9-1-2016

# A Mainly Circum-Mediterranean Origin for West Eurasian and North African mtDNAs in Puerto Rico with Strong Contributions from the Canary Islands and West Africa

Héctor J. Díaz-Zabala
*Ponce Health Sciences University*, diazhectorjoel@yahoo.com

María A. Nieves-Colón
*Arizona State University, Tempe*, mnievesc@asu.edu

Juan C. Martínez-Cruzado
*University of Puerto Rico, Mayagüez*, juancarlos.martinez@upr.edu

# A Mainly Circum-Mediterranean Origin for West Eurasian and North African mtDNAs in Puerto Rico with Strong Contributions from the Canary Islands and West Africa

Héctor J. Díaz-Zabala (*Ponce Health Sciences University*), María A. Nieves-Colón (*Arizona State University, Tempe*), Juan C. Martínez-Cruzado (*University of Puerto Rico, Mayagüez*)

Corresponding Author: Juan C. Martinez-Cruzado, Ph.D. Department of Biology University of Puerto Rico at Mayaguez Call Box 9000 Mayaguez, PR 00681-9000

Abstract

Maternal lineages of West Eurasian and North African origin account for 11.5% of total mitochondrial ancestry in Puerto Rico. Historical sources suggest that this ancestry arrived mostly from European migrations that took place during the four centuries of the Spanish colonization of Puerto Rico. This study analyzed 101 mitochondrial control region sequences and diagnostic coding region variants from a sample set randomly and systematically selected using a census-based sampling frame to be representative of the Puerto Rican population, with the goal of defining West Eurasian-North African maternal clades and estimating their possible geographical origin. Median-joining haplotype networks were constructed using HVR-I and –II sequences from various reference populations in search of shared haplotypes. A posterior probability analysis was performed to estimate the percentage of possible origins across wide geographic regions for the entire sample set and for the most common haplogroups on the island. Principal component analyses were conducted to place the Puerto Rican mtDNA set within the variation present amongst all reference populations. Our study shows that up to 38% of West Eurasian and North African mitochondrial ancestry in Puerto Rico most likely migrated from the Canary Islands. However, most of those haplotypes had previously migrated to the Canary Islands from elsewhere, and there are substantial contributions from various populations across the circum-Mediterranean region and from West African populations related to the modern Wolof and Serer peoples from Senegal and the nomad Fulani who extend up to Cameroon. In conclusion, the West Eurasian mitochondrial ancestry in Puerto Ricans is geographically diverse. However, haplotype diversity seems to be low and frequencies have been shaped by population bottlenecks, migration waves, and random genetic drift. Consequently, approximately 47% of mtDNAs of West Eurasian and North African ancestry in Puerto Rico probably arrived early in its colonial history.

Keywords: West Eurasian-North African-West African ancestry, mitochondrial DNA, haplotype networks, principal component analysis, posterior probability analysis, Tajima's D test, Puerto Rican population, Canary Islands

Short Title: Origins of West Eurasian mtDNAs in Puerto Rico

Submission Date: January 23 2016

Acceptance Date: March 16 2017

Puerto Ricans are an admixed population composed of three main ancestral subcontinental groups: Sub-Saharan African, European and Native American (Bryc et al. 2010; Moreno-Estrada et al. 2013). Documentary sources supported by the archaeological record indicate that the Spanish occupation of Puerto Rico began in 1506, and that by 1513 each of these ancestral populations coexisted on the island. Subsequent migration waves of mostly European and Sub-Saharan African individuals occurred for at least four centuries after the beginning of colonization (Emmer 1999; Picó 2004). A previous study based on a sample set representative of the population of Puerto Rico showed that 61.3% of Puerto Rican mitochondrial DNAs (mtDNAs) are of Native American origin, whereas 27.2% are Sub-Saharan African and 11.5% West Eurasian (Martínez-Cruzado et al. 2005). Subsequent studies have produced consistent results (Vilar et al. 2014). However, a study on the autosomal ancestry of the same sample set identified average genome ancestry proportions of around 63.7% European, 21.2% Sub-Saharan African, and 15.2% Native American (Via et al. 2011). Other studies conducted with Puerto Ricans living in the continental USA have produced similar results (Bryc et al. 2010; Moreno-Estrada et al. 2013; Gravel et al. 2013).

In Puerto Rico, European migrations began with the Spanish colonization of the early 16$^{th}$ century. Historical sources indicate that the first migrations were fostered by the Spanish Crown's selective migration policy, which excluded "Jew", "Moor" and "Gypsy" men from settling on the island (Cifre De Loubriel 1964). Later during the 16$^{th}$ century, the Spanish Crown abandoned this policy and agreed to welcome all migrants (Cifre De Loubriel 1964). Documentary sources indicate that the majority of emigrants from Spain originated in Aragon, Basque Country, Valencia, Galicia, Extremadura, Catalonia, Andalusia, Canary Islands, and Castile (Fernández Méndez 1970).

The vast majority of Spaniards arriving to Puerto Rico were single men who, by 1506 had begun to intermarry with Native American women. The Governor reported to the Crown in 1530 that, of a total of 369 "white" men in Puerto Rico, only 57 were married to "white" women (Brau 1904). The Crown took measures to increase the number of "white" people on the island. This included ordering "white",

enslaved Christian women to be sent to Puerto Rico in 1512, and instituting a policy of subsidizing the migration of Spanish families by offering free tools and seeds (Thomas 1997; Díaz-Soler 2000). From this initiative, 50 families consisting of 207 people arrived to Puerto Rico in a single ship in 1520. However, constant conflict with local Native Americans, frequent hurricanes, the little gold found in Puerto Rico, and the gold rush that developed in Perú at the time stymied the development of a stable settler population (Fernández Méndez 1970).

In 1532, the Spanish crown enacted an order that limited the number of enslaved Sub-Saharan Africans on island farms to five per each "white" peasant. Hence, peasants were sent to Puerto Rico mostly from the poorest parts of Spain, especially the southern provinces of Andalusia and Extremadura, but also from Castile and the Canary Islands (Fernández Méndez 1970). In 1695, 20 families numbering 100 members were sent to Puerto Rico from the Canary Islands by request of the Governor (Brau 1904). This event marked the beginning of subsequent migration waves from the Canary Islands that lasted for the next two centuries (Cifre De Loubriel 1964).

Between 1720 and 1730, Puerto Rico received a group of 176 families from the Canary Islands, each formed by a minimum of five individuals (Borges Jacinto del Castillo 1969). During this time, most migrants from the Canary Islands settled in Venezuela, Mexico, Cuba, Puerto Rico, Uruguay, Dominican Republic, and colonial French Louisiana (Rodríguez Mendoza 2004; Santana Pérez 2008). The 18$^{th}$ century ended with a total population of 153,234 individuals settled in Puerto Rico (Cifre De Loubriel 1964).

Migratory activity to Puerto Rico peaked in the 19$^{th}$ century due to the approval of the Royal Decree of Grace of 1815 and the crumbling of the Spanish empire in the Americas. The Decree incentivized Spanish and European migration to Puerto Rico by permitting the settlement of foreign Catholics with their wealth and slaves (Fernández Méndez 1970). According to Cifre De Loubriel (1964), soon after the implementation of the Royal Decree, French and French Haitians were the most common non-Hispanic migrants arriving to Puerto Rico, followed by people from Italy, Portugal and Corsica. Other migrants arrived in minor proportions from Germany, Ireland, North America, England, Switzerland, Austria, Denmark, the Netherlands, Scotland, Sweden, and Belgium. Furthermore, due to local political struggles in Venezuela, Cuba and Puerto Rico became the most viable destinations for Canary Islanders (Rodríguez Mendoza 2004; Santana Pérez 2008). Additional influxes of Canary Islanders occurred between 1855 and1860, when the Spanish government promoted their migration to Puerto Rico after the island suffered a cholera epidemic which resulted in the death of 26,820 people (Fernández Méndez 1970).

Whereas a Spanish origin for a large number of the Puerto Rican mtDNAs may be expected, the relative contribution of Spain *versus* other sources, and among

the other sources themselves is a matter of debate. In this study, we use the most extensive Puerto Rican mtDNA data to date, obtained through a census-based sampling frame designed to produce a sample set representative of the Puerto Rican population. We aim to determine the geographic origins of those mtDNAs in Puerto Rico that belong to haplogroups typically associated with West Eurasian and North African (WE-NA) populations (superhaplogroup N(xABFOPSY)). After subhaplogroup classification through control region sequencing and RFLP analysis, we applied clustering, probability and phylogenetic analytical methods to identify those populations most likely to have contributed these maternal lineages to Puerto Rico. We find strong genetic drift manifested by the presence of certain defined subhaplogroups at excess frequency in the Puerto Rico mtDNA pool.

## Materials and Methods

**DNA samples.** We used 101 West Eurasian mtDNA samples that were previously collected, extracted and tested with RFLP by Martínez-Cruzado et al. (2005). Through a procedure approved by the Institutional Review Committee of the University of Puerto Rico at Mayagüez (UPRM), voluntary participants completed an informed consent document and answered a questionnaire providing information about their place of birth and their oldest known maternal female ancestor.

**PCR.** MtDNA HVR-I and HVR-II fragments were amplified from DNA samples using flanking PCR primers (Table 1). The amplification reaction mix had the following chemical components: 5µL PCR 10x Buffer, 3µL MgCl2 (25mM), 8µL dNTP (2.5mM), 1.2µL of each primer (20µM), 1.5µL Bovine Serum Albumin (100µg/µL), 19.1µL dd$H_2O$, 10µL of purified DNA and 1µL of *Taq* Polymerase (1U/µL). The total reaction volume was 50µL. The PCR mix was heated at 94ºC for 2.5 minutes and then subjected to 35 cycles of 30 seconds at 94ºC, 1 minute at 56ºC and 70 seconds at 72ºC. To complete amplification, an extension cycle of 10 minutes at 72ºC was added. PCR products were cleaned using the Roche® DNA Purification Kit.

**Table 1.** HVR-I and HVR-II Amplification and Sequencing Primers

| mtDNA fragment | Amplification primer | Sequencing primer |
|---|---|---|
| HVR-I | L15829<br>5'catccgtactatacttcacaac3' | L15854<br>5'cctaatcctaataccaactat3' |
| | H34<br>5'accaaatgcatggagagctcc3' | L16219<br>5'tgcttacaagcaagtacagca3' |
| | | L191<br>5'cgttcaatattacaggagaac3' |
| | | H394<br>5'ccgccaaaagataaaatttg3' |

| | | |
|---|---|---|
| HVR-II | L16491<br>5'ggggtagctaaagtgaactg3' | H16526<br>5'gggaacgtgtgggctatttagg3' |
| | H501<br>5'gtgtgtgctgggtaggatg3' | L16504<br>5'gtgaactgtatccgacatctgg3' |

**DNA sequencing.** HVR-I and HVR-II fragments were sequenced using the primers shown in Table 1. DNA sequencing was performed using the Applied Biosystems ABI-Prism Big Dye Terminator v3.1. The sequencing reaction mix contained the following components: 3μL purified PCR product (5ng/μL), 0.5μL 5X Buffer, 0.8μL primer (1μM), and 0.7μL Big Dye v3.1. Total volume of reaction was 5μL. The Big Dye mix was heated at 96°C for 1 minute and then subjected to 35 cycles of 15 seconds at 96°C, 15 seconds at 50°C and 4 minutes at 60°C. After the completions of 35 cycles, samples were maintained at 4°C. The reaction product was cleaned with ethanol 90% and sodium acetate 3M to avoid sequence read errors. DNA sequences were read using the ABI 3130 Genetic Analyzer of the UPRM Department of Biology. Coding region fragments were amplified and sequenced to identify diagnostic SNPs for clades belonging to haplogroups H, K, and U. These coding region primers are listed in Table 2.

