



**Y-chromosome diversity in Native Mexicans reveals
continental transition of genetic structure in the Americas**

Journal:	<i>American Journal of Physical Anthropology</i>
Manuscript ID:	AJPA-2011-00227.R1
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Sandoval, Karla; Stanford University, Department of Genetics; Institut de Biologia Evolutiva, Moreno, Andrés; Stanford University, Genetics Mendizabal, Isabel; Institute of Evolutionary Biology (UPF-CSIC), CEXS-UPF-PRBB, Underhill, Peter; Stanford University, Genetics Lopez-Valenzuela, Maria; Institut de Biologia Evolutiva, Peñaloza-Espinosa, Rosenda; Centro Medico Nacional, Buentello-Malo, Leonora; Universidad Nacional Autonoma de Mexico, Avelino, Heriberto; Max Planck Institute for Evolutionary Anthropology, Calafell, Francesc; Universitat Pompeu Fabra, Unitat de Biologia Evolutiva; Comas, David; Institut de Biologia Evolutiva,
Key Words:	Y-chromosome haplogroups, Native Mexican populations, Mesoamerica, genetic diversity, male lineages

SCHOLARONE™
Manuscripts

Y-chromosome diversity in Native Mexicans reveals continental transition of genetic structure in the Americas

Karla Sandoval^{1,2}, Andres Moreno-Estrada², Isabel Mendizabal¹, Peter A. Underhill², Maria Lopez-Valenzuela¹, Rosenda Peñaloza-Espinosa³, Leonor Buentello-Malo⁴, Heriberto Avelino⁵, Francesc Calafell¹, David Comas¹

¹Institut de Biologia Evolutiva (UPF-CSIC), CEXS-UPF-PRBB, Barcelona, Catalonia, Spain.

²Department of Genetics, Stanford University, Stanford, CA 94305

³Unidad de Investigación Médica en Genética Humana, Centro Médico Nacional, Siglo XXI, IMSS, 06703 Mexico City, Mexico.

⁴Instituto de Investigaciones Antropológicas, Universidad Nacional Autónoma de México, Mexico City, Mexico.

⁵Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

Corresponding author: David Comas

Tel. +34 93 316 0843

Fax. +34 93 316 0901

E-mail: david.comas@upf.edu

Key words: Y-chromosome haplogroups, Native Mexican populations, Mesoamerica, genetic diversity, male lineages

Running headline: *Y-chromosome diversity of Native Mexican Populations*

Abstract

The genetic characterization of Native Mexicans is important to understand multi-ethnic based features influencing the medical genetics of present Mexican populations, as well as to the reconstruct the peopling of the Americas. We describe the Y-chromosome genetic diversity of 197 Native Mexicans from eleven populations and 1044 individuals from 44 Native American populations after combining with publicly available data. We found extensive heterogeneity among Native Mexican populations and ample segregation of Q-M242* (46%) and Q-M3 (54%) haplogroups within Mexico. The northernmost sampled populations falling outside Mesoamerica (Pima and Tarahumara) showed a clear differentiation with respect to the other populations, which is in agreement with previous results from mtDNA lineages. However, our results point towards a complex genetic makeup of Native Mexicans whose maternal and paternal lineages reveal different narratives of their population history, with sex-biased continental contributions and different admixture proportions. At a continental scale we found that Arctic populations and the northernmost groups from North America cluster together, but we did not find a clear differentiation within Mesoamerica and the rest of the continent, which coupled with the fact that the majority of individuals from Central and South American samples are restricted to the Q-M3 branch, supports the notion that most Native Americans from Mesoamerica southwards are descendants from a single wave of migration. This observation is compatible with the idea that present day Mexico might have constituted an area of transition in the diversification of paternal lineages during the colonization of the Americas.

Introduction

The ability to reconstruct human population history using uniparental systems has motivated a deluge of genetic studies during the last decades. However, several geographic regions remain underrepresented in comprehensive surveys of human genetic variation. One such region is Mesoamerica, a wide area covering central-southern Mexico and part of Central America (including regions from Guatemala, Belize, El Salvador, Honduras, Nicaragua and Costa Rica). Despite different civilizations arising in this region, Mesoamerica comprised an interrelated cultural area that constituted one of the two centers of the New World civilization (Carmack et al. 1996; López-Austin and López- Luján 2001). The impact of post-Columbian European and African admixture with Native Mexicans is of great interest to understand disease susceptibility in Mexican populations (Bryc et al. 2010), as well as to reconstruct the history of Mesoamerica and the peopling of the Americas (Mulligan et al. 2008).

Numerous studies on this region have focused on Mestizo populations using autosomal markers (Cerdeira-Flores et al. 2002a; 1989; 1987; 2002b; Felix-Lopez et al. 2006; Rangel-Villalobos et al. 1999; Silva-Zolezzi et al. 2009; Wang et al. 2008) as well as Y-chromosome markers (Gorostiza et al. 2007; Gutierrez-Alarcon et al. 2007; Hernandez-Gutierrez et al. 2005; Licea-Cadena et al. 2006; Luna-Vazquez et al. 2008; 2005; and 2003; Rangel-Villalobos et al. 2001a; 2001b; 2008; and 2000; Rubi-Castellanos et al. 2009a; 2009b; Salazar-Flores et al. 2010; Tarazona-Santos et al. 2001; Torres-Rodriguez et al. 2006). These studies have shown a substantial contribution of Native American ancestry to extant Mexican Mestizo populations, supported by additional studies based on mtDNA sequences (González-Oliver et al. 2001; Green et al. 2000; Kemp et al. 2010; Peñaloza-Espinosa et al. 2007; Sandoval et al. 2009; Torroni et al. 1994). Other studies have also described the paternal genetic diversity in Mexican

indigenous groups (Barrot et al. 2005; Felix-Lopez et al. 2006; Kemp et al. 2010; Paez-Riberos et al. 2006; Rangel-Villalobos et al. 2000); however, a better understanding could be gained by putting this characterization into context with the continental landscape. Previous studies on continental genetic structure identified by autosomal markers (Wang et al. 2007; Yang et al. 2010), have reported a North-to-South reduction of diversity in the Americas, consistent with early studies based on Y-chromosome diversity showing patterns of low haplotypic diversity in Native American populations (Ruiz-Linares et al. 1999; Salzano 2002; Underhill et al. 2001). Concerning Y-chromosome variation, components of both Q1a3 and C3 sub-haplogroups are viewed as prehistoric Native paternal lineages in the Americas while all other haplogroups are considered to reflect the consequences of post-Columbian contact. The nearest molecular ancestors to Q and C sub-haplogroups in Native Americans are rare elsewhere, with the exception of northern Asia, where they reach frequencies of 28% and 18%, respectively (Mulligan et al. 2004). With the identification of the haplogroup Q defining M242 SNP (Seielstad et al. 2003), the age of Q haplotypes in the Americas was estimated at 15,000–18,000 ybp (Bortolini et al. 2003). Although haplogroup Q is present in 76% of Native Americans (Zegura et al. 2004), the Native American C3-M217 associated haplogroups have only been detected among the Na-Dene-speaking Tanana, Navajo, Chipewyan, and the Amerindian Cheyenne (Bergen et al. 1999; Bortolini et al. 2002; 2003; Karafet et al. 1999; Zegura et al. 2004). The higher haplotypic diversity of C-M130 lineages in eastern Siberians and Asians (Lell et al. 2002) and an estimated coalescent for C-M130 of ~25,000–30,000 ybp (Karafet et al. 1999; Underhill et al. 2000), is consistent with an Asian origin for C-M130. This inference is bolstered by other age estimates for the entire C lineage and the Native

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

American-specific C3-P39 as 27,500±10,100 and 2,550±1910 ybp, respectively (Hammer and Zegura 2002; Karafet et al. 2002).

Here, we report the haplogroup composition based on Y-chromosome STR and SNP data for 197 unrelated individuals from eleven native populations from Mexico belonging to two independent cultural and geographical areas: Tarahumara and Pima from northern Mexico (outside Mesoamerica); and Triqui, Mixtec, Otomi, Xochimilco Nahua, Zitlala Nahua, Atocpan Nahua, Ocotitlan Nahua, Purepecha, and Yucatec Maya, all from Central-South Mexico and within the Mesoamerican area. These populations belong to four different linguistic families (Campbell 1997; 1986; and Mithun 1990; 1999), Uto-Aztecan including Tarahumara, Pima and the Nahua languages; Otomanguean including Triqui, Mixtec and Otomi; Mayan including Yucatec Maya, and finally Tarascan including the isolate Purepecha.

