR loci. The *p* values shown in bold indicate significant values.

Table S5. RST values estimated from haplogroup Q 17 Y-STR haplotypes in Native Mexican populations categorized by ling

	Oto-Manguean	Uto-Aztecan	Totonaca-Tepehuan
Oto-Manguean	-	0.19582	**0.02213**
Uto-Aztecan	0.00525	-	**0.02213**
Totonaca-Tepehuan	0.02418	0.02553	-

SUPPLEMENTAL TABLE 5 LEGEND

Title: R_{ST} values estimated from haplogroup Q Y-STR haplotypes in Native Mexican populations categorized by linguistic affiliation; related to Figure 2.

Note: Y-STR haplotypes were defined by 17 STR loci. GTO, Guanajuato; HGO, Hidalgo; MOR, Morelos; PUE, Puebla; QRO, Querétaro; SLP, San Luis Potosí; VER, Veracruz. Oto-Manguean: *Éza'r*, *Hñahñús* (from HGO, GTO, and QRO), and Popolocas; Totonaca-Tepehua: Tepehuas and Totonacas from PUE and VER; Uto-Aztecans: Nahuas from HGO, MOR and SLP. The R_{ST} values are presented in the lower diagonal of the table, while the *p* values appear in the upper matrix. These estimates were generated with Arlequin v3.5 (10,000 permutations) and adjusted by false discovery rate test in R-software. The *p* values in boldface indicate significantly different values.

Table S6. RST values estimated from haplogroup Q Y-STR haplotypes in Native Mexican populations based on ethnicity.

	Éza'r	Hñähñús	Nahuas-HGO	Nahuas-MOR	Nahuas-SLP	Popolocas	Tepehuas	Totonacas
Éza'r	-	**0.0273**	**0.0126**	0.5889	0.1666	**0.0273**	0.0755	**0.0218**
Hñähñús	0.0926	-	0.3887	0.1249	0.4791	**0.0055**	0.2875	0.2440
Nahuas-HGO	0.1284	0.0091	-	**0.0273**	0.2440	**≤ 0.0001**	0.2440	**0.0436**
Nahuas-MOR	-0.0143	0.0736	0.1276	-	0.3250	0.2156	0.2440	0.1249
Nahuas-SLP	0.0481	0.0010	0.0164	0.0235	-	0.0553	0.1574	0.0947
Popolocas	0.1818	0.1497	0.2473	0.0965	0.1346	-	0.0218	**0.0273**
Tepehuas	0.0898	0.0113	0.0192	0.0469	0.0428	0.1724	-	0.2078
Totonacas	0.1742	0.0271	0.0448	0.1296	0.0603	0.2071	0.0413	-

SUPPLEMENTAL TABLE 6 LEGEND

Title: R_{ST} values estimated from haplogroup Q Y-STR haplotypes in Native Mexican populations based on ethnicity; related to Figure 2.

Note: Y-STR haplotypes were defined by 17 STR loci. HGO, Hidalgo; MOR, Morelos; SLP, San Luis Potosí. R_{ST} values are presented in the lower diagonal of the table, whereas *p* values appear in the upper diagonal. These estimates were obtained with Arlequin v3.5 (10,000 steps in Markov chain and 1000 demorization steps) and adjusted by false discovery rate test in R-software. The *p* values shown in boldface indicate significant differences.

Table S7. Non-Amerindian Y-STR haplotypes for the Central Valley of Mexico ethnicities and lineage data.