**Table 2.** Primers Used to Identify Defining SNPs in Coding Region Fragments for Subhaplogroup Characterization

| Primer | Defining SNP | Subhaplogroup | Additional SNPs |
|---|---|---|---|
| H3346<br>5'attaggaatgccattgcgatta3' | G3010A<br>A2851G | H1<br>H1h2 | |
| L4485<br>5'gtactaattaatcccctggcc3' | G4769A | H2a | A4985G |
| H7104<br>5'tggtctagggtgtagcctg3' | T6776C | H3 | T3027C |
| L3517<br>5'cacatctaccatcaccctc3' | G3915A | H6a | |
| L4485<br>5'gtactaattaatcccctggcc3' | A4727G | H6a1 | T4639C<br>A4769G<br>A4985G |
| L4754 | A4793G | H7 | |

| | | | |
|---|---|---|---|
| 5’atactaccaatcaatactcatc3’ | C5348T | H7b | A4985G |
| H9304<br>5’tggagtggaagtgaaatcac3’ | G9055A<br>A9093G | U8b and K<br>K1c1 | |

MtDNA sequences were aligned to the revised Cambridge Reference Sequence (rCRS) using the ClustalW tool in MEGA5 (Anderson et al. 1981; Andrews et al. 1999; Tamura et al. 2011). Nucleotide positions considered for analysis of HVR-I were 15854 – 16503 and for HVR-II were 16504 – 501. Sequences from 15969 to 16503 and from 33 to 501 were reliably obtained except when noted (Table S1). Manually-implemented imputation was performed on eleven occasions for population parameter, posterior probability and principal component (PC) analyses (Table S2).

**Population genetic parameters.** Population genetic parameters were estimated for diversity and population expansion measurements using the haplotype counts shown in Table S2. Haplotype sequences in HVR-I (positions 15969-16503), HVR-II (positions 33-501) and a concatenation of both HVR-I + II were used to estimate the number of polymorphic segregating sites (*S*), the number of haplotypes in sample (H), Nei’s haplotype diversity parameter (Hd) (Nei 1987), Nei’s nucleotide diversity parameter ($\pi$) (Nei 1987), and Tajima’s *D* (Tajima 1989a,b). Analyses were performed using DNAsp v5.10 (Librado and Rozas 2009).

**Subhaplogroup characterization.** WE-NA mtDNA clades were characterized based on population haplogroup definitions from the literature using both control and coding region sequences. Haplogroup H clades were classified according to Roostalu et al. (2007), Loogvali et al. (2004) and Pereira et al. (2006); haplogroup J clades were classified according to Richards et al. (2000); haplogroup T clades were classified according to Abu-Amero et al. (2008), González et al. (2003), Maca-Meyer et al. (2004) and Behar et al. (2008); haplogroup K clades were classified according to Behar et al. (2006) and Behar et al. (2008); haplogroup U clades were classified according to Achilli et al. (2005), Álvarez et al. (2007), Turchi et al. (2008), Behar et al. (2008), Rando et al. (1999) and Maca-Meyer et al. (2004); haplogroup V clades were classified according to Álvarez et al. (2007) and Richards et al. (2000); haplogroups HV, R0a and R clades were classified according to Palanichamy et al. (2004), Kivisild et al. (2004) and Richards et al. (2000). An update to every haplogroup classification was performed according to the February 18, 2016 version of the mtDNA tree build 17 in PhyloTree database (van Oven and Kayser 2009). The motifs used to classify clades are shown in Table S3.

**Fisher’s Exact tests.** Fisher’s Exact tests were conducted comparing subhaplogroup frequencies between Puerto Rico and those populations gathered from the literature and listed in Table S4, except for those in which most samples

lacked all or a critical part of the HVR-II (specifically Mauritanians, Malians, and Tuscans). Subhaplogroup characterization was based on control region sequences and limited RFLPs in the coding region, as specified in Table S5. Comparisons were made as in Sarno et al. (2014). Exact tests were performed under the null hypothesis that subhaplogroup frequencies were not significantly different between populations. A significant test ($p < 0.05$) suggests that the null can be rejected and statistically significant differences exist in subhaplogroup frequencies between Puerto Rico and tested populations. The tests were conducted in R version 3.2.4 using the fisher.multicomp( ) function in the RVAideMemoirepackage (http://cran.rproject.org/web/packages/RVAideMemoire/). This function performs a pairwise comparison of multi-column contingency tables and corrects for multiple comparisons using the Bonferroni method. We also calculated statistical power for each comparison using the power.fisher.test() function in the R statmod package (https://cran.r-project.org/web/packages/statmod/index.html) using 10,000 simulations.

**Phylogenetic networks.** Median-joining networks based on HVR-I and HVR-II sequences were constructed for the most frequent subhaplogroups in Puerto Rico (frequency $\geq 4$) and the populations listed in Table S4 using Network 5.0 (www.fluxus-engineering.com) (Bandelt et al. 2000). For simplification of some networks, a transition relative to the CRS at position 263 was assumed for sequences not covering that site (Dubut et al. 2004; Achilli et al. 2007). Transitions at hypervariable site 16519, A-to-C transversions at positions from 16182 to 16184 associated to the 16189 mutation, and indels occurring in the interrupted C-stretch between positions 303 and 315 were not considered. Pertinent nucleotide positions were weighted according to the relative number of occurrences of each mutation in the worldwide human mitochondrial phylogenetic tree (Soares et al. 2009).

**Posterior probability analyses.** To widen the geographic span of populations compared to the Puerto Rican cohort, we used samples in the literature with HVR-I sequences only, or both HVR-I and –II, for a total of 61 populations (Table S6). For consistency with the comparative data, sample categorization was based only on HVR-I sequences for this analysis (Table S7). The 61 populations were pooled into 30 metapopulations or subcontinental regions as described below, and the probability of origin of each Puerto Rican HVR-I-based clade was calculated using a Bayesian approach to estimate the posterior probability of finding it in any one of the 30 metapopulations or subcontinental regions. The calculation was made using the following statistical equation from Mendizabal et al. (2008): $P_{0s} = \frac{1}{n}\sum_{i=1}^{n} k_i \frac{P_{is}}{P_{ic}}$, where $n$ is the total number of Puerto Rican samples; $k_i$ is the number of times each clade ($i$) is found in the Puerto Rican sample set; $P_{is}$ is the frequency of clade $i$ in a specific population or subcontinental region; and $P_{ic}$ is the frequency of clade $i$ in the metapopulation. The standard deviation for each calculation was also obtained

from Mendizabal et al. (2008): $SD(P_{os}) = \sqrt{\frac{P_{os}(1-P_{os})}{n}}$. To define subcontinental regions, Europe was first divided as in Richards et al. (2000), and data was added to the regions as follows (Table S6): to the Mediterranean West (*n* = 209 in Richards et al. (2000)), Spain (*n* = 301) and Portugal (*n* = 273); to the Mediterranean Central (*n* = 296), North Italy (*n* = 388) and Sardinia (*n* = 42); to the Mediterranean East (*n* = 165), Crete (*n* = 283) and Bosnia (*n* = 142); to the North West (*n* = 453), France (*n* = 210); to the Alps (*n* = 215), Austria (*n* = 272); to the South East (*n* = 229), Slovenia (*n* = 104) and Hungary (*n* = 73); to the North Central (*n* = 328), Northeast Germany (*n* = 212), Poland (*n* = 430) and Czech Republic (*n* = 177); to the North East (*n* = 398), Finland (*n* = 200) and Russia (*n* = 200); to the North Caucasus (*n* = 191), Georgia (*n* = 45), Armenia (*n* = 190) and Azerbaijan (*n* = 46) and renamed the region as Caucasus. Northwest Africa was defined including Mauritania (*n* = 47), Morocco (*n* = 61), West Sahara (*n* = 14), Tunisia (*n* = 46), Moroccan Berbers (*n* = 53) and Tunisian Berbers (*n* = 26). Because of their high representation of haplogroup U, a group of West African populations was made from Fulani, Wolofs and Serers (*n* = 25). Middle East is composed of Dubaians (*n* = 193), Palestinians (*n* = 198), Bedouin (*n* = 73), Druze (*n* = 120), Nubians (*n* = 34), Egyptians (*n* = 54), Iraqis (*n* = 105), Iranians (*n* = 12) and Syrians (*n* = 62). Other populations were maintained individually, including the Basques (*n* = 155), Canarians (*n* = 278), Kurds (*n* = 78), Scandinavians (*n* = 312), Turks (*n* = 202) and 13 non-Ashkenazi Jewish populations (*n* = 1047) defined in Behar et al. (2008). Posterior probability calculations were performed for the global sample set, and then individually for each of the most common WE-NA haplogroups in Puerto Rico, H, J, and U, which together make up 84% of the WE-NA mtDNAs in Puerto Rico (see RESULTS). Posterior probability geographic distribution figures were produced in R version 3.2.4 using the rworldmap package (South 2011).

**Principal component analyses.** Principal component analyses (PCA) can highlight similarities in subhaplogroup frequency distributions among populations, providing another means to assess relationships between populations. PCAs were performed comparing Puerto Rican WE-NA mtDNA subhaplogroup frequencies to the same set of populations and subcontinental regions used in the posterior probability analysis. The Belmonte and Georgian Jew populations (Behar et al. 2008) were not included because these populations were outliers, providing most of the variation within PC1 and reducing resolution. PCAs were also performed individually for the three most common WE-NA haplogroups in Puerto Rico (H, J and U). All PCAs were conducted using the prcomp() function in R version 3.2.4 (R Development Core Team 2013).

## Results

**WE-NA mtDNA variability in Puerto Rico.** We found a total of 34 WE-NA

mtDNA clades in our population sample of 101 unrelated individuals. Samples belonging to haplogroup U, initially identified by the *Hin*fI site at nucleotide position (np) 12308 when using a mismatched primer (Martínez-Cruzado et al. 2005), were the most frequent, accounting for 33% of our total sample. Samples belonging to haplogroup H were the second most frequent contributing around 28%of the total sample, and samples belonging to haplogroup J accounted for 23%. Thus, whereas haplogroups J and U combined account only for approximately 26.5% of the mtDNAs in the region encompassing Europe and the Caucasus, in Puerto Rico they account for 56% of all mtDNAs of WE-NA ancestry. More specifically, the frequency of haplogroup U averages 18.9%, and varies from 11.9% in Tuscans (Achilli et al. 2007) to 24.4% in Georgians (Comas et al. 2000), whereas that of haplogroup J averages 7.6% and varies from 4.4% among Georgians to 10.9% in Poles (Malyarchuk et al. 2002). Hence, both haplogroups are overly represented in the Puerto Rican sample set. The very high proportions of these haplogroups suggest a major role of genetic drift in determining maternal lineage frequencies of WE-NA ancestry in Puerto Rico.

The obtained RFLP data together with control regions sequences (Table S1) were used to further classify the mtDNAs into clades according to the worldwide mitochondrial phylogenetic tree (van Oven and Kayser 2009). The obtained control region sequence haplotypes, their frequency, and the resulting subhaplogroup distribution are shown in Table S2 and Figure 1. Clades J1b1a1 (16%), U5b1b1b (10%), U5b2b3a (9%) and H1b (9%) are the most frequent subhaplogroups and together account for 44% of the WE-NA sample set. Other frequent clades were H1(xH1b), H3 and T2b with 5% each. Additional mitochondrial lineages were found at lower frequencies varying from 1 to 4% (Figure 1).