We combined our data with additional continental populations to analyze the Y-chromosome diversity in a total of 1044 individuals from 44 Native American populations in order to determine the distribution of paternal lineages within Mexico in contrast with the continental landscape and evaluated its correlation to geographical, cultural, and linguistic factors. In addition, we contrasted our results with the female counterpart for the same populations as revealed by mtDNA analysis (Sandoval et al. 2009) and found a remarkable differential contribution between maternal and paternal lineages in Mesoamerica.

Materials and Methods

Populations studied

A total of 197 Native Mexican male individuals were analyzed: 22 Triquis, 18 Tarahumaras, 2 Mixtecs, 7 Otomis, 22 Nahuas from Xochimilco, 22 Nahuas from Zitlala, 9 Nahuas from San Pedro Atocpan, 17 Nahuas from Santo Domingo Ocotitlan, 8 Purepechas, 19 Yucatec Mayas, 51 Pimas. An additional sample of 11 Quechuas from Peru were also analyzed and used for inter population comparisons. DNAs from Pima, Yucatec Maya and Quechua were provided by the laboratory of Judith R. Kidd and Kenneth K. Kidd (Yale University, New Haven, Conn., USA) where lymphoblastoid cell lines are maintained. All samples were collected with the appropriate informed consent and all individuals were native speakers with local ancestors; their geographic location is shown in Figure 1.

Y chromosome genotyping

DNA concentration of all samples was normalized to 1ng/μl using the Quantifiler™ Human DNA Quantification Kit (Applied Biosystems). Three SNPs (M45, which defines the major haplogroup P; M242, which defines haplogroup Q; and M207, which defines haplogroup R) were genotyped in all individuals in 5μl reactions. The reactions contained 1μl of DNA template, 0.25μl of 20x SNP Genotyping Assay, 1.25μl of sterile-filtered water and 2.5μl of TaqMan® Universal PCR Master Mix. Each individual was amplified under standard conditions with TaqMan® probes (Applied Biosystems). Those individuals not assigned within the major haplogroup P, were genotyped for marker M130 (which defines haplogroup C), and a subset of 69 Native Mexicans and 7 Quechua individuals belonging to haplogroup Q were genotyped for marker M3 (Underhill et al. 1996), which defines haplogroup Q1a3a within clade Q (hereafter referred to as Q-M3). After the PCR amplification, allelic discrimination was

done with SDS Software TM v2.3 (Applied Biosystems) in order to determine the haplogroups defined by the SNPs previously described.

A total of 17 Y-STR loci (DYS19, *DYS385a/b*, *DYS389I*, *DYS389II*, *DYS389II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS437*, *DYS438*, *DYS439*, *DYS448*, *DYS456*, *DYS458*, *DYS635*, *GATA-H4*) were amplified with the AmpFISTR® Yfiler™ PCR Amplification Kit (Applied Biosystems). PCR products were discriminated in an ABI 3130xl genetic analyzer and the allele sizes were determined with GeneMapper v3.2. Software.

Statistical Analysis

STR-haplotype data was used to estimate intra population genetic diversity parameters such as: gene diversity, mean pairwise differences and theta values using Arlequin program v3.11 (Schneider et al. 2005). To compare the Mexican data with other Native American populations, haplotype information of seven Y-STRs (*DYS19*, *DYS389I*, *DYS389II*, *DYS390*, *DYS391*, *DYS392* and *DYS393*) were collected from the literature. After combining with our Native Mexicans and filtering populations with extreme low sample sizes ($N < 9$) a total of 44 populations were included in the database and used for further analyses (Supplementary Table S1). Individuals belonging to non-Native American haplogroups according to (Bortolini et al. 2003) were excluded for all estimations except for the Correspondence Analysis (see below). For some analyses, population samples were grouped into major sub-continental areas: North (675 individuals from 28 populations from Alaska to Mexico), Central (79 individuals from 4 populations from Guatemala to Panama) and South America (290 individuals from 12 populations from Colombia southwards) for a total of 1044 individuals analyzed (see Supplementary Table S1).

Population genetic structure was tested in Mexican and Native American populations with the analysis of molecular variance (AMOVA) (Excoffier et al. 1992) using the Arlequin program v3.11 (Excoffier et al. 2005). Genetic relationships among Mexican and continental populations were analyzed with two approaches. The first approach was based on genetic distances (using R_{ST}) calculated with Arlequin program v3.11 (Excoffier et al. 2005) and represented in a Multidimensional Scaling plot (MDS). The second approach was based on haplogroup absolute frequencies in order to construct a Correspondence Analysis (CoA) plot. Both analyses were performed with STATISTICA 7 package (<http://www.statsoft.com>).

Median joining networks were constructed using Network 4.5.1.0 package (Bandelt et al. 1999) for the Native American haplogroup Q (Bortolini et al. 2003). The time to the most recent common ancestor (TMRCA) was estimated for the Q lineage. Y-STRs weights were determined according to Hurles et al. (2002). The mutation rate considered was one mutation per 571 years, calibrated on the evolutionary rate of each locus reported in the Y-chromosome haplotype reference database (YHRD, www.yhrd.org) as of March 23, 2010.

Results

Paternal lineage composition in Native Mexicans

Few SNPs are needed to define the major haplogroups found in the Americas since only two major haplogroups are associated with Native American populations: haplogroup Q; and at lower frequency, haplogroup C (Zegura et al. 2004). After genotyping four Y-chromosome SNPs, the 197 Native Mexican individuals were classified into major haplogroups (Table 1). The most frequent haplogroup in Native Mexican populations was haplogroup Q (85.3%), that is defined by the M242 mutation and constitutes the major lineage among Native Americans, but is also widely distributed in North Eurasia and some Siberian populations (Karafet et al. 2008). In contrast, within clade Q, haplogroup Q-M3 (Q1a3a) is almost completely restricted to the Americas (Zegura et al. 2004). In a subset of 69 Native Mexicans additionally genotyped for Q-M3 we observed that 54% belong to Q1a3a (Table 1), whereas a much higher proportion (86%) was found in a subset of Quechua individuals, indicating that the ancestral lineages of Native Mexicans are less restricted to Q-M3 bearers compared to some South American populations. Interestingly, the population within Mexico with the largest count of ancestral states for Q-M3 was the northernmost group, namely the Pima. Conversely, the Maya group showed the largest count of derived states for Q-M3. Haplogroup C, the other major lineage found in Native Americans (so far observed exclusively in North America), was not observed in our Native Mexican sample. Other haplogroups with non-Native origin were found in our sample set, including haplogroups outside clade P (12.7%), and R (2.03%), which is the most common haplogroup in European populations (Karafet et al. 2008) (see Table 1). The presence of these non-native lineages in Native Mexicans is most plausibly the result of recent admixture since colonial times. In almost all Mexican populations the admixture

1
2
3 proportion was similar except for the Xochimilco Nahuas, where significant differences
4
5 were observed when compared to Triqui (exact test of differentiation, $p=0.003$) and
6
7 Pima ($p=0.0007$) samples. These last two populations showed a high percentage of Q
8
9 native haplogroup (100% and 96%, respectively).
10

11
12 Of the 176 individuals belonging to haplogroup Q, a total of 106 different
13
14 haplotypes (Supplementary Table S2) were observed. Table 2 summarizes the results
15
16 for the 17 STRs analyzed on 168 Native Mexican samples and the eight Quechua
17
18 samples used for comparison. Interestingly, Pima and Tarahumara, the northernmost
19
20 samples, exhibited the lowest diversity values (average genetic diversity and pairwise
21
22 differences), which are the only non-Mesoamerican populations.
23