		FREQUENCY						Y-STR Haplotype																			SNPs	
#	Hg	HÑÄHÑÚ	NAHUA	POPOLOCA	TEPEHUA	TOTONACA	ÚZA'	DYS19	DYS385a	DYS385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	YGATAH4	DYS426	DYS388	M17 G	M343 A
3	E-M96	1						13	15	18	13	17	23	10	11	13	14	10	12	20	17	16	23	12	11	12	-	-
5		1						13	17	18	12	17	24	10	11	13	14	10	10	20	14	14	22	11	11	12	-	-
6		1						13	17	18	12	17	24	10	11	13	14	10	10	20	14	15	22	11	11	12	-	-
52		1						15	18	19	14	18	24	10	11	13	14	10	11	20	16	17	20	13	11	12	-	-
4	E1b1b						1	13	16	17	13	17	24	10	11	13	14	10	13	20	15	14	22	12	11	12	-	-
36	G-M201	1						14	14	14	14	17	22	10	11	13	16	11	11	19	14	17	22	12	11	13	-	-
45		1						15	13	16	12	17	21	10	11	15	16	10	13	24	15	16	21	11	12	12	-	-
46		2						15	13	16	12	17	21	10	11	15	16	10	13	24	16	16	21	11	11	12	-	-
41	G2a	1						15	10	17	12	17	22	10	11	13	16	10	11	21	16	17	21	12	11	12	-	-
50	I-M170	1						15	16	18	14	18	22	10	12	14	15	10	11	20	13	18	21	11	11	13	-	-
57	I2a1	1						17	12	12	14	15	25	9	11	13	15	10	12	21	14	15	22	11	11	13	-	-
49	I2b1				1			15	16	17	13	16	23	10	12	15	14	10	11	20	14	18	23	11	12	12	-	-
51		1						15	16	18	14	18	22	10	12	14	15	10	11	20	13	18	21	11	12	12	-	-
28	J-M304	1						14	11	15	13	17	23	10	11	12	14	10	12	21	15	17	21	11	12	12	-	-
34		1						14	13	18	13	16	23	10	11	13	14	10	12	20	14	19	22	11	11	17	-	-
37		1						14	14	16	13	16	23	10	11	12	15	9	12	21	15	17	21	12	11	15	-	-
38		1						14	14	18	13	16	23	10	11	12	14	10	11	20	15	18	20	11	11	16	-	-
40		1						14	16	18	12	16	24	10	11	12	14	9	12	19	14	17	20	11	11	15	-	-
27	J1	2						14	11	15	13	16	23	10	11	12	14	10	12	21	15	17	21	11	11	16	-	-
35	J2a1b				1			14	13	18	13	16	23	10	11	12	16	9	11	20	15	17	23	11	11	15	-	-
39	L				1			14	16	17	14	16	23	10	14	12	15	10	11	20	14	16	21	10	11	12	-	-
53		1						16	11	14	12	16	25	10	12	13	14	11	10	20	14	16	23	13	12	12	+	-

54	R1a1	1						16	11	14	13	16	25	10	12	13	14	11	10	20	14	16	23	13	12	12	+	-
55		1						16	11	15	13	16	25	10	12	13	14	11	10	20	14	16	23	13	12	12	+	-
1							1	13	11	15	13	16	24	10	13	13	15	10	12	19	14	17	23	12	12	12	-	+
2			1					13	14	18	12	16	22	10	13	14	14	11	12	20	15	19	22	12	12	12	-	+
7			1					14	11	13	13	16	24	11	13	13	15	12	12	19	15	17	23	12	12	12	-	+
8		1						14	11	13	13	16	24	11	14	13	15	12	11	20	15	20	23	12	12	12	-	+
9		1						14	11	13	13	16	25	10	13	13	15	12	12	18	15	18	23	12	12	12	-	+
10			1					14	11	14	12	16	23	11	13	13	15	12	11	19	15	15	24	12	12	12	-	+
11			1					14	11	14	12	16	24	11	13	13	15	12	11	19	15	15	24	12	12	12	-	+
12		1						14	11	14	12	16	24	10	13	13	15	12	11	19	16	15	23	12	12	12	-	+
14		1						14	11	14	13	16	24	11	13	13	15	12	13	19	16	17	23	13	12	12	-	+
15			1					14	11	14	13	16	23	10	14	13	14	12	13	18	15	17	23	12	12	12	-	+
16		1						14	11	14	13	16	24	10	13	13	15	12	12	19	15	18	23	12	12	12	-	+
17		1						14	11	14	13	16	24	11	13	13	15	12	12	19	15	15	23	12	12	12	-	+
18					1			14	11	14	13	16	24	11	13	13	15	12	12	19	15	16	23	12	12	12	-	+
19	R1b	1						14	11	14	13	16	24	11	13	13	15	12	11	19	17	17	23	12	12	12	-	+
20					1			14	11	14	13	16	24	11	13	13	15	12	12	19	16	16	23	12	12	12	-	+
21					1			14	11	14	13	16	24	11	13	13	15	12	12	19	16	16	24	12	12	12	-	+
22		1						14	11	14	13	16	24	11	13	13	15	13	13	20	16	17	23	11	12	12	-	+
23		2						14	11	14	13	16	24	10	13	13	15	12	12	19	16	17	24	12	12	12	-	+
24					1			14	11	14	13	16	24	11	13	13	15	12	11	19	16	16	23	12	12	12	-	+
25		1						14	11	14	14	16	24	11	13	13	15	12	12	19	16	16	23	12	12	12	-	+
26		1						14	11	14	14	16	24	11	13	13	15	12	13	19	16	17	24	11	12	12	-	+
29		1						14	11	15	14	16	24	11	14	13	15	12	12	19	16	17	23	12	12	12	-	+
30								14	11	15	15	16	24	11	13	13	14	12	12	18	15	17	23	11	12	12	-	+
31		1						14	12	14	13	16	23	11	13	13	14	12	12	19	15	16	21	12	12	12	-	+
42		1						15	11	13	13	16	24	12	13	12	15	12	12	19	15	17	23	12	11	16	-	+
43			1					15	11	14	13	16	25	10	13	13	15	12	12	19	15	17	23	11	12	12	-	+
44					1			15	11	14	14	16	24	11	13	13	15	12	12	19	14	18	23	11	12	12	-	+
56	T		2					16	14	15	12	16	24	11	14	13	14	11	11	19	15	15	22	12	12	13	-	-
47	T-M70	1						15	14	16	12	16	24	10	14	13	14	9	11	19	15	16	21	11	11	12	-	-
	Totals	**37**	**8**	**0**	**8**	**0**	**2**																					