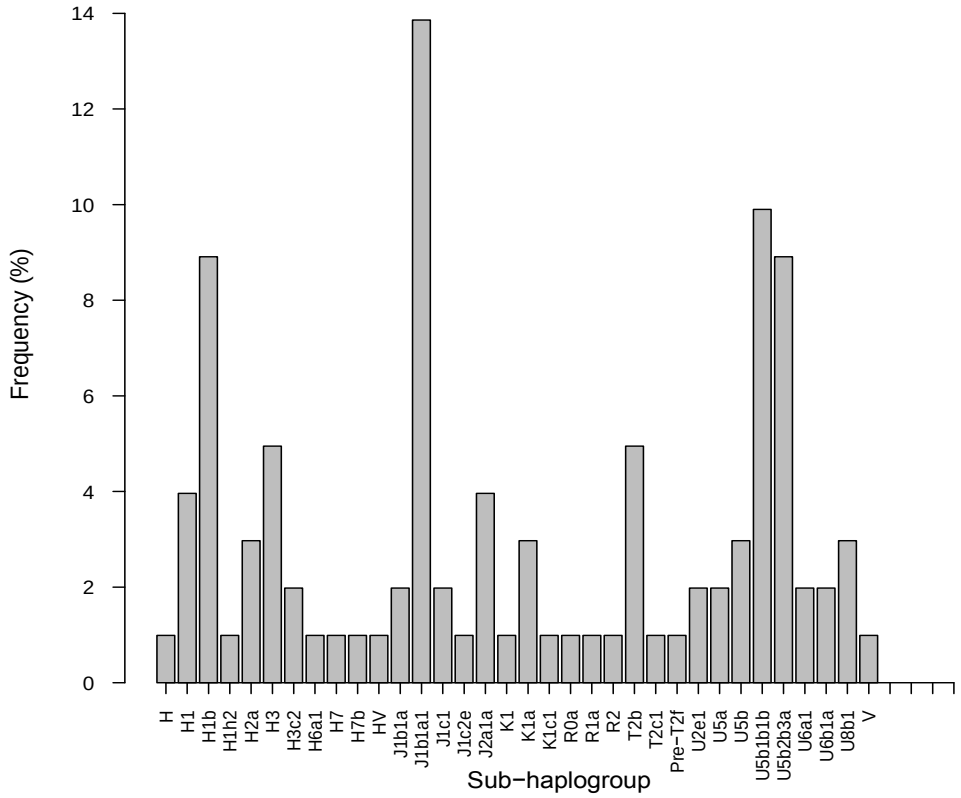


**Figure 1.** WE-NA mtDNA subhaplogroup frequency in Puerto Rico.

**Population genetic parameters.** We found equivalent haplotype diversity between HVR-I and HVR-II, and a substantial increase when both regions are concatenated (Table 3). This suggests that HVR-II haplotypes may be highly informative of mtDNA identity and probable origin. However, HVR-II has a higher concentration of hypermutable sites (Soares et al. 2009), and probably for this reason nucleotide diversity was lower (Table 3). The combination of these results suggests that HVR-II sequencing is highly useful but only when in addition to HVR-I information. We also find negative Tajima's *D* values of -0.8400, -0.9293 and -0.9256 for HVR-I, HVR-II and HVR-I + II fragments, respectively (Table 3). However, p-values for the Tajima's *D* test were not significant ($p > 0.10$).

**Table 3.** Summary Statistics for 101 Puerto Rican MtDNAs of WE-NA Ancestry

| Region | Positions | No. Sites | S[1] | H[2] | Hd[3] (SD)[4] | π[5] (SD)[4] | Tajima's D[6] | *P*-value |
|---|---|---|---|---|---|---|---|---|
| HVR-I | 15969-16503 | 535 | 37 | 37 | 0.9420 (0.0100) | 0.00972 (0.0004) | -0.8400 | P>0.10 |
| HVR-II | 33-501 | 469 | 28 | 32 | 0.9420 (0.0090) | 0.00826 (0.0005) | -0.9293 | P>0.10 |
| HVR-I + -II | 33-501,15969-16503 | 1004 | 65 | 46 | 0.9610 (0.0080) | 0.00904 (0.0004) | -0.9256 | P>0.10 |

[1]S = Number of polymorphic segregating sites. [2]H = Number of haplotypes in sample.
[3]Hd = Haplotype diversity (Nei, 1987). [4]SD = Standard deviation. [5]π = Nucleotide diversity (Nei, 1987). [6]Tajima, 1989a,b.

**Fisher's Exact tests.** We conducted Fisher's Exact tests comparing Puerto Rico with each one of 29 populations in Europe, North Africa and the Middle East, testing the null hypothesis that the subhaplogroup frequencies in both populations were not significantly different. The frequency of subhaplogroup J1b1a1 in Puerto Rico was found to be significantly different to 20 of the 29 populations compared, whereas the frequency of "Others" (E-G,I,M-Q,S,W-Z), U5b1b1b and U5b2b3a subhaplogroups were significantly different in 14, 13, and 12 occasions, respectively (Table S8). This observation suggests that the high frequencies of J1b1a1, U5b1b1b and U5b2b3a, as well as the absence of some haplogroups common in the Middle East, such as haplogroup M, are distinctive traits of the Puerto Rican WE-NA mitochondrial pool. We also calculated the power of each test via simulation (Table S9), and observed that all comparisons in which significant differences were found had $\geq$ 0.985 statistical power. In total, there were 100 instances in which the test had a power $\geq$ 0.985, and significant differences were found in 85 of them. Only for two populations was the null hypothesis not rejected in half of the times or more in which test power was $\geq$ 0.985 (Iranian Jews, 2 non-rejections in 4 tests, and Azeri Jews, 1 non-rejection in 2 tests). Hence, none of the populations used in this study produced a pattern fitting the expectations for a population contributing strongly to the Puerto Rico WE-NA maternal pool.

**Median-joining network analysis.** We constructed a haplotype frequency spectrum (Figure 2) from Table S2. A group of haplotypes of very high frequency (9-to-11) was easily identified. Another group of medium frequency (4-to-5) can also be separated from the vast majority of haplotypes which are of low-frequency (1-to-3). In an effort to identify the origin of the most frequent haplotypes in Puerto Rico, we constructed median-joining haplotype networks for the clades they belong to (Figures S1 – S5).

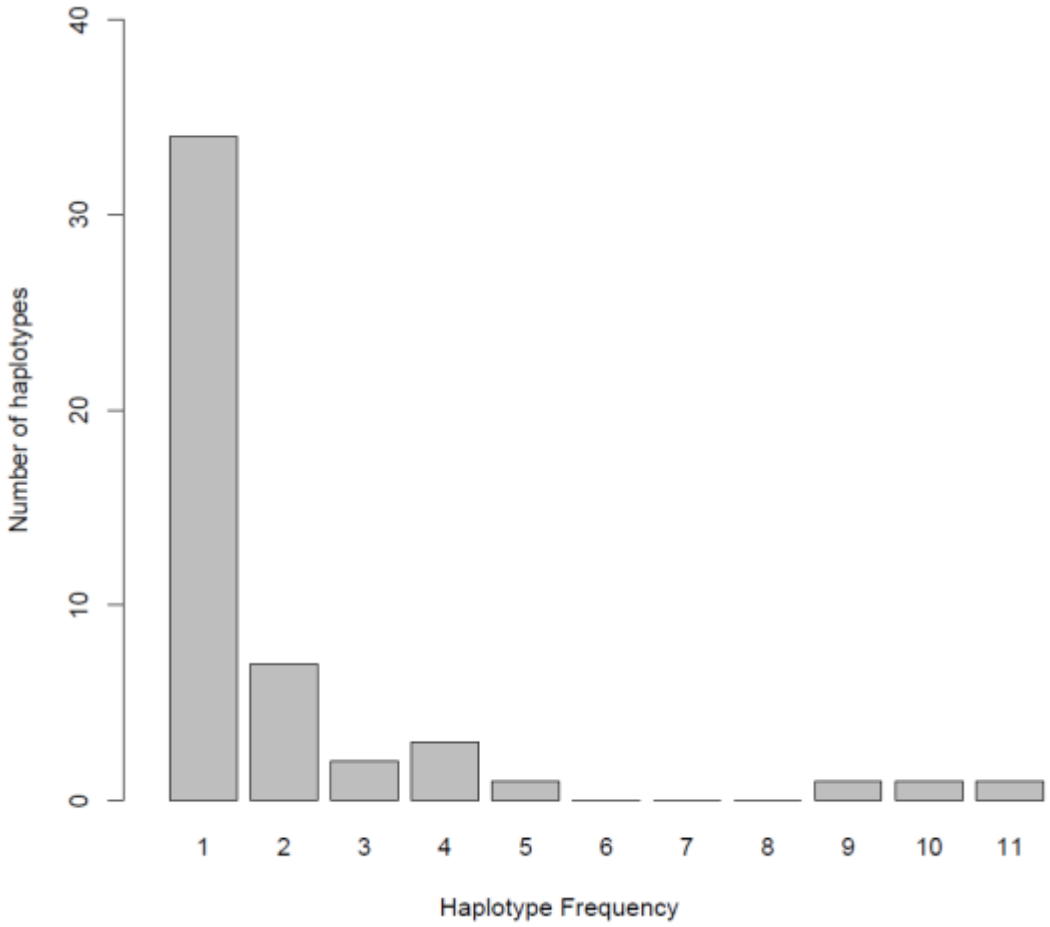


**Figure 2**. HVR-I and -II haplotype frequency spectrum of WE-NA mtDNAs in Puerto Rico.

*J1b1a1*. The three Puerto Rican J1b1a1 haplotypes (Figure S1) form an exclusive cluster with the second most frequent J1b1a1 haplotype in Puerto Rico ($n$ = 4, Table S2) at a central position, differing from the remainder haplotypes by single transitions. This J1b1a1 haplotype differs from the central node of the network by mutations at nps 462 and 489. These sites were not sequenced in the referenced studies and thus it is reasonable to assume that the second most frequent Puerto Rican haplotype corresponds to the central node. The central node is represented mostly by samples from Russia and Poland, but also from Italy and Spain, suggesting that the most represented WE-NA subclade in Puerto Rico may have its origin in a population ancestral to any of these countries.

*U5b*. Subhaplogroup U5b shows five control region haplotypes in Puerto Rico.

The most frequent haplotype ($n$ = 10), corresponding to the U5b1b1b clade, is one mutational step away (np 16320) from a node with equal amount of samples from Crete, Austria and Germany (Figure S2). Clade U5b1b1b is regarded as specific to West Africa, associated to the Serer, Wolof and Fulani populations from Senegal and northern Cameroon (Rando et al. 1998; Achilli et al. 2005; Coia et al. 2005; Cerny et al. 2006). The scarcity of HVR-II sequences of samples from these populations explains why this most frequent U5b control region haplotype in Puerto Rico finds only a match with a Spaniard sample.

The second most frequent haplotype ($n$ = 9) is separated from its nearest node by transitions at nps 16224 and 279, which together define the U5b2b3a clade. The nearest node is shared only by Spain, France and Saudi Arabia, and this haplotype is thus likely to have originated in Western Europe. One of the remaining Puerto Rican U5b haplotypes shares the central node of the network with Italy, Russia, Germany, Crete and France, and another is unique, located one mutation away (np 16311) from this central node. The last haplotype is also unique, and is separated by another transition at np 16311 from a Bosnian sample, thus suggesting a Balkan origin.

*H1b.* In the H1b network (Figure S3), two nodes are found distinguished by a single transition at np 16362, which groups subclades H1b1a, b, c, d, and h (van Oven and Kayser 2009). All Puerto Rico H1b samples lack the 16362 transition. Two of the Puerto Rican samples are distinguished by a unique transition at np 199. Six others possess a transition at np 152 which is shared only with Polish mtDNAs.

*J2a.* The J2a median-joining network is divided into two well-defined clusters separated by transitions at nps 16231, 16261 and 152 representing subhaplogroups J2a1a1 to the left and J2a2 to the right of the network (Figure S4). Because all samples within J2a1a1 lacking the transitions at nps 319 and 489 were not sequenced at these sites, it is highly probable that they form a large central node for J2a1a1 shared by all Puerto Rican samples that is likely the founder haplotype of J2a1a1. This haplotype is common among non-Ashkenazi Jewish populations: of the 11 non-Puerto Rican samples, 7 are known Jewish samples, including four Spanish exilers, two Libyan Jews and one Moroccan Jew (Behar et al. 2008).

*T-16304.* All haplogroup T samples exhibiting a transition at np 16304 were chosen to construct the T2b network (Figure S5). A transition at np 16304 defines the T2b clade (van Oven and Kayser 2009). Although 16304 also defines subclade T1a1n, T1 is defined by a transition at np 16186, which appeared in only one sample in the network. Thus, the presence of two main nodes separated by a single transition at np 16296 (which defines subhaplogroup T2 (van Oven and Kayser 2009)) is the result of the instability of the 16296 site in the T background, and not of having large numbers of T1 and T2 samples in the network. Indeed, the 16296 transition plays a role in numerous reticulations in the network (Figure S5). The transition at

np 16344 forms a monophyletic group in the network composed only of Puerto Rican and Jewish samples (from Tunisia and from a Spanish exiler in Turkey). In addition, the only haplotype different to all other T2b haplotypes in Puerto Rico derived from the latter accumulating four additional mutations, suggesting this could represent a very old, as yet non-described clade of Jewish origin.