24
25 An analysis of molecular variance (AMOVA) based on R_{st} genetic distances was
26
27 performed in order to define the population structure of Mexican populations according
28
29 to sample size, geographic and linguistic criteria (Table 3). In order to avoid biases due
30
31 to admixture of different non-native haplogroups, AMOVA estimations were performed
32
33 only with Native American haplotypes. To determine whether the four Nahua groups
34
35 had to be considered as a single group, an AMOVA was performed among the Nahua
36
37 samples. Certain heterogeneity ($p<0.05$) among Nahuas was found (10.18%). To
38
39 corroborate this result, an exact test of population differentiation was performed
40
41 ($p<0.05$). Since both results showed a clear differentiation within the Nahua
42
43 populations, for most of the analyses they were considered as four separate groups,
44
45 except for those analyses in which sample size was relevant. Particularly, Xochimilco
46
47 and San Pedro Nahuas were treated as a single group, since they do not represent
48
49 separate dialects, are close geographically, and no statistical differences were observed
50
51 when an AMOVA was performed between these two populations (3.58%, $p=0.2$).
52
53
54
55
56
57
58
59
60

The AMOVA showed that when all Mexican populations were considered as a single group (11 populations), 16.30% ($p < 0.05$) of the genetic variance was found between populations, showing a notable genetic heterogeneity among Native Mexican populations. When two groups (northern samples versus Mesoamerican samples) were compared, the AMOVA showed significant differences between the two groups: 8.43% ($p < 0.05$). Non-significant proportions of the variance (-2.17%, $p = 0.7$) were explained by the fact of belonging to one of the four language families in the sample; this suggests a lack of correlation between Y-chromosome STR haplotype diversity and linguistic affiliation. When only populations with at least 10 individuals (7 groups) were considered (Mixtec, Purepechas and Otomi were discarded due to their low sample size, and Nahuas from Xochimilco and San Pedro were considered together as a single group), 16.80% ($p < 0.05$) of the genetic variance was found between all populations, showing again notable genetic heterogeneity among Native Mexican populations. When cultural groups (northern versus Mesoamericans) were considered, significant values were found (8.22%, $p < 0.05$). When the linguistic classification (three families represented) was considered, again non-significant values were found (-3.09%, $p = 0.8$).

In the MDS calculated with Mexican populations that only belong to haplogroup Q based on R_{ST} genetic distances (obtained from the full set of 17 Y-STR loci), Tarahumara and Pima were located apart from the rest in the first dimension, whereas the rest of the populations did not show a clear clustering (Figure 2A). Additionally, for the Correspondence Analysis (CoA) performed with Mexican populations (Supplementary Figure S1), only samples with more than ten individuals (only 7 groups) were considered (Mixtec, Otomi and Purepecha were excluded). The first dimension, which explains 88.19% of the variation, separates Q-M242 and Q-M3 haplogroups. Pimas and Mayas were projected closer to Q-M242 and Q-M3,

1
2
3 respectively. The second dimension was driven by the presence of non-native (R and
4 non-P) lineages in some populations, mainly Tarahumara, Nahua from Xochimilco, and
5 Maya, denoting higher proportions of European admixture in these populations
6
7
8
9
10 (Supplementary Figure S1).

11
12 *Mexican populations within the American genetic landscape*

13
14 To compare the genetic diversity found in Mexico with the rest of Native
15
16 American populations at a continental scale, a comprehensive database of previously
17
18 reported Y-STR and SNP data was constructed and combined with our data. A total of
19
20 44 populations with available data for a common subset of 7 STRs were included in the
21
22 database. Out of those, 27 populations (7 from North America, 12 from Mexico, 4 from
23
24 Central America, and 4 from South America) had all three major native haplogroups
25
26 reported (C-M130, Q-M242, and Q-M3), and were thus used to perform a
27
28 Correspondence Analysis (CoA) based on absolute haplogroup frequencies (Figure 3).
29
30 The first dimension of the CoA (60.53%) separates Q-M3 from the rest, and the second
31
32 dimension (39.47%) C-M130 from the rest. In agreement with the known distribution of
33
34 haplogroup C, we observed that the two northernmost populations of this panel
35
36 (Chippewa and Sioux) cluster next to C-M130 and the rest of populations show varying
37
38 proportions of Q-M242 and Q-M3. It is noteworthy that Native Mexicans are the only
39
40 regional group with populations represented both next to the Q-M242 cluster and the Q-
41
42 M3 cluster. In contrast, all Central and South American samples were grouped
43
44 significantly closer to the Q-M3 haplogroup (Figure 3).
45
46
47
48

49
50 A broader picture was obtained when MDS analyses were performed at a
51
52 continental scale using R_{ST} genetic distances based on 7 STRs from the full panel of 44
53
54 populations (Figure 2B). In the two-dimension plot (stress value = 0.142) it is clear that
55
56 the first dimension is driven by a strong differentiation of the northernmost populations
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

from the rest: Dogrib and Tanana, both belonging to the same linguistic family (Athapaskan), Aleuts (Alaska Peninsula), Minnesota Chippewa (Northeast USA), and Sisseton Wahpeton-Sioux (Northeast USA). The second dimension appears to be driven by highly isolated tribes, such as the two Yanomami groups from the Amazon, and Bari from Venezuela. No clear clusters were observed among the rest of populations.

In order to formally test whether Native American samples clustered, an AMOVA considering the American database was performed (Table 3). When samples were considered as a single group, a high degree of genetic heterogeneity between populations 20.47% was found ($p<0.05$). In order to ascertain whether this heterogeneity was caused by geographical factors, three continental groups (North, Central and South) were considered. The AMOVA showed that non-significant 1.22% ($p>0.1$) of the genetic variance was due to differences between these three sub-continental groups, whereas the genetic variance within the three groups remained highly significant (19.73%; $p<0.01$), suggesting an important genetic heterogeneity within the sub-continental areas and a lack of continental structure. One plausible explanation for this lack of structure is that when considering all territories from Mexico northwards as a part of the same group, the existing differences between the northern populations and the rest is homogenized by the inclusion of Mesoamerican populations. In fact, when continental populations were grouped into only two categories (North America versus all other), highly significant differences were obtained (18.51%; $p<0.01$), in agreement with the observed clustering pattern in the MDS analysis (Figure 2B). Furthermore, when populations were grouped into 6 major cultural areas (North America, Aridoamerica, Mesoamerica, Central America, Amazon, and Andes), the differences among groups (6.63%) remained significant ($p<0.05$).

1
2
3 In order to test the correlation between paternal lineages and continental
4 linguistic groups, an additional AMOVA was performed classifying the 44 Native
5 American populations into 22 different linguistic families (see Supplementary Table
6 S1). It is noteworthy that significant differences between linguistic groups were found:
7 11.60% ($p < 0.01$), suggesting a genetic differentiation of linguistic groups in the
8 Americas when a broader representation of continental linguistic families are
9 considered, but not within linguistic Mexican groups.
10
11
12
13
14
15
16
17

18 *Analysis of haplogroup Q in Native Mexicans*

19
20 In order to establish the genetic relationships between the Y-STRs haplotypes observed
21 in our sample set, a median joining network was constructed for haplogroup Q, in which
22 only native Mexican haplotypes with 17 STRs were included. A star-like pattern
23 without population clustering was observed, with the exception of one cluster mainly
24 formed by Pima individuals (Figure 4). The estimated age for haplogroup Q in Native
25 Mexicans was 6,436 (± 913) ybp. This cluster of mainly Pima individuals (shaded area
26 in Figure 4) comprised 10 different haplotypes, from which nine were found in Pimas
27 (30 individuals) and one in Tarahumaras (4 individuals). It is noteworthy that these two
28 populations are the northernmost samples in the analysis and that they do not belong to
29 what is strictly defined as Mesoamerica. This cluster was characterized by one
30 haplotype shared between 12 Pimas (13, 13, 17, 12, 16, 23, 10, 14, 13, 14, 11, 12, 19,
31 17, 15, 22, 11, same STR order as the one described in *Y chromosome genotyping*
32 section). The age estimation for this cluster was 2,486 (± 785) ybp.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Discussion

We have explored the Y-chromosome diversity of Native Mexican populations and contrasted to continental patterns of variation and distribution across the Americas. Due to the presence of a substantial proportion of non-Native American (mainly European) male lineages in our samples (see Table 1), the Y-chromosome haplogroup composition of Native Mexican populations shows a notable contrast with previous results from the mtDNA counterpart for the same populations (less than 0.5%, Sandoval et al. 2009). These results suggest a sexual bias in the introgression of non-Native American lineages in native populations, consistent with previous studies performed in different Latin-American populations (Carvajal-Carmona et al. 2003; Dipierri et al. 1998; Gonzalez-Andrade et al. 2007; Mendizabal et al. 2008; Sans et al. 2002). In addition, there is variation in the Y-chromosome admixture proportions among Native Mexican populations, with the highest proportion of non-Native haplotypes (32%) in the urban cosmopolitan sample of Xochimilco Nahuas, whereas Triqui and Pima (with the highest proportion of Native American haplotypes) are two indigenous groups well known to be characterized by their geographic and cultural isolation.