SUPPLEMENTAL TABLE 7 LEGEND

Title: Non-native Y-STR haplotypes for Native Mexican populations from the Central Valley of Mexico; related to Figure 3.

Note: These samples were also screened for the M17 and M343 SNPs that define R1a1 and R1b, respectively. The haplogroups to which these haplotypes belong and the ethnic groups in which they appear are indicated therein.

Table S10. Rst values estimated from Q-M3 Y-STR haplotypes in Native Mexican and Mesoamerican populations.

	Éza'r	Nahuas-HGO	Nahuas-MOR	Nahuas-SLP	Hñähñús	Panama	Popolocas	Tepehuas
Éza'r	-	**≤ 0.0001**	0.9315	0.3193	0.1239	**≤ 0.0001**	**0.0396**	0.0991
Nahuas-HGO	0.1582	-	**≤ 0.0001**	0.1416	0.2803	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**
Nahuas-MOR	-0.0655	0.2018	-	0.2939	0.2201	0.0811	0.2027	0.0811
Nahuas-SLP	0.0203	0.0236	0.0294	-	0.9575	**≤ 0.0001**	**0.0240**	0.0577
Hñähñús	0.0541	0.0054	0.0405	-0.0147	-	**≤ 0.0001**	0.0631	0.1419
Panama	0.1353	0.1897	0.1419	0.1073	0.1056	-	**≤ 0.0001**	**≤ 0.0001**
Popolocas	0.2057	0.2749	0.1166	0.1503	0.1150	0.2188	-	**≤ 0.0001**
Tepehuas	0.1365	0.0976	0.1682	0.0696	0.0420	0.1340	0.2255	-
Totonacas	0.1707	0.0387	0.1802	0.0832	0.0647	0.2875	0.2724	0.1592
El Salvador	-0.1672	0.0656	-0.0244	-0.1151	-0.2920	-0.1212	-0.0117	0.0823
Nicaragua	0.0941	0.0409	0.1105	0.0429	0.0194	0.0989	0.1080	0.0839
Guatemala	0.1955	0.2796	0.3378	0.0795	0.0736	0.2778	0.1876	0.2789

Totonacas	El Salvador	Nicaragua	Guatemala
0.0396	0.9910	**0.0496**	**0.0396**
0.1416	0.9910	0.0721	0.2803
0.1297	0.9910	0.1216	0.0811
0.0240	0.9910	0.0577	0.8649
≤ 0.0001	0.9910	0.1766	0.3048
≤ 0.0001	0.9910	**≤ 0.0001**	0.1081
≤ 0.0001	0.9910	**≤ 0.0001**	0.0676
0.0451	0.9910	**≤ 0.0001**	0.0601
-	0.9910	**≤ 0.0001**	0.4460
-0.1264	-	0.6396	0.6396
0.1183	-0.2123	-	0.6396
0.1136	-0.1539	0.0871	-

SUPPLEMENTAL TABLE 10 LEGEND

Title: R_{ST} values estimated from Q-M3 Y-STR haplotypes in Native Mexican and Mesoamerican populations; related to Figure 5 and Figure S6.