In summary, only three of the seven most frequent WE-NA haplotypes in Puerto Rico, belonging to clades H1b, T2b and U5b1b1b, were shared by a single population or ethnic group in the median-joining networks, and we believe that in one of these cases (U5b1b1b) a shared haplotype is not a reliable indicator of origin because of recent back-migrations (see DISCUSSION). Furthermore, seven of the remaining eight, low-frequency, Puerto Rican haplotypes in these networks were not shared with any population or ethnicity and the one who did, belonging to U5b, was shared with several populations. Median-joining networks are highly informative phylogenetic analytical tools that nevertheless require highly informative sequences for which the HVR-I region alone is insufficient. Hence, to include in our analyses geographic regions for which most sequences available in the literature are restricted to HVR-I and are not suitable for median-joining network analysis, we employed posterior probability and principal component methods.

**Posterior probability analysis.** Our posterior probability (Pos) analysis of all subhaplogroups combined does not fit the historical expectation of a major contribution to the Puerto Rican WE-NA mtDNAs from the Iberian Peninsula (Table 4, Figure S6). Instead, WE-NA mtDNA sequences in Puerto Rico are in general more likely to originate in the Canary Islands with a Pos value of 14.23% (SD=3.48%), or from the Wolof and Serer from Senegal and the nomad Fulani of West Africa (Pos = 10.38%, SD = 3.03%). However, the contribution of these West African populations seems to be largely restricted to haplogroup U (Table 4, Figure S6), suggesting there exists no population in the database which carries a similarly diverse distribution of haplogroups as found in the Puerto Rican WE-NA mtDNAs, with the possible exception of the Canary Islands.

In consistency with this pattern, the most common Puerto Rican WE-NA haplogroups present origin probabilities widely different for different haplogroups. Whereas haplogroup U has a very strong Northwestern African component, including the Canary Islands and the West African populations mentioned previously, Puerto Rican haplogroup J mtDNAs are better represented by Scandinavia (Pos = 13.10%, SD = 7.03%), Turkey (Pos = 12.23%, SD = 6.83%), and Moroccan Jews (Pos = 10.15%, SD = 6.30%), followed by Northern European regions such as North Central (Pos = 8.77%, SD = 5.90%), North Eastern (Pos = 7.62%, SD = 5.53%), and North Western Europe (Pos = 7.11%, SD = 5.36%). Haplogroup H, on the other hand shows no region or population with a particularly high probability of origin (Table 4). Hence, our posterior probability analysis

suggests that WE-NA mtDNAs in Puerto Rico are likely derived from a wide geographical range, with the Canary Islands as the largest contributor.

**Table 4.** Posterior Probability of Origin (Pos) for Puerto Rican WE-NA mtDNAs
[a]Sample numbers are presented in the order: Global, H, J, U

| Population | *n*[a] | Rank | Global Pos (SD), % | Rank | H Pos (SD), % | Rank | J Pos (SD), % | Rank | U Pos (SD), % |
|---|---|---|---|---|---|---|---|---|---|
| Canarian | 278, 113, 21, 73 | 1 | 14.23 (3.48) | 2 | 6.57 (4.68) | 17 | 1.40 (2.45) | 1 | 34.04 (8.25) |
| Fulani + Wolof + Serer | 27, 1, 7, 18 | 2 | 10.38 (3.03) | 23 | 2.12 (2.72) | 21 | 0.56 (1.56) | 2 | 30.11 (7.99) |
| North Central European | 1147, 528, 106, 174 | 3 | 5.19 (2.21) | 4 | 5.77 (4.41) | 4 | 8.77 (5.90) | 17 | 1.18 (1.88) |
| Scandinavian | 312, 163, 27, 55 | 4 | 4.85 (2.14) | 15 | 2.52 (2.96) | 1 | 13.10 (7.03) | 15 | 1.26 (1.94) |
| North Western European | 661, 309, 74, 80 | 5 | 4.61 (2.09) | 8 | 4.35 (3.86) | 6 | 7.11 (5.36) | 7 | 1.92 (2.39) |
| North Eastern European | 798, 335, 57, 182 | 6 | 4.50 (2.06) | 5 | 5.47 (4.30) | 5 | 7.62 (5.53) | 9 | 1.58 (2.17) |
| Turk | 202, 58, 22, 42 | 7 | 4.40 (2.04) | 20 | 2.16 (2.75) | 2 | 12.23 (6.83) | 22 | 0.81 (1.56) |
| Moroccan Jew | 146, 61, 14, 10 | 8 | 3.91 (1.93) | 19 | 2.16 (2.75) | 3 | 10.15 (6.30) | 10 | 1.57 (2.16) |
| Alpine | 487, 220, 48, 91 | 9 | 3.39 (1.80) | 7 | 4.63 (3.97) | 15 | 1.91 (2.85) | 14 | 1.36 (2.02) |
| Basque | 155, 92, 4, 22 | 10 | 3.38 (1.80) | 3 | 5.77 (4.41) | 14 | 2.03 (2.94) | 8 | 1.66 (2.23) |
| North Western African | 243, 110, 17, 50 | 11 | 3.36 (1.79) | 22 | 2.14 (2.73) | 18 | 0.91 (1.98) | 4 | 3.95 (3.39) |
| Iraqi Jew | 135, 16, 30, 21 | 12 | 3.25 (1.77) | 23 | 2.12 (2.72) | 10 | 3.88 (4.03) | 18 | 1.18 (1.88) |
| West Mediterranean | 783, 390, 56, 122 | 13 | 3.23 (1.76) | 11 | 3.12 (3.29) | 7 | 5.35 (4.69) | 6 | 1.98 (2.43) |
| Spanish Exile Jew | 213, 82, 24, 23 | 14 | 3.14 (1.74) | 21 | 2.15 (2.74) | 8 | 4.46 (4.30) | 5 | 2.21 (2.56) |
| Algerian Jew | 20, 10, 1, 2 | 15 | 2.69 (1.61) | 18 | 2.16 (2.75) | 21 | 0.56 (1.56) | 3 | 5.37 (3.92) |
| Caucasian | 472, 134, 36, 107 | 16 | 2.64 (1.59) | 12 | 3.08 (3.27) | 9 | 4.45 (4.30) | 16 | 1.18 (1.88) |
| East Mediterranean | 590, 249, 55, 92 | 17 | 2.55 (1.57) | 6 | 4.90 (4.08) | 19 | 0.78 (1.83) | 23 | 0.80 (1.55) |
| Middle Eastern | 851, 195, 91, 124 | 18 | 2.41 (1.53) | 10 | 4.01 (3.71) | 12 | 2.94 (3.52) | 20 | 1.01 (1.74) |
| Kurd | 78, 24, 5, 14 | 19 | 2.36 (1.51) | 9 | 4.01 (3.71) | 13 | 2.91 (3.51) | 11 | 1.46 (2.09) |
| Central Mediterranean | 726, 305, 59, 102 | 20 | 2.35 (1.51) | 14 | 2.85 (3.14) | 16 | 1.85 (2.81) | 13 | 1.39 (2.04) |
| South Eastern European | 407, 165, 36, 76 | 21 | 2.03 (1.40) | 13 | 2.90 (3.17) | 20 | 0.72 (1.77) | 19 | 1.06 (1.79) |
| Azeri Jew | 58, 8, 35, 5 | 22 | 1.78 (1.32) | 1 | 9.98 (5.66) | 21 | 0.56 (1.56) | 25 | 0.00 (0.00) |
| Tunisian Jew | 36, 15, 0, 5 | 23 | 1.77 (1.31) | 17 | 2.18 (2.76) | 28 | 0.00 (0.00) | 12 | 1.45 (2.08) |
| Libyan Jew | 80, 27, 4, 4 | 24 | 1.44 (1.19) | 23 | 2.12 (2.72) | 11 | 3.50 (3.83) | 21 | 0.88 (1.62) |
| Yemeni Jew | 139, 5, 30, 14 | 25 | 1.36 (1.15) | 23 | 2.12 (2.72) | 21 | 0.56 (1.56) | 25 | 0.00 (0.00) |
| Iranian Jew | 75, 25, 13, 12 | 26 | 1.27 (1.12) | 16 | 2.27 (2.82) | 21 | 0.56 (1.56) | 24 | 0.59 (1.33) |
| Georgian Jew | 74, 8, 4, 3 | 27 | 1.13 (1.05) | 23 | 2.12 (2.72) | 21 | 0.56 (1.56) | 25 | 0.00 (0.00) |
| Near & Mid Eastern Jew | 34, 8, 7, 2 | 28 | 1.08 (1.03) | 23 | 2.12 (2.72) | 21 | 0.56 (1.56) | 25 | 0.00 (0.00) |
| Cochin Jew | 44, 0, 0, 6 | 29 | 0.86 (0.92) | 30 | 0.00 (0.00) | 28 | 0.00 (0.00) | 25 | 0.00 (0.00) |
| Mumbaikar Jew | 34, 3, 0, 0 | 30 | 0.43 (0.65) | 23 | 2.12 (2.72) | 28 | 0.00 (0.00) | 25 | 0.00 (0.00) |

**Principal component analyses.** In our first global PCA the first two principal components captured more than half of the genetic variability, PC1 38.54% and PC2 24.93%. PC1 seems to have been driven by populations likely having undergone genetic drift, separating mostly the Indian Jew populations to the right from the bulk of the other populations, from which the Basque protrude to the left (Figure S7). PC2 separates the West African populations and the Azeri Jews to the bottom of the plot from the Mumbaikar and Libya Jews to the opposite pole. Puerto Rico is close to several populations in the PC1-PC2 plot, as well as when PC1 is plotted against PC3, which explains 14.02% of the variation (Figure S8). To better understand the relationship of Puerto Rico WE-NA mtDNAs to these populations, we added the absolute differences between the eigenvalues of Puerto Rico and any given population for all 25 PCs calculated, which combined accounted for 100% of the variation. The results are shown in Table 5.

The composition of Puerto Rico WE-NA mtDNA subhaplogroups resembles more closely that of populations located in the northeastern coast of the Mediterranean Sea and further inland. Four of the five populations with the least absolute differences are from this region including Caucasians (0.928), Turks (0.930), East Mediterraneans (0.982) and South East Europe (0.984). North Central Europe was the only population outside of the region with similar total eigenvalue differences with Puerto Rico (0.984).

**Table 5**. Sum of Egenvalues Absolute Differences with Puerto Rico for the First 25 PCs

| Population | Sum |
|---|---|
| Puerto Rico | 0 |
| Caucasian | 0.928492535 |
| Turk | 0.930114982 |
| East Mediterranean | 0.982070254 |
| North Central European | 0.984187878 |
| South East European | 0.984365802 |
| Spanish Exile Jew | 0.994187894 |
| North East European | 1.013600768 |
| Middle Eastern | 1.021054132 |
| Central Mediterranean | 1.037297849 |
| Iranian Jew | 1.048964361 |
| Kurd | 1.059468436 |
| Alpine | 1.061142711 |
| North West European | 1.075248659 |
| Iraqi Jew | 1.097384013 |
| Canarian | 1.111158047 |
| Moroccan Jew | 1.128515084 |
| West Mediterranean | 1.146220672 |
| Scandinavian | 1.178065801 |
| Northwestern African | 1.184606213 |
| Tunisian Jew | 1.194190322 |
| Algerian Jew | 1.253456929 |
| Near and Middle Eastern Jew | 1.263396212 |
| Yemeni Jew | 1.355252559 |
| Basque | 1.379198264 |
| Libyan Jew | 1.417913777 |
| Georgian Jew | 1.449869361 |
| Fulani-Wolof-Serer | 1.533771411 |
| Azeri Jew | 1.600948512 |
| Cochin Jew | 1.634802469 |
| Mumbaikar Jew | 1.782730821 |

We also performed individual principal component analyses for haplogroups H, J and U. For haplogroup H, PC1 contained 91.21% of the variation, separating several Jewish populations to the far left of the plot from Near Eastern populations at the right pole (Figure S7). With the exception of Scandinavians, all European populations clustered next to each other between PC1 values 0.162 (West Mediterranean) and 0.267 (North East Europeans). Puerto Rico was located very close to the cluster at PC1 value 0.072, and closest within the cluster to the West Mediterranean population. For haplogroup J, PC1 captured 84.25% of the variation and produced a tightly knit cluster at the left of the plot containing, between values -0.349 and -0.408, all European populations with the exception of Scandinavians and Basques, including North Western Africans, Kurds, Caucasians, and Middle

Easterners, and excluding all Jewish populations, Turks, Canarians and West Africans. The PC1 value of Puerto Rico (-0.320) located it very close to this cluster at the PC1 axis; however, most of the variation contained in PC2, which encompassed 10.3% of the total variation, separated Puerto Rico away from the rest of the populations. The haplogroup U PCA produced another tight cluster of populations close to Puerto Rico in principal components one to three, which contained 62.14%, 15.51%, and 10.56% of the variation, respectively (Figures S7 and S8). The cluster was maintained in all three PCs and contained all European populations except Scandinavia, Northwestern Africa, Spanish Exile and Yemeni Jews, Kurds, Middle Easterners, and Caucasians.