Interestingly, Pima and Tarahumara, the northernmost samples, exhibit the lowest diversity values, based on haplogroup Q composition (average genetic diversity and pairwise differences, see Table 2). This observation is in total agreement with our previous results based on mtDNA diversity of the same set of samples, where we found a clear differentiation of the two northern populations from the rest (Sandoval et al. 2009). In addition, they present significant differences in Y-chromosome variation compared to the Mesoamerican samples as shown in the AMOVAs, and some of their lineages clustered differentially in the haplogroup Q network (see the highlight cluster in Figure 4). Pima and Tarahumara are the only sampled populations outside the

1
2
3 cultural, linguistic and geographic region of Mesoamerica. Different geographic and
4
5 climate conditions between North America and Mesoamerica might have created over
6
7 millennia this genetic discontinuity between both geographical areas.
8

9
10 An additional evidence of differentiation was obtained by directly genotyping
11
12 the M3 mutation that defines the Q1a3a haplogroup within clade Q. We found ample
13
14 segregation of both Q-M242* and Q-M3 haplogroups within Mexico, with Pima and
15
16 Triqui showing the lowest and Maya the highest counts of Q-M3 individuals. This
17
18 scenario differs drastically with other regions at a continental scale. The fact that we
19
20 observed the vast majority of individuals from Central and South American populations
21
22 to be restricted to the Q-M3 branch, is compatible with the notion that most Native
23
24 Americans from Mesoamerica southwards are descendants from a single wave of
25
26 migration. In contrast, paternal lineages in Mexico trace back their origin not only to the
27
28 Q-M3 branch but also to other branches outside Q-M3 within haplogroup Q. This
29
30 observation supports the idea that present day Mexico might have constituted an area of
31
32 transition in the diversification of paternal lineages during the colonization of the
33
34 Americas, with a larger proportion of individuals being more remotely related to the Q-
35
36 M3 primogenitor, in particular those from Northern Mexico such as the Pima or
37
38 extremely isolated groups such as the Triqui.
39
40
41
42

43 Overall, we found extensive heterogeneity among Native Mexican populations
44
45 where the strongest signal of differentiation is driven by the northern populations
46
47 outside Mesoamerica, both within Mexico and at the continental scale. With the modest
48
49 resolution provided by a limited number of markers, we found a lack of genetic
50
51 structure within Mesoamerica and the rest of the continent, which is likely the result of
52
53 a shared recent ancestry from Mesoamerica southwards resulting in lower levels of
54
55 differentiation compared to the more divergent northern populations from the Arctic and
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

North America. Likewise, we did not find a correlation between linguistic affiliation and Y-chromosome diversity within Mexico, which is also consistent with previous observations of mtDNA diversity in Native Mexicans (Sandoval et al., 2009). However, some correlation between linguistic and genetic diversity is found for the paternal lineages at a continental scale, suggesting a genetic link between linguistically close populations regardless their geographical location. Most of the genetic and linguistic correlations found in human populations have been attributed to their correlation with geography; i.e. genes and languages are correlated as a result of geography (Comas et al. 2008). In the analysis of male lineages in Native Americans, this correlation is primarily captured at the very broad continental scale and when North American populations are contrasted with the rest, probably as a result of distinct migration waves leaving a deep genetic signature in these major groups. Most likely, many more markers (even hundreds of thousands and genome-wide distributed) are needed to have enough power to explore fine-scale patterns of genetic structure within sub-continental regions of the Americas.

Finally, we conducted a phylogenetic analysis on the native haplotypes in order to perform some historical inferences in Mexican populations. Haplogroup Q (defined by M242 mutation) is one of the two main branches of the major Y-chromosome lineage P, which originated approximately 26,600 to 41,400 ybp (Karafet et al. 2008). It has been established that the bearers of M242 mutation migrated eastward across Siberia until they reached the north-eastern point of Asia trough the Bering ice corridor about 15,000 to 20,000 years ago (Cavalli-Sforza et al. 1994; Ray et al. 2010; Torroni et al. 1994). Haplogroup Q represents a recent paternal founder for the Native American populations into the peopling of the Americas (Bortolini et al. 2003; Seielstad et al. 2003). The diversity of haplogroup Q in Native Mexicans is around 6,436 (± 913) ybp,

that is, well within the aforementioned ranges, pointing to a recent origin of these haplotypes. It is noteworthy the common and recent origin of a cluster in Pima and Tarahumara, which could be related to a common expansion in the area located northern of the Mesoamerica border.

Overall, our results offer a detailed picture of the male-specific genetic composition of a dense panel of Native Mexican populations and reveal that Mesoamerican populations played a crucial role in patterning the subsequent diversification of paternal lineages during the colonization of the Americas. A north to south gradient of decreasing diversity has been well documented before, yet here we are able to locate a region in which one of the major breakpoints of diversity took place at a continental scale as revealed by Y-chromosome markers. Further studies at larger, genome-wide, scales integrated with further linguistic and archaeological knowledge are necessary in order to extend the characterization of Native American genomes to a much finer scale.

Acknowledgements

We would like to thank Mònica Vallés (Unitat de Biologia Evolutiva, UPF), Stéphanie Plaza and Roger Anglada (Servei Genòmica, UPF) for their technical support; and Kenneth Kidd and Judith Kidd (Yale University) for kindly providing Pima, Maya and Quechua samples. Isabel Mendizabal was supported by a PhD fellowship by the Basque Government (Hezkuntza, Unibertsitate eta Ikerketa Saila, Eusko Jaurlaritza, BFI107.4). This research was supported by the Agencia Española de Cooperación Internacional (AECI, Programa de Cooperación Interuniversitaria e Investigación Científica entre España e Iberoamérica, A/7694/07), Dirección General de

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Investigación, Ministerio de Ciencia e Innovación, Spain (CGL2010-14944/BOS), and
Direcció General de Recerca, Generalitat de Catalunya (2009SGR1101).

Figure Legends

Figure 1. Geographical location of Native Mexican populations sampled and the additional populations included in the continental database. Population codes as shown in Supplementary Table 1.

Figure 2. Genetic relationships among Mexican and continental populations. A) Two-dimensional MDS plot of Native Mexicans based on R_{ST} genetic distances using 17 Y-STR loci (stress value = 0.090). Labels in bold indicate non-Mesoamerican populations. B) Two-dimensional MDS plot of Native American populations from the continental database based on R_{ST} genetic distances using 7 Y-STR loci reported in all available datasets (stress value = 0.142). A total of 44 populations were grouped into 6 major cultural areas.

Figure 3. Two-dimension Correspondence Analysis (CoA) based on absolute frequencies of the three major native haplogroups (Q-M242, Q-M3, and C) reported in all available datasets. 27 Native American populations included. Mexican populations were classified separately to contrast their haplogroup composition against North, Central, and South American populations. Selected populations are labeled within each group to ease interpretation.

Figure 4. Network of haplogroup Q in Native Mexican samples for 17 Y-STRs. Color code: Triqui (red), Tarahumara (yellow), Mixtec (black), Otomi (aqua), Nahua (blue), Purepechas (pink), Yucatec Maya (orange), and Pima (green). Area of the network in pale blue indicates a branch restricted to Pima and Tarahumara individuals and was dated separately (see main text for details).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supplementary Material

Supplementary Table S1. References and general information of the 44 Native American populations included in the continental database used for comparison to the Mexican Y-chromosome haplotypes analyzed.