Note: Y-STR haplotypes were defined by 14 STR loci. HGO, Hidalgo; MOR, Morelos; SLP, San Luis Potosí. R_{ST} values are presented in the lower diagonal, while *p* values appear in the upper diagonal. All estimates were obtained with Arlequin v3.5 (10,000 steps in Markov chain and 1000 demorization steps) and adjusted by false discovery rate test in R-software. The *p* values shown in boldface indicate significant differences.

Table S11. Rst values estimated Q-M3 Y-STR haplotypes in Native Mexican and indigenous and

	Hñähñús	Nahuas-HGO	Nahuas-SLP	Popolocas	Totonacas
Hñähñús	-	0.2626	0.9910	0.0609	0.0168
Nahuas-HGO	0.0094	-	0.3764	**≤ 0.0001**	0.0852
Nahuas-SLP	-0.0231	0.0106	-	0.0316	0.0852
Popolocas	0.1102	0.2708	0.1358	-	**≤ 0.0001**
Totonacas	0.0784	0.0546	0.0789	0.2724	-
Tepehuas	0.0485	0.0642	0.0520	0.2140	0.1013
Éza'r	0.0438	0.1574	0.0371	0.2408	0.1940
Nahuas-MOR	0.0665	0.2276	0.0472	0.1184	0.2262
Bolivia-Z5915	0.1371	0.4088	0.2598	0.2474	0.4048
Peru-Z5915	-0.0086	0.2365	0.0301	0.3218	0.2191
Peru-SA04	-0.2307	0.0576	-0.2105	0.1308	-0.0134
Ecuador-SA04	0.1418	0.2985	0.1799	0.4082	0.2443
Bolivia-SA05	0.1828	0.2672	0.2505	0.4036	0.3813
Peru-CTS11357	0.1195	0.2502	0.1910	0.3629	0.2141
Bolivia-CTS11357	-0.2038	-0.1315	-0.1907	0.1609	-0.0838

admixed South American populations.

Tepehuas	Éza'r	Nahuas-MOR	Bolivia-Z5915	Peru-Z5915	Peru-SA04
0.0748	0.2539	0.1330	0.9910	0.9910	0.9910
0.0168	**≤ 0.0001**	**≤ 0.0001**	0.9910	0.9910	0.9910
0.1099	0.2062	0.1833	0.9910	0.9910	0.9910
≤ 0.0001	**0.0168**	0.1099	0.9910	0.9910	0.9910
0.0852	**≤ 0.0001**	**0.0316**	0.9910	0.9910	0.9910
-	0.1216	**0.0466**	0.9910	0.9910	0.9910
0.0923	-	0.8218	0.9910	0.9910	0.9910
0.1278	-0.0352	-	0.9910	0.9910	0.9910
0.3622	0.4452	0.5882	-	0.9910	0.9910
0.2940	0.2833	0.5254	1.0000	-	0.9910
-0.0018	0.2037	0.3636	1.0000	1.0000	-
0.2552	0.2842	0.4013	0.6417	0.4275	0.2020
0.3987	0.4507	0.5115	0.5170	0.4647	0.5627
0.2754	0.4982	0.6158	0.8449	0.7829	0.7286
0.0017	0.0575	0.2373	0.3333	0.2414	-0.1579

Ecuador-SA04	Bolivia-SA05	Peru-CTS11357	Bolivia-CTS11357
≤ 0.0001	**≤ 0.0001**	**0.0316**	0.9910
≤ 0.0001	**≤ 0.0001**	**≤ 0.0001**	0.9910
≤ 0.0001	**≤ 0.0001**	**0.0168**	0.9910
≤ 0.0001	**≤ 0.0001**	**≤ 0.0001**	0.2834
≤ 0.0001	**≤ 0.0001**	**≤ 0.0001**	0.8142
≤ 0.0001	**≤ 0.0001**	**≤ 0.0001**	0.4309
0.0195	**≤ 0.0001**	**≤ 0.0001**	0.4438
≤ 0.0001	**≤ 0.0001**	**0.0195**	0.1578
0.1578	0.1791	0.1791	0.4438
0.2683	0.1495	0.1856	0.9907
0.4725	0.1405	0.1495	0.9910
-	**≤ 0.0001**	**≤ 0.0001**	0.1856
0.5534	-	**≤ 0.0001**	**0.0195**
0.5409	0.5703	-	0.2683
0.3815	0.4044	0.4358	-