## Discussion

The total WE-NA genetic composition of Puerto Rico depends on the number of migrants, their geographic origin, and their reproductive success in Puerto Rico, which in turn is a function of the number of generations in Puerto Rico and the average number of descendants per generation. Moreno-Estrada et al. (2013) showed that in spite of the large number of migrants arriving to Puerto Rico from Europe during the first half of the 18$^{th}$ century (Cifre de Loubriel 1964, Fernández Méndez 1970), migrations from Sub-Saharan Africa produced a larger change in the ancestral composition of the Puerto Rican population. This observation can be explained by a population expansion produced by an increase in the reproductive success of the admixed locals as opposed to migration from Europe. In such a case, the WE-NA origins and composition of Puerto Rico may be driven more by the small number of early migrants than by the large number of migrants arriving later in history.

However, this study focuses on WE-NA ancestry of Puerto Rico mediated by women, whose migrations to Puerto Rico were notably rare in the early centuries of European colonization (Fernández Méndez 1970). The colonization of the Americas by the Spaniards was characterized predominantly by the migration of men, with the result of strong asymmetries between male (Y-DNA) and female (mtDNA) ancestries in the admixed populations (Carvalho-Silva et al. 2001; Bryc et al. 2010). This has been confirmed for Puerto Rico (Vilar et al. 2014), and also the Canary Islands (Pinto et al. 1994, 1996), where a large proportion of mtDNAs, but not nuclear markers, were found to be of Guanche origin, in addition to those belonging to other subjugated populations such as Berbers or Guineans.

Thus, the bimodality of the obtained haplotype frequency distribution (Figure 2) can be explained by the presence of a few early-arriving haplotypes that gained large numbers through population expansion, and a large number of low-frequency haplotypes representing late arrivals. This very large number of low-frequency haplotypes may be the reason why a population expansion was not detected by the Tajima's *D* test (Table 3). However, the very high frequency of a low number of haplotypes is outstanding (Figure 2). Drift has played a big role in

the make-up of the Puerto Rican WE-NA composition.

WE-NA mitochondrial ancestry accounts for only 11.5% of maternal haplogroups in the modern Puerto Rican population. Most of this maternal ancestry has been attributed to Western European sources (Vilar et al. 2014). However, using Fisher's Exact tests, median-joining networks, posterior probability analyses, and principal component analyses, we found that the origins of the WE-NA female ancestry of modern Puerto Ricans may be varied, with strong contributions from Spain, the Canary Islands and populations related to contemporary Fulani, Serer and Wolof populations of West Africa. Substantial contributions were also detected from elsewhere in the circum-Mediterranean region and inside Europe. Some of these inputs may have originated among Jewish populations.

Because haplotypes shared between populations are strong evidence for recent migrations, haplotype networks are powerful tools to highlight such processes. Of the 7 major (most frequent) Puerto Rican WE-NA haplotypes ($\geq$4), two were shared with only one population. The H1b haplotype 16189-16356-152-263 was shared only with two Polish samples (Malyarchuk et al. 2002), and U5b1b1b (16189-16192-16270-16320-73-150-263) was shared only with one Spanish sample (Álvarez et al. 2007). The latter connection is likely due to back-migration. U5b1b1b is common among the nomad Fulani of West Africa and the Wolof and Serer populations of Senegambia (Rando et al. 1998, Rosa et al. 2004, Achilli et al. 2005, Coia et al. 2005). Its ancestral clade U5b1b (control region motif 16189-16192-16270-73-150-263), has been proposed together with haplogroups H1, H3 and V to represent hunter-gatherer migrations from the Franco-Cantabrian refuge to North Africa during the Ice Age (Achilli et al. 2005). The lineage diverged in Africa into U5b1b1e, characterized by a transition at np 152 and found almost exclusively among Berbers (Achilli et al. 2005), and into U5b1b1b, characterized by a transition at np 16320. U5b1b1b thus arose in the African continent, and because it is virtually absent north of Senegambia, it probably arose only after a coastal migration led to admixture events with Sub-Saharan African populations in Senegambia.

According to documentary sources, Senegambia, where the Wolof, Serer, and some Fulani can be found, was the first region of Sub-Saharan Africa to be exploited for the African Slave Trade. This region was abandoned by the traders in the second half of the 16$^{th}$ century, who moved their trade to the Bight of Biafra and nearby coasts (Thomas 1997). It has also been documented that Wolofs were among the first enslaved Africans sold in the new colony of San Juan (Puerto Rico) early in the 16$^{th}$ century, but were later imported in lower numbers because of their renowned resistance to the conditions of slavery (Alegría 1985). In addition, molecular and chromosome recombination evidence based on the probability of a population of being the source of short *vs* long chromosomal fragments of Sub-Saharan African ancestry supports Senegambia as a source of enslaved Africans for

Puerto Rico and the Caribbean prior to the Bight of Biafra (Moreno-Estrada et al. 2013). It is noteworthy that despite its West African origin, U5b1b1b has not been found among African-Americans (Just et al. 2008), nor in the English-speaking Caribbean (Benn-Torres et al. 2007), Cuba (Mendizabal et al. 2008), or Dominican Republic (unpublished results). Thus, it is apparent that the arrival of the U5b1b1b haplotype to Puerto Rico was a rare event that occurred in the first decades of the Spanish colonization, and that its numbers expanded through the centuries not by additional migrations, but by reproduction. Its high frequency in Puerto Rico is likely due to genetic drift.

Another major Puerto Rican WE-NA haplotype was shared by only two populations, both Jewish. Haplotype T2b 16126-16294-16296-16304-16344-73-151-152-263 was shared with a Tunisian Jew and with a Spanish Jew exile in Turkey, suggesting migration of Jewish women in the early colonization of Puerto Rico. Of the four remaining major Puerto Rican WE-NA haplotypes, two were shared with multiple populations and two with none. The J2a1a1 founder haplotype (16069-16126-16145-16231-16261-73-150-152-195-215-263-295) was shared with four Spanish Jew exiles, two Libyan Jews and one Moroccan Jew, in addition to two Poles, one Bosnian and one Spaniard, and is thus likely to be of Jewish ancestry. The J1b1a1 haplotype (16069-16126-16145-16172-16222-16261-73-242-263-295) was shared with three Poles, three Russians, one Spaniard and one Italian. The two major Puerto Rican haplotypes that were not shared belonged, together with U5b1b1b, to the trio of WE-NA haplotypes most frequent in Puerto Rico (Figure 2): the J1b1a1 haplotype (16069-16126-16145-16172-16222-16261-73-152-242-263-295), and the U5b2b3a haplotype (16224-16270-73-150-263-279).

In an effort to increase our certainty on the origin of the latter four haplotypes, we searched the 1000 Genomes Project phase 3 database for Puerto Rico, which contains the complete mtDNA sequence of 104 unrelated samples, but found useful information only for both J1b1a1 haplotypes. Although these two haplotypes differ by only one transition at np 152, they differ by five transitions in their coding region. Specifically, HG01098 and HG01111, who correspond to the second most frequent J1b1a1 haplotype in Puerto Rico ($n$ = 4, Table S2), share four private transitions at nps 3324, 9438, 14560 and 15740. On the other hand, HG01302 and HG01308, with a control region sequence identical to the most frequent haplotype in Puerto Rico ($n$ = 11), share a private transition at 13943 (Table S10). This extensive divergence suggests subhaplogroup J1b1a1 may represent a very old lineage that, in spite of its low frequency worldwide, has become geographically widespread, having been reported as far east as Iraq, and as far west as Spain and Morocco (Behar et al. 2008). Despite their highly similar coding region sequences, the two most frequent J1b1a1 haplotypes in Puerto Rico could well have originated from far apart regions.

Three of the four mutations that characterize the second most frequent

J1b1a1 haplotype in Puerto Rico (3324, 9438 and 15740) appeared in only one 1000 Genomes Project sample outside of Puerto Rico: HG01694, from the IBS (Iberian populations in Spain) sample set. We believe it is unlikely that this haplotype is widespread outside of Spain because its 15740 mutation is predicted by Polyphen-2 to be probably damaging with a score of 1.00 (Adzhubei et al. 2010). Damaging mutations are usually recent and restricted to their respective populations (The 1000 Genomes Consortium 2012). It is thus likely that the second most frequent J1b1a1 haplotype in Puerto Rico originated in Spain, which was one of the candidate countries suggested by the haplotype network analysis (Figure S1). The network contains several Jewish samples, including one Spanish Jew exile sequence located one mutational step away from the central haplotype. We conclude that the second most frequent J1b1a1 haplotype in Puerto Rico may have originated in Spain, and cannot rule out a Spanish Jewish origin.

On the other hand, none of the remaining 2502 mtDNA samples of the 1000 Genomes Phase 3 database possesses the np 13943 transition unique to HG01302 and HG01308. It is thus apparent that the transition at np 13943 characterizes the most frequent J1b1a1 haplotype of Puerto Rico. Furthermore, its absence in the IBS sample set increases the likelihood of an origin outside of Spain, more probably where its clade exists at higher frequencies and variability. Ancient DNA studies have associated J1b1a1 with the diffusion of Proto-Germanic and Proto-Celto-Italic speakers, and its frequencies are higher in North-Central Europe (Richards et al. 1996, 1998, 2000, 2002; Parson et al. 2004). This distribution may explain the large posterior probability for this region as one of the likely sources of Puerto Rican J mtDNAs (Table 4). However, the posterior probability analysis suggested other probable sources such as Moroccan Jews, or Turks. Battles between the Ottoman Empire and Spanish-led Catholic coalitions were common during the 16$^{th}$ century, and could have led to the introduction of Near Eastern mtDNAs to Spanish colonies in the Americas (Crowley 2008).

A search in the EMPOP Haplogroup Browser Database (empop.online) showed that U5b2b3a, defined by transitions at nps 16224 and 279 (van Oven and Kayser 2009), is found only in the Iberian Peninsula and in the Americas, with a much higher frequency in the Iberian Peninsula. Its derived clade U5b2b3a1 has been found only in the Americas and is defined by a transition in the coding region that was not tested in this study (np 9494). However, none of the Puerto Rican U5b2b3a samples in the 1000 Genomes database possess the transition that defines U5b2b3a1 and are thus likely to belong to U5b2b3a. In conclusion, U5b2b3a and its derived clades seem to represent a lineage probably restricted to the Iberian Peninsula. A substantial group of women settlers carrying these clades may have participated in the early decades of the Spanish colonization of the Americas, but only U5b2b3a made it to Puerto Rico.

We also searched the 1000 Genome Project Puerto Rico database for the

T2b, U5b1b1b and H1b Puerto Rican haplotypes of high frequency but found only the H1b haplotype. HG01167 lacked all transitions that define H1b subclades, suggesting that the most common H1b haplotype in Puerto Rico is basal in its phylogeny. Subhaplogroup H1b is widespread across Europe and present in very low frequencies in North African populations (Ennafaa et al. 2009; Richards et al. 2000). Since all H1b mtDNAs found in the North-Central Caucasus possess the transition at np 3796 that defines H1b1 (Roostalu et al. 2007), it is reasonable to conclude that the most common H1b haplotype of Puerto Rico does not originate in the North-Central Caucasus. In the H1b median network (Figure S3) the most common H1b Puerto Rican haplotype is shared only with Polish samples, differing from the central node of the network only by a transition at np 152, which is the most unstable site in the whole mitochondrial genome after np 16519 (Soares et al. 2009). The central node of this network is shared only by Eastern European or Eastern Mediterranean samples (Poland, Finland, Russia, Bosnia and Crete). On the other hand, two H1b samples sharing their HVR-I haplotype with the most common H1b haplotype in Puerto Rico have been described among Moroccan Berbers (Ennafaa et al. 2009). Thus, the most common H1b haplotype in Puerto Rico may originate from Eastern Europe, the Eastern Mediterranean, or Moroccan Berbers. On its way to Puerto Rico, it could have passed by the Canary Islands, as its HVR-I haplotype is found there (see below). Other minor H1b Puerto Rican haplotypes may have arrived through the Canary Islands as well, such as the H1b sample carrying the 16257 transition found only in this population (Rando et al. 1999).