Supplementary Table S2. Haplogroup Q associated haplotypes based on 17 Y-chromosome STRs of the AmpFISTR® Yfiler™ kit for Mexican populations reported in this study: Triqui, Tarahumara, Purepecha, Otomi, Mixtec, Nahua from Xochimilco, Nahua from Zitlala, Nahua from Ixhuatlancillo, Nahua from Necoxtla, Yucatec Maya, Pima. One Quechua population was also included for comparison.

Supplementary Figure S1. Two-dimension Correspondence Analysis (CoA) plot of Native Mexican samples with more than 10 individuals, based on absolute frequencies of the two native and other non-native haplogroups found. Open circles are populations and solid circles represent haplogroups.

References

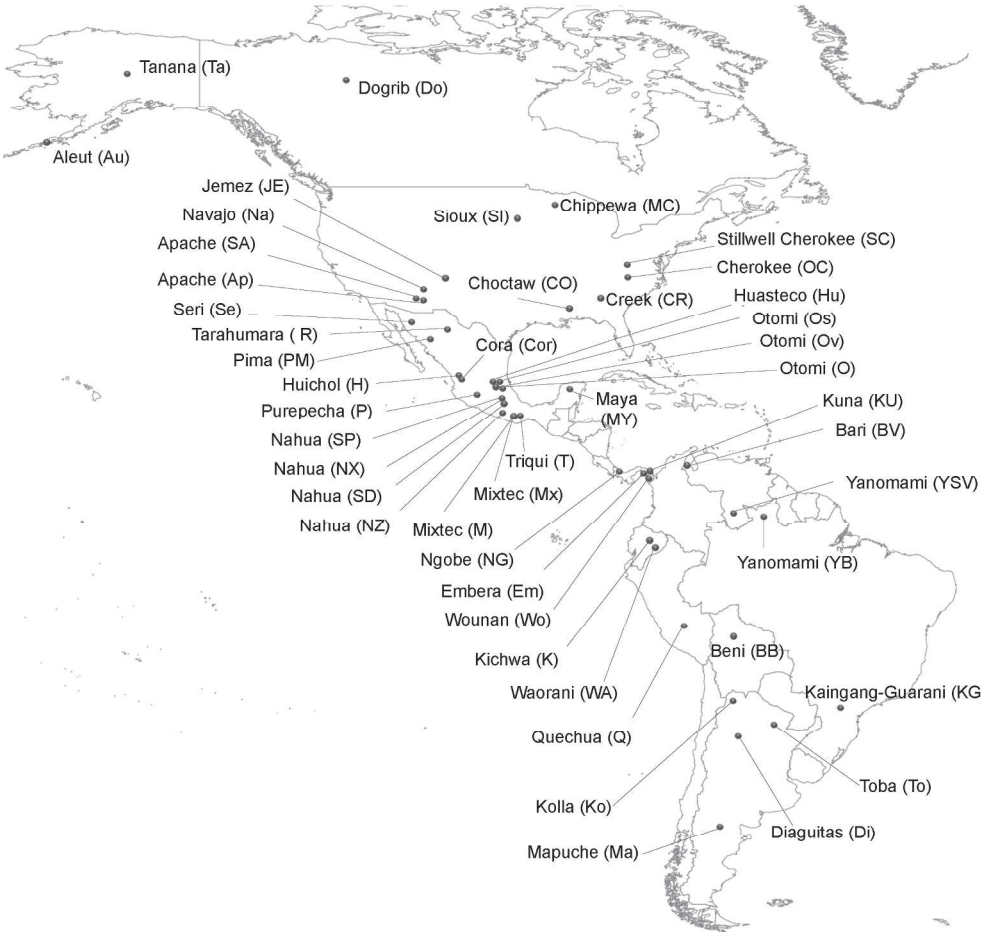
- Bandelt HJ, Forster P, and Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. 16(1):37-48.
- Barrot C, Sánchez C, Ortega M, Gonzalez-Martin A, Brand-Casadevall C, Gorostiza A, Huguet E, Corbella J, and Gene M. 2005. Characterisation of three Amerindian populations from Hidalgo State (Mexico) by 15 STR-PCR polymorphisms. *Int J Legal Med* 119(2):111-115.
- Bergen AW, Wang CY, Tsai J, Jefferson K, Dey C, Smith KD, Park SC, Tsai SJ, and Goldman D. 1999. An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann Hum Genet* 63(Pt 1):63-80.
- Bortolini MC, Salzano FM, Bau CH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-De-Guerra D, Bedoya G and others. 2002. Y-chromosome biallelic polymorphisms and Native American population structure. *Ann Hum Genet* 66(Pt 4):255-259.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT and others. 2003. Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73(3):524-539.
- Bryc K, Velez C, Karafet T, Moreno-Estrada A, Reynolds A, Auton A, Hammer M, Bustamante CD, and Ostrer H. 2010. Colloquium paper: genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci U S A* 107 Suppl 2:8954-8961.
- Campbell L. 1997. American Indian languages: The historical linguistics of Native America. Oxford, editor. New York: Oxford University Press.
- Campbell L, Kaufman T, and Smith-Stark T. 1986. Meso-America as a linguistic area. *Language* 62(3):530-558.
- Carmack RM, Janine L G, and Gary H G. 1996. Legacy of Mesoamerica: The History and Culture of a Native American Civilization. New Jersey: Prentice Hall.
- Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P, Ospina J, Bedoya G, Freimer N, and Ruiz-Linares A. 2003. Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum Genet* 112(5-6):534-541.
- Cavalli-Sforza L, Menozzi P, and Piazza A. 1994. The History and Geography of Human Genes. Princeton: Princeton University Press.
- Cerda-Flores RM, Budowle B, Jin L, Barton SA, Dekka R, and Chakraborty R. 2002a. Maximum likelihood estimates of admixture in Northeastern Mexico using 13 short tandem repeat loci. *Am J Hum Biol* 14(4):429-439.
- Cerda-Flores RM, and Garza-Chapa R. 1989. Variation in the gene frequencies of three generations of humans from Monterrey, Nuevo Leon, Mexico. *Hum Biol* 61(2):249-261.
- Cerda-Flores RM, Ramirez-Fernandez E, and Garza-Chapa R. 1987. Genetic admixture and distances between populations from Monterrey, Nuevo Leon, Mexico, and their putative ancestral populations. *Hum Biol* 59(1):31-49.
- Cerda-Flores RM, Villalobos-Torres MC, Barrera-Saldana HA, Cortes-Prieto LM, Barajas LO, Rivas F, Carracedo A, Zhong Y, Barton SA, and Chakraborty R.

- 2002b. Genetic admixture in three Mexican Mestizo populations based on D1S80 and HLA-DQA1 loci. *Am J Hum Biol* 14(2):257-263.
- Comas D, Bosch E, and Calafell F. 2008. Human Genetics and Languages. *ENCYCLOPEDIA OF LIFE SCIENCES* John Wiley & Sons, Ltd: Chichester <http://www.welsnet/> [DOI: 101002/9780470015902a0020810].
- Dipierri JE, Alfaro E, Martinez-Marignac VL, Bailliet G, Bravi CM, Cejas S, and Bianchi NO. 1998. Paternal directional mating in two Amerindian subpopulations located at different altitudes in northwestern Argentina. *Hum Biol* 70(6):1001-1010.
- Excoffier L, Laval G, and Schneider S. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47-50.
- Excoffier L, Smouse PE, and Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131(2):479-491.
- Felix-Lopez XA, Arguello-Garcia R, Cerda-Flores RM, Penaloza-Espinoza RI, Buentello-Malo L, Estrada-Mena FJ, Ramos-Kuri M, Gomez FS, and Arenas-Aranda DJ. 2006. FMR1 CGG repeat distribution and linked microsatellite-SNP haplotypes in normal Mexican Mestizo and indigenous populations. *Hum Biol* 78(5):579-598.
- Gonzalez-Andrade F, Sanchez D, Gonzalez-Solorzano J, Gascon S, and Martinez-Jarreta B. 2007. Sex-specific genetic admixture of Mestizos, Amerindian Kichwas, and Afro-Ecuadorans from Ecuador. *Hum Biol* 79(1):51-77.
- González-Oliver A, Márquez-Morfin L, Jimenez JC, and Torre-Blanco A. 2001. Founding Amerindian mitochondrial DNA lineages in ancient Maya from Xcaret, Quintana Roo. *Am J Phys Anthropol* 116(3):230-235.
- Gorostiza A, Gonzalez-Martin A, Ramirez CL, Sanchez C, Barrot C, Ortega M, Huguet E, Corbella J, and Gene M. 2007. Allele frequencies of the 15 AmpF/Str Identifier loci in the population of Metztitlan (Estado de Hidalgo), Mexico. *Forensic Sci Int* 166(2-3):230-232.
- Green LD, Derr JN, and Knight A. 2000. mtDNA affinities of the peoples of North-Central Mexico. *Am J Hum Genet* 66(3):989-998.
- Gutierrez-Alarcon AB, Moguel-Torres M, Leon-Jimenez AK, Cuellar-Nevarez GE, and Rangel-Villalobos H. 2007. Allele and haplotype distribution for 16 Y-STRs (AmpFISTR Y-filer kit) in the state of Chihuahua at North Center of Mexico. *Leg Med (Tokyo)* 9(3):154-157.
- Hammer MF, and Zegura SL. 2002. THE HUMAN Y CHROMOSOME HAPLOGROUP TREE: Nomenclature and Phylogeography of Its Major Divisions. *Annual Review of Anthropology* 31:303-321.
- Hernandez-Gutierrez S, Hernandez-Franco P, Martinez-Tripp S, Ramos-Kuri M, and Rangel-Villalobos H. 2005. STR data for 15 loci in a population sample from the central region of Mexico. *Forensic Sci Int* 151(1):97-100.
- Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, and Jobling MA. 2002. Y chromosomal evidence for the origins of oceanic-speaking peoples. *Genetics* 160(1):289-303.
- Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, and Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18(5):830-838.
- Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, and Hammer MF. 2002. High levels of Y-chromosome differentiation among native Siberian populations