SUPPLEMENTAL TABLE 11 LEGEND

Title: R_{ST} values estimated Q-M3 Y-STR haplotypes in Native Mexican and indigenous South American populations; related to Figure 5 and Figure S7.

Note: Y-STR haplotypes were defined by 14 STR loci. HGO, Hidalgo; MOR, Morelos; SLP, San Luis Potosí. R_{ST} values are presented in the lower diagonal, and *p* values appear in the upper diagonal. All estimates were obtained with Arlequin v3.5 (10,000 steps in Markov chain and 1000 demorization steps) and adjusted by false discovery rate test in R-software. The *p* values shown in boldface indicate significant differences. Populations with CTS11357, SA04, SA05, and Z5915 SNPs in their Q-M3 haplotypes are indicated in the table header.

Table S12. RST values estimated from Q-L54 Y-STR haplotypes in Native Mexicans, indigenous American and Asian populati

	Altai-Kizhi	Bolivia	Éza'r	Colombia	Ecuador	Gwich'in	Mexico	Nahuas	Nicaragua
Altai-Kizhi	-	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**	**0.0097**
Bolivia	0.4355	-	0.0878	**≤ 0.0001**	0.2538	**≤ 0.0001**	**≤ 0.0001**	0.1562	0.5045
Éza'r	0.5961	0.2249	-	0.0691	0.0797	0.1684	0.1957	0.8469	0.4351
Colombia	0.7941	0.5205	0.6348	-	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**	0.0565
Ecuador	0.6210	0.0188	0.2768	0.6162	-	**≤ 0.0001**	**0.0243**	0.3181	0.3514
Gwich'in	0.8485	0.4159	0.7149	0.7728	0.5812	-	**≤ 0.0001**	**0.0374**	0.0716
Mexico	0.3758	0.1453	0.1307	0.4975	0.0986	0.4001	-	0.3114	0.2115
Nahuas	0.4393	0.0437	-0.1256	0.5247	0.0115	0.3438	0.0137	-	0.6073
Nicaragua	0.6430	0.0310	0.1539	0.5926	0.0984	0.6667	0.0986	-0.0187	-
Hñähñús	0.4459	0.0752	-0.1081	0.4835	0.0866	0.3967	0.1016	-0.0272	-0.0299
Panama	0.7301	0.1446	0.0192	0.1281	0.3442	0.5909	0.1780	0.0480	0.0417
Peru	0.4036	0.0071	0.1534	0.5296	0.0217	0.4330	0.1321	0.0221	0.0825
Tepehuas	0.4840	0.1431	0.1361	0.6094	0.2218	0.7029	0.1070	0.0809	0.0506
Totonacas	0.6419	0.1172	0.2937	0.6411	0.1377	0.8419	0.1110	0.0934	0.1498
Tuva Repul	0.1221	0.4805	0.6894	0.8279	0.6963	0.8815	0.4129	0.5292	0.7393

ons.

Hñähñús	Panama	Peru	Tepehuas	Totonacas	Tuva
≤ 0.0001	**0.0180**	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**	**0.0097**
0.0390	0.2236	0.2236	**0.0390**	0.0502	**≤ 0.0001**
0.6781	0.6781	0.1105	0.3054	0.1036	**0.0207**
≤ 0.0001	0.1957	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**
0.1158	**0.0374**	0.2230	0.0608	0.1024	**≤ 0.0001**
≤ 0.0001	0.1095	**≤ 0.0001**	**0.0487**	**0.0374**	**≤ 0.0001**
0.0243	0.1216	**≤ 0.0001**	0.0811	**0.0487**	**≤ 0.0001**
0.6073	0.3027	0.3211	0.3211	0.3211	**≤ 0.0001**
0.4742	0.7297	0.3211	0.4742	0.1261	**≤ 0.0001**
-	0.4606	0.1261	0.3211	0.3211	**≤ 0.0001**
0.0205	-	0.2911	0.3211	0.1261	**≤ 0.0001**
0.0436	0.1592	-	0.1261	0.1261	**≤ 0.0001**
0.0300	0.1904	0.1067	-	0.3027	**≤ 0.0001**
0.0351	0.4161	0.0783	0.1366	-	**≤ 0.0001**
0.4958	0.7687	0.4403	0.5226	0.7278	-

SUPPLEMENTAL TABLE 12 LEGEND

Title: R_{ST} values estimated from Q-L54 Y-STR haplotypes in Native Mexicans, indigenous American and Asian populations; related to Figure 6 and Figure S9.