In summary, for the three haplotypes with the highest frequencies in Puerto Rico, the likeliest origins include Sub-Saharan Africa (U5b1b1b), Spain (U5b2b3a), and other regions sharing a coast to the Mediterranean Sea or in North Central Europe (J1b1a1). Of the remaining four Puerto Rican haplotypes with frequencies $\geq$ 4, the J1b1a1 haplotype differed by only one coding region transition from a Spanish sample, H1b is likely from Eastern Europe, the Eastern Mediterranean, or Moroccan Berbers, and the T2b and J2a1a1 haplotypes are likely of Jewish origin, especially T2b. Sephardic Jews are known to have been among the ethnic groups engaging in migrations to the Puerto Rico (Fernández Méndez 1970), and their imprint is probably present as well among the less frequent haplotypes on the island. For instance, three Puerto Rican samples possess the transition at np 497 in a K1a lineage that has been found in high frequencies in Ashkenazi Jewish populations (Table S2, Behar et al. 2006).

It is noteworthy that the Posterior Probability and PCA analyses gave very different results despite using the exact same data (Table S11). Whereas PCA identifies the variables and covariants that are responsible for most of the variation among all populations in the database, the posterior probability analysis estimates the relative probability of origin for each Puerto Rican subhaplogroup among the

populations in the database, and is therefore heavily influenced by subhaplogroups that are highly frequent in Puerto Rico. For example, the West African group in the reference database (Fulani + Wolf + Serer) was the only metapopulation with more than one U5b1b1b mtDNA (15 of its 27 WE-NA samples). Hence, it is highly probable that U5b1b1b mtDNAs in Puerto Rico originated from that population. Thus, even though other subhaplogroups highly frequent in Puerto Rico, such as J1b1a1, H1b and U5b2b3a, were absent in this West African population, the very high frequency of U5b1b1b in Puerto Rico gave weight to its West African connection to the point that this population obtained the second highest overall probability of origin (Table 4). Its 10.4% value is very close to the frequency of U5b1b1b in Puerto Rico (9.9%), and implies that approximately 10.4% of the Puerto Rican WE-NA mtDNAs originate from West Africa. By contrast, in the combination of the 25 principal components of the global PCA analysis, the West African population was not closer to Puerto Rico relative to other populations (Table 5).

Our posterior probability analyses show that WE-NA mtDNA sequences in Puerto Rico are generally more likely to originate in the Canary Islands with a probability of 14.2%. This is not surprising as U6b1a and the haplogroup H HVR-I haplotype 16260 found in Puerto Rico (Table S2) are considered founders of the Guanche population. In addition, two samples bearing the North African Berber motif (16172-16189-16219-16278), corresponding to U6a1 and found in the Canary Islands (Rando et al. 1999; Maca-Meyer et al. 2004), were also identified in Puerto Rico. Several HVR-I sequences found in modern admixed Canary Islanders are also present in Puerto Ricans (Pinto et al. 1996; Rando et al. 1999; Maca-Meyer et al. 2004; Martínez-Cruzado et al. 2005; Fregel et al. 2009). More specifically, 14 of our 37 Puerto Rican HVR-I haplotypes are shared with some Canarian samples from Rando et al. (1999), represented by 38 of our 101 WE-NA samples. These findings are consistent with historical sources describing multiple migration events from the Canary Islands to Puerto Rico after the 16$^{th}$ century, and especially at the beginning of the 18$^{th}$ century. Unlike migrations from other European sources, which consisted largely of single men, many groups migrating from the Canary Islands were composed of complete, farmer families (Cifre de Loubriel 1964; Fernández Méndez 1970; Borges Jacinto del Castillo 1969; Rodríguez Mendoza 2004). Hence, the data as a whole suggests that migrations of women under different conditions from Spain and elsewhere (including North Africa and Eastern Europe or the Eastern Mediterranean) to the Canary Islands eventually gave rise to women who extended their migration further to the Caribbean as part of whole families.

Regarding haplogroup H mtDNA sequences, H1 and H3 Iberian samples encompass most of the haplogroups H1 and H3 diversity (Pereira et al. 2006; Roostalu et al. 2007), and our PCA analysis for haplogroup H in Puerto Rico

supports a predominantly Iberian origin (Figure S7). The diversity within these haplogroups in Puerto Rico, 3 haplotypes in 5 samples in H1(xH1b) and 5 samples with different haplotypes in H3, is in stark contrast with the diversity in J1b1a1 (3 haplotypes in 16 samples), U5b1b1b (1 haplotype in 10 samples) and U5b2b3a (1 haplotype in 9 samples), suggesting they mostly represent later migrations to Puerto Rico such as those occurring during the 19th century, which reportedly entailed mostly the migration of families from Western Europe (Cifre de Loubriel 1964).

In conclusion, the Canary Islands seem to have been a major source of WE-NA mtDNAs to Puerto Rico. However, most of the WE-NA mtDNAs arriving to Puerto Rico from the Canary Islands were not native to the Canary Islands. Furthermore, most of the Puerto Rican WE-NA haplotypes may not have arrived to Puerto Rico through the Canary Islands, as they are not found there today. It appears there are multiple ulterior origins for the Puerto Rican mtDNA haplotypes. Specific examples are U5b1b1b ($n$ =10) from Senegambia, U5b2b3a ($n$ = 9) from the Iberian Peninsula, T2b ($n$ = 4) from Jews residing in Tunisia or exiled from Spain, and U6a1 ($n$ = 2) from North Africa. As suggested from haplotypes belonging to J2a1a and K1a in addition to T2b, Jewish populations from diverse geographic regions are probably one of the main contributors to the mtDNA variation of admixed Puerto Ricans. As suggested by the H1b network and the global PCA results (Table 5), other contributors could have been from Eastern Europe or the northeastern coasts of the Mediterranean Sea. It is noteworthy that Turkey and Jews from Morocco in addition to North Central Europe are among the likeliest sources for the most frequent WE-NA haplotype in Puerto Rico, belonging to J1b1a1. The likely geographic origins of WE-NA mtDNAs in Puerto Rico span coasts north and south of the Mediterranean Sea.

---

Acknowledgments
We thank the people who, after being randomly selected, received us at their homes and willingly participated in the study with informed consent. This study was supported by the National Science Foundation Physical Anthropology Program grant SBR-9904252 to JCMC and the Faculty of Arts and Sciences of the University of Puerto Rico at Mayagüez.

## Literature Cited

Abu-Amero, K.K., J.M. Larruga, V.M. Cabrera et al. 2008. Mitochondrial DNA structure in the Arabian Peninsula. *BMC Evol. Biol.* 8:45.

Achilli, A., C. Rengo, V. Battaglia et al. 2005. Saami and Berbers: An unexpected mitochondrial DNA link. *Am. J. Hum. Genet.* 76:883-886.

Achilli, A., A. Olivieri, M. Pala et al. 2007. Mitochondrial DNA variation of modern Tuscans supports the Near Eastern origin of Etruscans. *Am. J. Hum. Genet*. 80:759-768.

Adzhubei, I.A., S. Schmidt, L. Peshkin et al. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7:248-249.

Alegría, R. 1985. Notas sobre la procedencia cultural de los esclavos negros de Puerto Rico durante la segunda mitad del siglo XVI. *Revista del Centro de Estudios Avanzados de Puerto Rico y el Caribe* 1:58-79.

Alshamali, F., A. Brandstatter, B. Zimmermann et al. 2008. Mitochondrial DNA control region variation in Dubai, United Arab Emirates. *Forensic Sci. Int. Genetics* 2:e9 – e10.

Álvarez, J.C., D.L.E. Johnson, J.A. Lorente et al. 2007. Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. *Legal Med*. 9:293-304.

Anderson, S., A.T. Bankier, B.G. Barrell et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457-465.

Andrews, R.M., I. Kubacka, P.F. Chinnery et al. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23:147.

Bandelt, H.J., V. Macaulay, and M. Richards. 2000. Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol. Phylogenet. Evol.* 16:8-28.

Behar, D.M., N. Metspalu, T. Kivisild et al. 2006. The matrilineal ancestry of Ashkenazi Jewry: Portrait of a Recent Founder Event. *Am. J. Hum. Genet.* 78:487-497.

Behar, D.M., N. Metspalu, T. Kivisild et al. 2008. Counting the founders: The matrilineal genetic ancestry of the Jewish Diaspora. *PLoS ONE* 3:e2062.

Benn Torres, J., R.A. Kittles, and A.C. Stone 2007. Mitochondrial and Y chromosome diversity in the English-Speaking Caribbean. *Ann. Hum. Genet.* 71:782-790.

Bogácsi-Szabó, E., T. Kalmar, T. Csanyi et al. 2005. Mitochondrial DNA of ancient Cumanians: Culturally Asian Steppe nomadic immigrants with substantially more

Western Eurasian mitochondrial DNA lineages. *Hum. Biol.* 77:639-662.

Borges Jacinto del Castillo, A.L. 1969. *El Archipiélago Canario y las Indias Occidentales, episodios históricos.* Ediciones Cultura Hispánica.

Brakez, Z., E. Bosch, H. Izaabel et al. 2001. Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann. Hum. Biol.* 28:295-307.

Brandstatter, A., H. Niederstatter, M. Pavlic et al. 2007. Generating population data for the EMPOP database: An overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. *Forensic Sci. Int.* 166:164-175.

Brau, S. 1904. *Historia de Puerto Rico.* New York: Appleton and Company.

Bryc, K., C. Vélez, T. Karafet et al. 2010. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc. Natl. Acad. Sci* (USA) 107:8954-8961.

Carvalho-Silva, D.R., F.R. Santos, J. Rocha et al. 2001. The phylogeography of Brazilian Y-chromosome lineages. *Am. J. Hum. Genet.* 68:281-286.

Cerny, V., M. Hajek, M. Bromova et al. 2006. MtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations. *Hum. Biol.* 78:9-27.

Cherni, L., B.Y. Loueslati, L. Pereira et al. 2005. Female gene pools of Berber and Arab neighboring communities in Central Tunisia: Microstructure of mtDNA variation in North Africa. *Hum. Biol.* 77:61-70.

Cifre de Loubriel, E. 1964. *La inmigración a Puerto Rico durante el siglo XIX*. San Juan, PR: Instituto de Cultura Puertorriqueña.

Coia, V., G. Destro-Bisol, F. Verginelli et al. 2005. MtDNA variation in North Cameroon: Lack of Asian lineages and implications for back migration from Asia to Sub-Saharan Africa. *Am. J. Phys. Anthropol.* 128:678-681.

Comas, D., F. Calafell, N. Bendukidze et al. 2000. Georgian and Kurd mtDNA sequence analysis shows a lack of correlation between languages and female genetic lineages. *Am. J. Phys. Anthropol.* 112:5-16.

Crowley, R. 2008. *Empires of the Sea: The Siege of Malta, the Battle of Lepanto, and the Contest for the Center of the World.* New York: Random House.

Díaz-Soler, L. M. 2000. *Historia de la esclavitud negra en Puerto Rico*. San Juan, Puerto Rico: Editorial de la Universidad de Puerto Rico.

Dubut, V., L. Chollet, P. Murail et al. 2004. MtDNA Polymorphysms in five French groups: importance of regional sampling. *Eur. J. Hum. Genet.* 12: 293-300.

Emmer, P. C. 1999. New Societies: The Caribbean in the long sixteenth century. In *General History of the Caribbean* (Vol. II). London: UNESCO Publishing.

Ennafaa, H., V.M. Cabrera, K.K. Abu-Amero et al. 2009. Mitochondrial DNA haplogroup H structure in North Africa. *BMC Genetics* 10:8.

Fernández Méndez, E. 1970. *Historia Cultural de Puerto Rico*. San Juan, PR: Ediciones El Cemí.

Fraumene, C., E. Petretto, A. Angius et al. 2003. Striking differentiation of subpopulations with a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis. *Hum. Genet.* 114:1-10.

Fregel, R., J. Pestano, M. Arnay et al. 2009. The maternal aborigine colonization of La Palma (Canary Islands). *Eur. J. Hum. Genet.* 17:1314-1324.

González, A.M., A. Brehm, J.A. Pérez et al. 2003. Mitochondrial DNA affinities at the Atlantic Fringe of Europe. *Am. J. Phys. Anthropol.* 120:391-404.