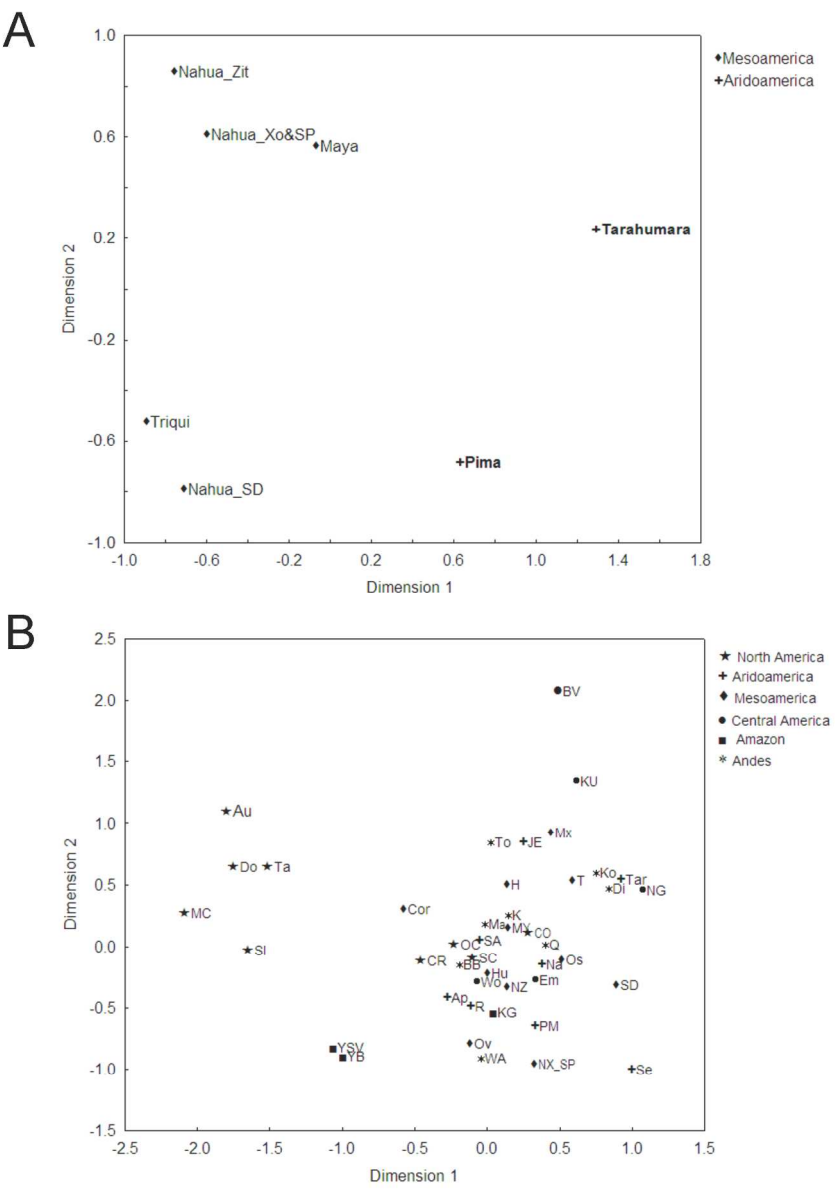
- and the genetic signature of a boreal hunter-gatherer way of life. *Hum Biol* 74(6):761-789.
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P and others. 1999. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64(3):817-831.
- Kemp BM, Gonzalez-Oliver A, Malhi RS, Monroe C, Schroeder KB, McDonough J, Rhett G, Resendez A, Penaloza-Espinosa RI, Buentello-Malo L and others. 2010. Evaluating the Farming/Language Dispersal Hypothesis with genetic variation exhibited by populations in the Southwest and Mesoamerica. *Proc Natl Acad Sci U S A* 107(15):6759-6764.
- Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, and Wallace DC. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 70(1):192-206.
- Licea-Cadena RA, Rizzo-Juarez RA, Muniz-Lozano E, Paez-Riberos LA, and Rangel-Villalobos H. 2006. Population data of nine STRs of Mexican-mestizos from Veracruz (Central South-Eastern, Mexico). *Leg Med (Tokyo)* 8(4):251-252.
- López-Austin A, and López- Luján L. 2001. El pasado indígena. Historia FdHdlAS, editor. México D.F.: Fondo de Cultura Económica y El Colegio de México.
- Luna-Vazquez A, Vilchis-Dorantes G, Aguilar-Ruiz MO, Bautista-Rivas A, Perez-Garcia A, Orea-Ochoa R, Villanueva-Hernandez D, Munoz-Valle JF, and Rangel-Villalobos H. 2008. Haplotype frequencies of the PowerPlex Y system in a Mexican-Mestizo population sample from Mexico City. *Forensic Sci Int Genet* 2(1):e11-13.
- Luna-Vazquez A, Vilchis-Dorantes G, Aguilar-Ruiz MO, Bautista-Rivas A, Rojo-Nava AL, Rios-Barrios E, and Rangel-Villalobos H. 2005. Population data for 15 loci (Identifiler Kit) in a sample from the Valley of Mexico. *Leg Med (Tokyo)* 7(5):331-333.
- Luna-Vazquez A, Vilchis-Dorantes G, Paez-Riberos LA, Munoz-Valle F, Gonzalez-Martin A, and Rangel-Villalobos H. 2003. Population data of nine STRs of Mexican-Mestizos from Mexico City. *Forensic Sci Int* 136(1-3):96-98.
- Mendizabal I, Sandoval K, Berniell-Lee G, Calafell F, Salas A, Martinez-Fuentes A, and Comas D. 2008. Genetic origin, admixture, and asymmetry in maternal and paternal human lineages in Cuba. *BMC Evol Biol* 8:213.
- Mithun M. 1990. Studies of North American Indian Languages. *Annual Review of Anthropology* 19(October):309-330.
- Mithun M. 1999. The languages of Native North America. Cambridge, editor. Cambridge, UK.: Cambridge University Press.
- Mulligan CJ, Hunley K, Cole S, and Long JC. 2004. Population genetics, history, and health patterns in native americans. *Annu Rev Genomics Hum Genet* 5:295-315.
- Mulligan CJ, Kitchen A, and Miyamoto MM. 2008. Updated three-stage model for the peopling of the Americas. *PLoS One* 3(9):e3199.
- Paez-Riberos LA, Munoz-Valle JF, Figuera LE, Nuno-Arana I, Sandoval-Ramirez L, Gonzalez-Martin A, Ibarra B, and Rangel-Villalobos H. 2006. Y-linked haplotypes in Amerindian chromosomes from Mexican populations: genetic evidence to the dual origin of the Huichol tribe. *Leg Med (Tokyo)* 8(4):220-225.
- Peñaloza-Espinosa RI, Arenas-Aranda D, Cerda-Flores RM, Buentello-Malo L, Gonzalez-Valencia G, Torres J, Alvarez B, Mendoza I, Flores M, Sandoval L and others. 2007. Characterization of mtDNA haplogroups in 14 Mexican indigenous populations. *Hum Biol* 79(3):313-320.