Note: Y-STR haplotypes were defined by 14 STR loci. R_{ST} values appear in the lower diagonal, whereas *p* values appear in the upper diagonal. All estimates were obtained with Arlequin v3.5 (10,000 steps in Markov chain and 1000 demorization steps) and adjusted by false discovery rate test in R-software. The *p* values shown in boldface indicate significant differences.

Table S13. RST values estimated Q-L54 Y-STR haplotypes in Native Mexican and indigenous South American populations.

	Nahuas-HGO	Nahuas-SLP	Tepehuas	Totonacas	Hñähñús	Nahuas-MOR	Éza'r	Colombia-SA02
Nahuas-HGO	-	0.99099	0.13215	0.09911	0.79279	0.99099	0.94144	**≤ 0.0001**
Nahuas-SLP	0.10687	-	0.31531	0.31531	0.31531	0.99099	0.31531	0.31531
Tepehuas	0.1605	0.44856	-	**0.01622**	**0.04054**	0.99099	0.09266	**≤ 0.0001**
Totonacas	0.37201	0.6	0.38677	-	**≤ 0.0001**	0.99099	**0.04805**	**≤ 0.0001**
Hñähñús	-0.00569	0.14195	0.14437	0.34757	-	0.99099	0.12613	**≤ 0.0001**
Nahuas-MOR	-0.17	1	-0.33333	0.58095	-0.18421	-	0.25676	0.25676
Éza'r	-0.03885	0.40299	0.17198	0.51678	0.12949	0.28571	-	**0.02253**
Colombia-SA02	0.6828	0.90426	0.74056	0.86278	0.59442	0.91469	0.84582	-
Colombia-SA03	0.33143	1	0.55333	0.79535	0.30651	1	0.61905	-1.25
Bolivia-CTS1780	0.02012	0.33128	0.23363	0.39472	0.09797	0.12348	0.28045	0.59066
Ecuador-CTS1780	0.04847	0.35948	0.23642	0.44603	0.12467	0.19008	0.3317	0.74386
Peru-CTS1780	-0.01793	0.33786	0.16945	0.33781	0.08173	0.00728	0.21821	0.60251

Colombia-SA03	Bolivia-CTS1780	Ecuador-CTS1780	Peru-CTS1780
0.99099	0.59459	0.4955	0.79279
0.99099	0.31531	0.31531	0.31531
0.99099	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**
0.99099	**≤ 0.0001**	**0.01442**	**≤ 0.0001**
0.99099	**0.02102**	0.09459	**≤ 0.0001**
0.99099	0.25676	0.25676	0.4
0.99099	**0.02253**	**0.02253**	**0.02253**
0.99099	**≤ 0.0001**	**≤ 0.0001**	**≤ 0.0001**
-	0.05406	0.0991	0.05406
0.46906	-	0.26126	0.26126
0.61265	0.0207	-	0.24324
0.48557	0.00631	0.01422	-

SUPPLEMENTAL TABLE 13 LEGEND

Title: R_{ST} values estimated Q-L54 Y-STR haplotypes in Native Mexican and indigenous South American populations; related to Figure 6 and Figure S10.

Note: Y-STR haplotypes were defined by 15 STR loci. HGO, Hidalgo; MOR, Morelos; SLP, San Luis Potosí. R_{ST} values appear in the lower diagonal, whereas *p* values appear in the upper diagonal. All estimates were obtained with Arlequin v3.5 (10,000 steps in Markov chain and 1000 demorization steps) and adjusted by false discovery rate test in R-software. The *p* values shown in boldface indicate significant differences. Populations with CTS1780, SA02, and SA03 SNPs in their Q-L54 haplotypes are indicated in the table header.