González, A.M., V.M. Cabrera, J.M. Larruga et al. 2006. Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations. *Ann. Hum. Genet.* 70:631-657.

Gravel, S., F. Zakharia, A. Moreno-Estrada et al. 2013. Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genet.* 9:e100023.

Hedman, M., A. Brandstatter, V. Pimenoff et al. 2007. Finnish mitochondrial DNA HVS-I and HVS-II population data. *Forensic Sci. Int.* 172:171-178.

Just, R.S., T.M. Diegoli, J.L. Saunier et al. 2008. Complete mitochondrial genome sequences for 265 African American and U.S. 'Hispanic' individuals. *Forensic Sci. Int. Genet.* 2:E45-E48.

Kivisild, T., M. Reidla, E. Metspalu et al. 2004. Ethiopian mitochondrial DNA Heritage: Tracking gene flow across and around the Gate of Tears. *Am. J. Hum. Genet.* 75:752-770.

Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.

Loogvali, E.L., U. Roostalu, B.A. Malyarchuk et al. 2004. Disuniting uniformity: A pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol. Biol. Evol.* 21:2012-2021.

Maca-Meyer, N., M. Arnay, J.C. Rando et al. 2004. Ancient mtDNA analysis and the origin of the Guanches. *Eur. J. Hum. Genet.* 12:155-162.

Malyarchuk, B.A., T. Grzybowsky, M.V. Derenko et al. 2002. Mitochondrial DNA variability in Poles and Russians. *Ann. Hum. Genet*. 66:261-283.

Malyarchuk, B.A., T. Grzybowsky, M.V. Derenko et al. 2003. Mitochondrial DNA variability in Bosnians and Slovenians. *Ann. Hum. Genet.* 67:412-425.

Malyarchuk, B.A., T. Vanecek, M.A. Perkova et al. 2006. Mitochondrial DNA variability in the Czech population, with application to the ethnic history of Slavs. *Hum. Biol.* 78:681-696.

Martínez-Cruzado, J.C., G. Toro-Labrador, J. Viera-Vera et al. 2005. Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis. *Am. J. Phys. Anthropol.* 128:131-155.

Martínez, L., S. Mirabal, J.R. Luis et al. 2008. Middle Eastern and European mtDNA lineages characterize populations from eastern Crete. *Am. J. Phys. Anthropol.* 137:213-223.

Mendizabal, I., K. Sandoval, G. Berniell-Lee et al. 2008. Genetic origin, admixture and asymmetry in maternal and paternal human lineages in Cuba. *BMC Evol. Biol.* 8:213.

Moreno-Estrada, A., S. Gravel, F. Zakharia et al. 2013. Reconstructing the population genetic history of the Caribbean. *PLoS Genetics* 9:e1003925.

Nei, M. 1987. *Molecular Evolutionary Genetics.* New York: Columbia University Press.

Palanichamy, M.G., C. Sun, S. Agrawal et al. 2004. Phylogeny of mitochondrial DNA Macrohaplogroup N in India, based on complete sequencing: Implications for the peopling of South Asia. *Am. J. Hum. Genet.* 75:966-978.

Parson, W., A. Brandstätter, A. Alonso et al. 2004. The EDNAP mitochondrial

DNA population database (EMPOP) collaborative exercises: organization results and perspectives. *Forensic Sci. Int.* 139:215-226.

Pereira, L., M. Richards, A. Goios et al. 2006. Evaluating the forensic informativeness of mtDNA Haplogroup H sub-typing on an Eurasian scale. *Forensic Sci. Int.* 159:43-50.

Picó, F. 2004. *Historia General de Puerto Rico*. San Juan, PR: Ediciones Huracán.

Pinto, F., V.M. Cabrera, A.M. González et al. 1994. Human enzyme polymorphism in the Canary Islands VI. Northwest African influence. *Hum. Hered.* 44:156-161.

Pinto, F., A.M. González, M. Hernández et al. 1996. Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequence. *Ann. Hum. Genet.* 60:321-330.

R Development Core Team. 2013. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing, 1.9.1.

Rando, J.C., F. Pinto, A.M. Gonzalez et al. 1998. Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Ann. Hum. Genet.* 62:531-550.

Rando, J.C., V.M. Cabrera, M. Larruga et al. 1999. Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann. Hum. Genet*. 63:413-428.

Richards, M., M. Corte-Real, P. Forster et al. 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* 59:185–203.

Richards, M.B., V.A. Macaulay, H.J. Bandelt et al. 1998. Phylogeography of mitochondrial DNA in Western Europe. *Ann. Hum. Genet.* 62:241-260.

Richards, M., V. Macaulay, E. Hickey et al. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67:1251-1276.

Richards, M., V. Macaulay, A. Torroni et al. 2002. In search of geographical patterns in European mitochondrial DNA. *Am. J. Hum. Genet.* 71:1168-1174.

Rodríguez Mendoza, F. 2004. *La emigración del noroeste de Tenerife a América durante 1750-1830*. Servicio de publicaciones, Universidad de La Laguna. Santa Cruz de Tenerife, Canary Islands, Spain: Serie tesis doctorales.

Roostalu, U., I. Kutuev, E.L. Loogvali et al. 2007. Origin and expansion of Haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: The Near Eastern and Caucasian Perspective. *Mol. Biol. Evol.* 24:436-448.

Rosa, A., A. Brehm, T. Kivisild et al. 2004. MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. *Ann. Hum. Genet.* 68:340–352 (erratum 68:658).

Santana Pérez, J.M. 2008. Isleños en Cuba y Puerto Rico (del siglo XVIII a mediados del siglo XIX). *Cuadernos Americanos* 126:173-192.

Sarno, S., A. Boattini, M. Carta et al. 2014. An ancient Mediterranean melting pot: investigating the uniparental genetic structure and population history of Sicily and Southern Italy. *PLoS ONE* 9:e96074.

Soares, P., L. Ermini, N. Thomson et al. 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* 84:740-759.

South, A. 2011. Rworldmap: A new R package for mapping global data. *The R Journal*. 3/1:35-43.

Tajima, F. 1989a. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585-595.

Tajima, F. 1989b. The effect of change in population size on DNA polymorphism. *Genetics* 123:597-601.

Tamura, K., D. Peterson, N. Peterson et al. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* 28:2731-2739.

Tetzlaff, S., A. Brandstatter, R. Wegener et al. 2007. Mitocondrial DNA population data of HVS-I and HVS-II sequences from a northeast German sample. *Forensic Sci. Int.* 172:218-224.

The 1000 Genomes Project Consortium 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56-65.

Thomas, H. 1997. *The slave trade: the story of the Atlantic slave trade: 1440-1870.* New York: Simon & Schuster.

Turchi, C., L. Buscemi, C. Previdere et al. 2008. Italian mitochondrial DNA

database: results of a collaborative exercise and proficiency testing. *Int. J. Legal Med.* 122:199-204.

van Oven, M., and M. Kayser 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mut.* 30:E386-E394.

Via, M., C.R. Gignoux, L.A. Roth et al. 2011. History Shaped the Geographic Distribution of Genomic Admixture on the Island of Puerto Rico. *PLoS ONE* 6:e16513.

Vilar, M.G., C. Melendez, A.B. Sanders et al. 2014. Genetic Diversity in Puerto Rico and Its Implications for the Peopling of the Island and the West Indies. *Am. J. Phys. Anthropol.* 155:352-368.

**Table S1.** HVR-I and HVR-II Haplotypes in 101 Samples from Modern Puerto Ricans

| Subhaplogroup | HVR-I | HVR-II | Frequency |
|---|---|---|---|
| H | 16260, 16519 | 263 | 1 |
| H1 | CRS | 263 | 1 |
| H1 | 16519 | 263 | 1 |
| H1 | CRS | 152, 199, 263 | 1 |
| H1 | 16296, 16519 | 263 | 1 |
| H1b | 16189, 16356 | 152, 263 | 4 |
| H1b | 16356[a1] | 152, 263 | 1 |
| H1b | 16189, 16356 | 73, 152, 263 | 1 |
| H1b | 16189, 16356 | 199, 263 | 2 |
| H1b | 16189, 16257, 16356 | 73, 263 | 1 |
| H1h2 | CRS | 263 | 1 |
| H2a | 16519 | 263 | 2 |
| H2a | 16218, 16362 | 263, 292 | 1 |
| H3 | CRS | 263 | 1 |
| H3 | 16519 | 152, 263 | 1 |
| H3 | 16093, 16362, 16519 | 263 | 1 |
| H3 | 16093, 16362 | 73, 263 | 1 |
| H3 | 16192, 16320, 16519 | 152, 263 | 1 |
| H3c2 | 16176 | 195, 263 | 1 |
| H3c2 | 16176 | 195, 242, 263 | 1 |
| H6a1 | 16192, 16362 | 239, 263 | 1 |
| H7 | 16519 | 263 | 1 |
| H7b | 16519 | 73, 263 | 1 |
| HV | 16209, 16223, 16278, 16362, 16519 | 73, 195, 263, 497 | 1 |
| J1b1a | 16222, 16261[a2] | 73, 242, 263, 295, 462, 489 | 2 |
| J1b1a1 | 16069, 16126, 16145, 16172, 16222, 16261 | 73, 152, 242, 263, 295, 462, 489 | 11 |
| J1b1a1 | 16069, 16126, 16145, 16172, 16222, 16261 | 73, 242, 263, 295, 462, 489 | 2 |
| J1b1a1 | 16069, 16126, 16145, 16189, 16172, 16222, 16261 | 73, 242, 263[a3] | 1 |
| J1c1 | 16069, 16126 | 64, 73, 185, 228, 263, 295, 462, 482, 489 | 1 |

| | | | |
|---|---|---|---|
| J1c1 | Did not amplify | 64, 73, 185, 228, 263, 295, 462, 482, 489 | 1 |
| J1c2e | 16069, 16126, 16366 | 73, 185, 188, 228, 263, 295, 462, 489 | 1 |
| J2a1a1 | 16069, 16126, 16145, 16231, 16261 | 73, 150, 152, 195, 215, 263, 295, 319, 489 | 3 |
| J2a1a1 | 16069,16126, 16145,16231, 16261 | 73, 150, 152, 195, 215, 263, 295[a4] | 1 |
| K1 | 16189, 16224, 16311 | 73, 263[b1] | 1 |
| K1a | 16224, 16311, 16519 | 73, 263, 497 | 1 |
| K1a | 15930, 16224, 16311, 16519 | 73, 263, 497 | 1 |
| K1a | 16093, 16224, 16311, 16519 | 73, 152, 263, 497 | 1 |
| K1c1 | 15928, 16311, 16519 | 73, 146, 152, 263 | 1 |
| R0a | 16126, 16189, 16362 | 64, 150, 263 | 1 |
| R1a | 16288, 16311, 16519 | 73, 150, 185, 189, 263, C295A | 1 |
| R2 | CRS | 73, 152, 195, 263 | 1 |
| T2b | 15928, 16126, 16294, 16296, 16304, 16344, 16519 | 73, 151, 152, 263 | 4 |
| T2b | 15928, 16126, 16294, 16296, 16304, 16344 | 73, 185, 228, 263 | 1 |
| T2c1 | 16126, 16292, 16294, 16296, 16519[b2] | 73, 146, 247, 263, 466 | 1 |
| Pre-T2f | 16126, 16189, 16294, 16296, 16304 | 73, 146, 263 | 1 |
| U2e1 | 16362, 16519 | 73, 152, 195, 217, 263, 340 | 1 |
| U2e1 | 16051, 16129C, 16189, 16362 | 73, 152, 217, 263, 340 | 1 |
| U5a | 16256, 16270 | 73, 146, 263 | 2 |
| U5b | 16192[b3] | 73, 150, 263 | 1 |
| U5b | 16189, 16270, 16311 | 73, 150, 263 | 1 |
| U5b | 16189, 16270 | Did not amplify | 1 |

| | | | |
|---|---|---|---|
| U5b1b1b | 16189, 16192, 16270, 16320 | 73, 150, 263 | 9 |
| U5b1b1b | 16189, 16192, 16270, 16320 | Did not amplify | 1 |
| U5b2b3a | 16224, 16270 | 73, 150, 263, 279 | 5 |
| U5b2b3a | 15905, 16224, 16270 | 73, 150, 263, 279 | 3 |
| U5b2b3a | 16270[a5] | 73, 150, 263, 279 | 1 |
| U6a1 | 16172, 16189, 16219, 16278 | 73, 263 | 1 |
| U6a1 | 16219, 16278[a6] | 73, 263 | 1 |
| U6b1a | 16163, 16172, 16219, 16311 | 73, 263 | 2 |
| U8b1 | 16172, 16189, 16234, 16311 | 73, 195, 263 | 2 |
| U8b1 | 16234, 16311, 16519[a7] | 73, 195, 263 | 1 |
| V | 16298 | 72, 263 | 1 |

[a]Transitions expected could not be confirmed due to low-quality sequences: (1) 16189; (2) 16069, 16126, 16145, 16172; (3) 295, 462, 489; (4) 319, 489; (5) 16224; (6) 16172, 16189; (7) 16172, 16189.