- Rangel-Villalobos H, Jaloma-Cruz AR, Cerda-Aguilar L, Rios-Angulo CD, Mendoza-Carrera F, Patino-Garcia B, Sandoval-Ramirez L, and Figuera-Villanueva LE. 2001a. [The genetic DNA trace in men: chromosome Y haplotypes in a Mexican population, analyzing 5 STRs]. *Rev Invest Clin* 53(5):401-406.
- Rangel-Villalobos H, Jaloma-Cruz AR, Sandoval-Ramirez L, Velarde-Felix JS, Gallegos-Arreola MP, and Figuera LE. 2001b. Y-chromosome haplotypes for six short tandem repeats (STRs) in a Mexican population. *Arch Med Res* 32(3):232-237.
- Rangel-Villalobos H, Munoz-Valle JF, Gonzalez-Martin A, Gorostiza A, Magana MT, and Paez-Riberos LA. 2008. Genetic admixture, relatedness, and structure patterns among Mexican populations revealed by the Y-chromosome. *Am J Phys Anthropol* 135(4):448-461.
- Rangel-Villalobos H, Rivas F, Sandoval L, Ibarra B, Garcia-Carvajal ZY, Cantu JM, and Figuera LE. 2000. Genetic variation among four Mexican populations (Huichol, Purepecha, Tarahumara, and Mestizo) revealed by two VNTRs and four STRs. *Hum Biol* 72(6):983-995.
- Rangel-Villalobos H, Rivas F, Torres-Rodriguez M, Jaloma-Cruz AR, Gallegos-Arreola MP, López-Satow J, Cantu JM, and Figuera LE. 1999. Allele frequency distributions of six Amp-FLPS (D1S80, APO-B, VWA, TH01, CSF1PO and HPRTB) in a Mexican population. *Forensic Sci Int* 105(2):125-129.
- Ray N, Wegmann D, Fagundes NJ, Wang S, Ruiz-Linares A, and Excoffier L. 2010. A statistical evaluation of models for the initial settlement of the american continent emphasizes the importance of gene flow with Asia. *Mol Biol Evol* 27(2):337-345.
- Rubi-Castellanos R, Anaya-Palafox M, Mena-Rojas E, Bautista-Espana D, Munoz-Valle JF, and Rangel-Villalobos H. 2009a. Genetic data of 15 autosomal STRs (Identifiler kit) of three Mexican Mestizo population samples from the States of Jalisco (West), Puebla (Center), and Yucatan (Southeast). *Forensic Sci Int Genet* 3(3):e71-76.
- Rubi-Castellanos R, Martinez-Cortes G, Munoz-Valle JF, Gonzalez-Martin A, Cerda-Flores RM, Anaya-Palafox M, and Rangel-Villalobos H. 2009b. Pre-Hispanic Mesoamerican demography approximates the present-day ancestry of Mestizos throughout the territory of Mexico. *Am J Phys Anthropol* 139(3):284-294.
- Ruiz-Linares A, Ortiz-Barrientos D, Figueroa M, Mesa N, Munera JG, Bedoya G, Velez ID, Garcia LF, Perez-Lezaun A, Bertranpetit J and others. 1999. Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. *Proc Natl Acad Sci U S A* 96(11):6312-6317.
- Salazar-Flores J, Dondiego-Aldape R, Rubi-Castellanos R, Anaya-Palafox M, Nuno-Arana I, Canseco-Avila LM, Flores-Flores G, Morales-Vallejo ME, Barojas-Perez N, Munoz-Valle JF and others. 2010. Population structure and paternal admixture landscape on present-day Mexican-Mestizos revealed by Y-STR haplotypes. *Am J Hum Biol* 22(3):401-409.
- Salzano FM. 2002. Molecular variability in Amerindians: widespread but uneven information. *An Acad Bras Cienc* 74(2):223-263.
- Sandoval K, Buentello-Malo L, Penaloza-Espinosa R, Avelino H, Salas A, Calafell F, and Comas D. 2009. Linguistic and maternal genetic diversity are not correlated in Native Mexicans. *Hum Genet* 126(4):521-531.
- Sans M, Weimer TA, Franco MH, Salzano FM, Bentancor N, Alvarez I, Bianchi NO, and Chakraborty R. 2002. Unequal contributions of male and female gene pools

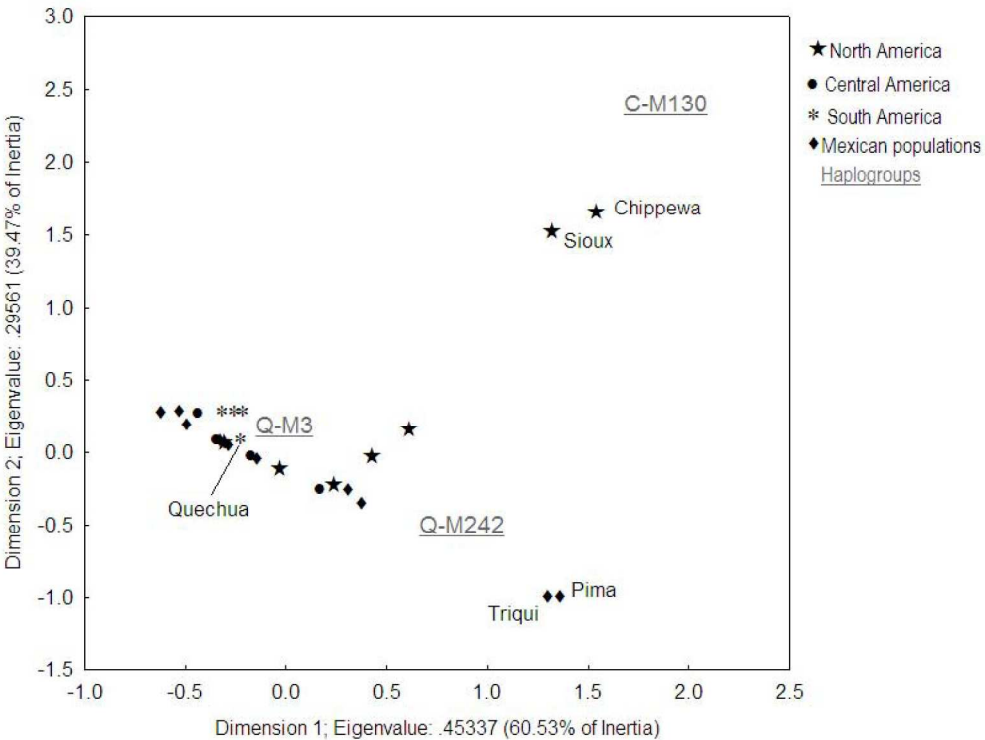
- from parental populations in the African descendants of the city of Melo, Uruguay. *Am J Phys Anthropol* 118(1):33-44.
- Schneider S, Excoffier L, and Laval G. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1:47-50. .
- Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, and Wells RS. 2003. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am J Hum Genet* 73(3):700-705.
- Silva-Zolezzi I, Hidalgo-Miranda A, Estrada-Gil J, Fernandez-Lopez JC, Uribe-Figueroa L, Contreras A, Balam-Ortiz E, del Bosque-Plata L, Velazquez-Fernandez D, Lara C and others. 2009. Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc Natl Acad Sci U S A* 106(21):8611-8616.
- Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, Labarga CM, Rickards O, Tyler-Smith C, Pena SD, and Santos FR. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am J Hum Genet* 68(6):1485-1496.
- Torres-Rodriguez M, Martinez-Cortes G, Paez-Riberos LA, Sandoval L, Munoz-Valle JF, Ceballos-Quintal JM, Pinto-Escalante D, and Rangel-Villalobos H. 2006. Forensic potential of the STR DXYS156 in Mexican populations: inference of X-linked allele null. *Leg Med (Tokyo)* 8(1):52-54.
- Torroni A, Chen YS, Semino O, Santachiara-Beneczeretti AS, Scott CR, Lott MT, Winter M, and Wallace DC. 1994. mtDNA and Y-chromosome polymorphisms in four Native American populations from southern Mexico. *Am J Hum Genet* 54(2):303-318.
- Underhill PA, Jin L, Zemans R, Oefner PJ, and Cavalli-Sforza LL. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci U S A* 93(1):196-200.
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, and Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65(Pt 1):43-62.
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P and others. 2000. Y chromosome sequence variation and the history of human populations. *Nat Genet* 26(3):358-361.
- Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C and others. 2007. Genetic variation and population structure in native Americans. *PLoS Genet* 3(11):e185.
- Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, Poletti G, Mazzotti G, Hill K, Hurtado AM and others. 2008. Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet* 4(3):e1000037.
- Yang NN, Mazieres S, Bravi C, Ray N, Wang S, Burley MW, Bedoya G, Rojas W, Parra MV, Molina JA and others. 2010. Contrasting patterns of nuclear and mtDNA diversity in Native American populations. *Ann Hum Genet* 74(6):525-538.
- Zegura SL, Karafet TM, Zhivotovsky LA, and Hammer MF. 2004. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Mol Biol Evol* 21(1):164-175.



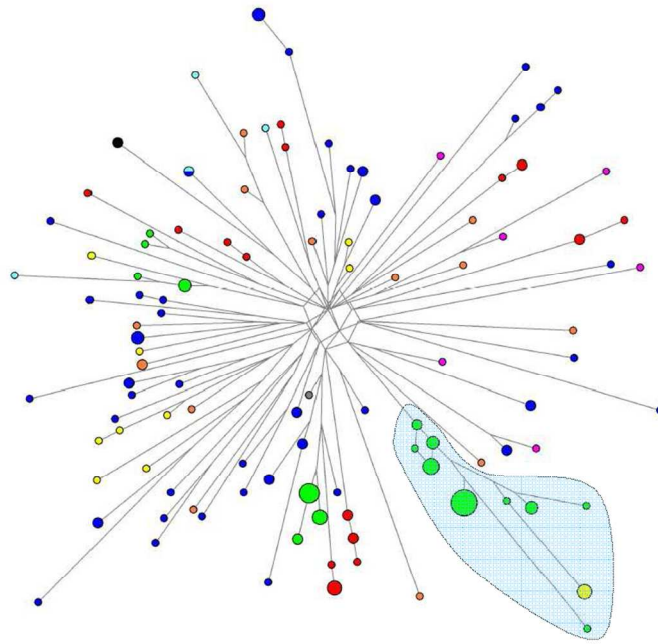
203x190mm (300 x 300 DPI)



153x214mm (300 x 300 DPI)



214x161mm (300 x 300 DPI)



297x210mm (99 x 99 DPI)

Table 1. Y-chromosome haplogroups found in eleven Native Mexican samples and one Quechua population.

SNPs genotyped					Haplogroup frequencies %					M3 genotyped				
Population	n	M242 ^a	M207 ^b	M45 ^c	M130 ^d	Q	R	Non P	C	n	M242* ^e	M3 ^f	M242 %	M3 %
Triqui	22	22	-	-	-	100	-	-	-	9	9	-	100	-
Tarahumara	18	13	1	4	-	72.2	5.6	22.2	-	0	-	-	-	-
Mixtec	2	2	-	-	-	100	-	-	-	0	-	-	-	-
Otomi	7	4	1	2	-	57.1	14.3	28.6	-	0	-	-	-	-
Nahua Xochimilco	22	15	-	7	-	68.1	-	31.8	-	7	-	7	-	100
Nahua-Zitlala	22	19	-	3	-	86.3	-	13.7	-	6	3	3	50	50
Nahua San Pedro	9	9	-	-	-	100	-	-	-	5	-	5	-	100
Nahua Santo Domingo	17	15	-	2	-	88.2	-	11.8	-	15	8	7	53.3	46.7
Purepecha	8	6	-	2	-	75	-	25	-	4	1	3	25	75
Yucatec Maya	19	14	2	3	-	73.7	10.5	15.8	-	13	2	11	15.4	84.7
Pima	51	49	-	2	-	96.0	-	3.9	-	10	9	1	90	10
Total	197	168	4	25	-					69	32	37	46	54

SNPs genotyped					Haplogroup frequencies %					M3 genotyped				
Quechua	8	8	-	-	-	100	-	-	-	7	1	6	14.3	85.8

^aDerived allele: haplogroup Q; ^bDerived allele: haplogroup R; ^cAncestral allele: individuals not belonging to major haplogroup P; ^dDerived allele: haplogroup C; ^eAncestral allele for M3; ^fDerived allele: haplogroup Q1a3a.

Table 2. Diversity parameters for 11 Native Mexican groups and one Quechua population based on 17 Y-STRs from haplogroup Q individuals.

Population	N	K	S	Haplotype diversity	Average GD	Mean PW	Theta*
Triqui	22	15	16	0.95+/-0.02	0.47+/-0.25	8.12+/-3.92	20.43 (14.54)
Tarahumara	13	10	14	0.92+/-0.06	0.43+/-0.24	7.38+/-3.69	10.56 (11.28)
Mixtec	2	1	0	0.00+/-0.00	0.00+/-0.00	0.00+/-0.00	UC
Otomí	4	4	12	1.00+/-0.17	0.50+/-0.35	8.50+/-4.98	UC
Xochimilco Nahua	15	10	14	0.95+/-0.03	0.48+/-0.26	8.16+/-4.03	20.08 (19.26)
Zitlala Nahua	19	17	17	0.98+/-0.02	0.55+/-0.29	9.40+/-4.52	82.59 (153)
San Pedro Nahua	9	7	15	1.00+/-0.07	0.50+/-0.30	8.57+/-4.51	UC
Santo Domingo Nahua	15	8	12	0.91+/-0.04	0.45+/-0.24	7.65+/-3.78	9.28 (5.52)
Purépecha	6	6	16	1.00+/-0.09	0.55+/-0.34	9.46+/-5.07	UC
Yucatec Maya	14	13	17	0.98+/-0.03	0.55+/-0.30	9.38+/-4.58	88.08 (260)
Pima	49	16	13	0.89+/-0.02	0.32+/-0.17	5.60+/-2.73	7.08 (2.25)
Total	168						
Quechua	8	7	15	0.96+/-0.07	0.44+/-0.26	7.50+/-3.92	25.27 (59.97)

N: number of individuals; K: number of different haplotypes; S: number of polymorphic loci; Average GD: average gene diversity; Mean PW: mean pairwise differences. *Theta, in parenthesis standard deviation

UC: Unable to compute theta (Hom) when all gene copies are different

Table 3. Analysis of Molecular Variance (AMOVA) using different grouping structures of Native American populations from Mexico and the Americas.

	Groups	Among groups	Among populations within groups	Within populations
Native Mexicans ^a				
	11 populations		16.30*	83.70*
	4 linguistic families	-2.17ns	17.64*	84.53*
	North vs Meso America	8.43*	11.03*	80.55*
Native Mexicans ^b				
	7 populations		16.80*	83.20*
	3 linguistic families	-3.09ns	18.53*	84.56*
	North vs Meso America	8.22*	11.38*	80.40*
America database ^c				
	Single group		20.47*	79.53*
	Geographic regions ^d (North, Central, South)	1.22ns	19.73*	79.05*
	North America ^e vs all other	18.51*	12.95*	68.54*
	6 Cultural regions	6.63*	14.76*	78.61*
	22 Linguistic families	11.60*	9.42*	78.98*

* P < 0.05; ns: non-significant
All AMOVAS were based on R_{ST} distances

^a Eleven populations were considered with none sample size restriction

^b Seven populations were considered with more than 10 individuals

^c Data from 44 Native American populations based on haplotype information from 7 Y-STRs (see Supplementary Table S1)

^d Geographic grouping structure as in Supplementary Table S1

^e In this case North America includes Aleut, Choctaw, Creek, Chippewa, Cherokee, Sioux, Dogrib, and Tanana (based on cultural groups as in Figure 2B).