[b]Phylogenetically probable or expected transitions were confirmed not to be present: (1) 497; (2) 15928; (3) 16270.

**Table S2.** MtDNA Haplotypes Used for Population Parameters, Posterior Probability of Origins and PC Analyses

| Subhaplogroups | HVR-I | HVR-II | Frequency |
|---|---|---|---|
| H | 16260 | 263 | 1 |
| H1, H2a, H3, H7 | CRS | 263 | 7 |
| H1 | CRS | 152, 199, 263 | 1 |
| H1 | 16296 | 263 | 1 |
| H1b | 16189, 16356 | 152, 263 | 5* |
| H1b | 16189, 16356 | 73, 152, 263 | 1 |
| H1b | 16189, 16356 | 199, 263 | 2 |
| H1b | 16189, 16257, 16356 | 73, 263 | 1 |
| H2a | 16218, 16362 | 263, 292 | 1 |
| H3 | CRS | 152, 263 | 1 |
| H3 | 16093, 16362 | 263 | 1 |
| H3 | 16093, 16362 | 73, 263 | 1 |
| H3 | 16192, 16320 | 152, 263 | 1 |
| H3c2 | 16176 | 195, 263 | 1 |
| H3c2 | 16176 | 195, 242, 263 | 1 |
| H6a1 | 16192, 16362 | 239, 263 | 1 |
| H7b | CRS | 73, 263 | 1 |
| HV | 16209, 16223, 16278, 16362 | 73, 195, 263, 497 | 1 |
| J1b1a1 | 16069, 16126, 16145, 16172, 16222, 16261 | 73, 242, 263, 295, 462, 489 | 4* |
| J1b1a1 | 16069, 16126, 16145, 16172, 16222, 16261 | 73, 152, 242, 263, 295, 462, 489 | 11* |
| J1b1a1 | 16069, 16126, 16145, 16172, 16189, 16222, 16261 | 73, 242, 263, 295, 462, 489 | 1* |
| J1c1 | 16069, 16126 | 64, 73, 185, 228, 263, 295, 462, 482, 489 | 2* |
| J1c2e | 16069, 16126, 16366 | 73, 185, 188, 228, 263, 295, 462, 489 | 1 |

| | | | |
|---|---|---|---|
| J2a1a1 | 16069, 16126, 16145, 16231, 16261 | 73, 150, 152, 195, 215, 263, 295, 319, 489 | 4* |
| K1 | 16189, 16224, 16311 | 73, 263 | 1 |
| K1a | 16224, 16311 | 73, 263, 497 | 2 |
| K1a | 16093, 16224, 16311 | 73, 152, 263, 497 | 1 |
| K1c1 | 16311 | 73, 146, 152, 263 | 1 |
| R0a | 16126, 16189, 16362 | 64, 150, 263 | 1 |
| R1a | 16288, 16311 | 73, 150, 185, 189, 263, C295A | 1 |
| R2 | CRS | 73, 152, 195, 263 | 1 |
| T2b | 16126, 16294, 16296, 16304, 16344 | 73, 151, 152, 263 | 4 |
| T2b | 16126, 16294, 16296, 16304, 16344 | 73, 185, 228, 263 | 1 |
| T2c1 | 16126, 16292, 16294, 16296 | 73, 146, 247, 263, 466 | 1 |
| Pre-T2f | 16126, 16189, 16294, 16296, 16304 | 73, 146, 263 | 1 |
| U2e1 | 16362 | 73, 152, 195, 217, 263, 340 | 1 |
| U2e1 | 16051, 16129C, 16189, 16362 | 73, 152, 217, 263, 340 | 1 |
| U5a | 16256, 16270 | 73, 146, 263 | 2 |
| U5b | 16192 | 73, 150, 263 | 1 |
| U5b | 16189, 16270, 16311 | 73, 150, 263 | 1 |

| | | | |
|---|---|---|---|
| U5b | 16189, 16270 | 73, 150, 263 | 1* |
| U5b1b1b | 16189, 16192, 16270, 16320 | 73, 150, 263 | 10* |
| U5b2b3a | 16224, 16270 | 73, 150, 263, 279 | 9* |
| U6a1 | 16172, 16189, 16219, 16278 | 73, 263 | 2* |
| U6b1a | 16163, 16172, 16219, 16311 | 73, 263 | 2 |
| U8b1 | 16172, 16189, 16234, 16311 | 73, 195, 263 | 3* |
| V | 16298 | 72, 263 | 1 |

*Includes imputed sequences.

**Table S3.** Motifs and Frequency of MtDNA Clades in Puerto Rico

| Clade | Motifs | Frequency (%) |
|---|---|---|
| H | -*Alu*I 7025, -*Mse*I 14766 | 1 (0.9) |
| H1 | 3010 | 4 (3.9) |
| H1b | 3010-16189-16356 | 9 (8.9) |
| H1h2 | 3010-2851 | 1 (0.9) |
| H2a | 4769 | 3 (2.9) |
| H3 | 6776 | 5 (4.9) |
| H3c2 | 6776-16176-195 | 2 (1.9) |
| H6a1 | 239-3915-4727-16362 | 1 (0.9) |
| H7 | 4793 | 1 (0.9) |
| H7b | 4793-5348 | 1 (0.9) |
| HV | -*Mse*I 14766, 73 | 1 (0.9) |
| V | -NlaIII 4577, 72-16298 | 1 (0.9) |
| J1b1a | 16222-16261 | 2 (1.9) |
| J1b1a1 | 16069-16126-16145-16172-16222-16261 | 14 (13.8) |
| J1c1 | 64-73-185-228-263-295-462-482-489 | 2 (1.9) |
| J1c2e | 16069-16126-16366 | 1 (0.9) |
| J2a1a | 16069-16126-16145-16231-16261 | 4 (3.9) |
| T2b | 16126-16294-16296-16304-16344 | 5 (4.9) |
| T2c1 | 16126-16292-16294-16296-16519 | 1 (0.9) |
| Pre-T2f | 8281-8289 del, 16126-16189-16294-16296-16304 | 1(0.9) |
| U2e1 | 16051-16129C-16189-16362 | 2 (1.9) |
| U5a | 146-16256-16270 | 2 (1.9) |
| U5b | 150-16189-16270 | 3 (2.9) |
| U5b2b3a | 150-279-16224-16270 | 9 (8.9) |
| U5b1b1b | 150-16189-16192-16270-16320 | 10 (9.9) |
| U6a1 | 16172-16189-16219-16278 | 2 (1.9) |
| U6b1a | 16163-16172-16219-16311 | 2 (1.9) |
| U8b1 | 16172-16189-16234-16311 | 3 (2.9) |
| K1 | -*Hae*II 9052, -*Dde*I 10394 | 1 (0.9) |
| K1a | -*Hae*II 9052, -*Dde*I 10394, 497 | 3 (2.9) |
| K1c1 | -*Hae*II 9052, -*Dde*I 10394, 9093 | 1 (0.9) |
| R0a | 64-16126-16189-16362 | 1 (0.9) |
| R1a | 73-150-185-189-263-C295A-16288-16311 | 1 (0.9) |
| R2 | 73-152-195-263 | 1 (0.9) |
| N, total samples | | 101 |

**Table S4.** Published HVR-I and –II Sequence Data Used for Fisher's Exact Tests and to Construct Phylogenetic Networks

| Ethnicity | Geographic origin | Total # of samples | Reference |
|---|---|---|---|
| Puerto Ricans | Puerto Rico | 101 | Present study |
| Austrians | Austria | 273 | Brandstätter et al. (2007) |
| Bedouins | Israel | 58 | Behar et al. (2008) |
| Bosnians | Bosnia-Herzegovina | 144 | Malyarchuk et al. (2003) |
| Cherkess | Israel | 8 | Behar et al. (2008) |
| Cretans | Eastern Crete | 283 | Martínez et al. (2008) |
| Czech | Czech Republic | 179 | Malyarchuk et al. (2006) |
| Druze | Israel | 77 | Behar et al. (2008) |
| Dubaians | Dubai | 249 | Alshamali et al. (2008) |
| Finns | Finland | 200 | Hedman et al. (2007) |
| French | France | 210 | Dubut et al. (2004) |
| Northeast Germans | Mecklenburg | 213 | Tetzlaff et al. 2007 |
| North Italians | North Central Italy | 395 | Turchi et al. (2008) |
| Malian | Mali | 124 | González et al.(2006) |
| Maure | Mauritania | 64 | González et al.(2006) |
| Non-Ashkenazi Jews | Dispersed (see ref. Table 1) | 1142 | Behar et al. (2008) |
| Palestinians | Israel | 110 | Behar et al. (2008) |
| Poles | Poland (Pomerania-Kujawy) | 436 | Malyarchuk et al. (2002) |
| Russians | European Russia | 201 | Malyarchuk et al. (2002) |
| Saudi Arabs | Saudi Arabia | 546 | Abu-Amero et al. (2008) |
| Slovenians | Slovenia | 104 | Malyarchuk et al. (2003) |
| Spaniards | Spain | 312 | Álvarez et al. (2007) |
| Tuscans | Italy | 322 | Achilli et al. (2007) |

**Table S5.** Diagnostic Motifs for Subhaplogroup Characterization Using HVR-I and –II Sequences

| Clade | Coding region state | Control region signature |
|---|---|---|
| H1b | -7025 *Alu*I | 16189-16356 |
| H3c2 | -7025 *Alu*I | 16176-195 |
| H6 | -7025 *Alu*I | 16362-239 |
| H(xH1b, H3c2, H6) | -7025 *Alu*I | --- |
| HV (includes V) | -14766 *Mse*I | --- |
| R0a | +7025 *Alu*I, -14766 *Mse*I | 16126-16362 |
| R1a | +7025 *Alu*I, -14766 *Mse*I | 16311-C295A |
| R2 | +7025 *Alu*I, -14766 *Mse*I +4216 *Nla*III | 152 *(lack of 16126)* |
| R | +7025 *Alu*I, -14766 *Mse*I | --- |
| J1b1a1 | +4216 *Nla*III | 16069-16126-16145-16172-16261-242-295 |
| J1c | +4216 *Nla*III | 16069-16126-185-228-295 |
| J1c2 | +4216 *Nla*III | 16069-16126-185-188-228-295 |
| J2a1a1 | +4216 *Nla*III | 16069-16126-16145-16231-16261-150-152-195-215-295 |
| J(xJ1b1a1, J1c, J2a1a1) | +4216 *Nla*III | 16069-16126-295 |
| T2b | +4216 *Nla*III | 16126-16294-16296-16304 |
| T2c1 | +4216 *Nla*III | 16126-16292-16294-16296 |
| Pre-T2f | +4216 *Nla*III | 16126-16189-16294-16296 |
| T(xT2b, T2c1, Pre-T2f) | +4216 *Nla*III | 16126-16294 |
| U2e1'2'3 | +12308 *Hin*fI[a] | 16051-G16129C-16189-16362-152-217 |
| U5a | +12308 *Hin*fI[a] | 16256-16270 |
| U5b | +12308 *Hin*fI[a] | 16270-150 |
| U5b1b1b | +12308 *Hin*fI[a] | 16189-16192-16270-16320-150 |
| U5b2b3a | +12308 *Hin*fI[a] | 16224-16279-150-279 |
| U6a1 | +12308 *Hin*fI[a] | 16172-16219-16278 |
| U6b1a | +12308 *Hin*fI[a] | 16163-16172-16219-16311 |
| U8b1 | +12308 *Hin*fI[a] | 16189-16234-195 |
| U(xU2e1'2'3, U5a ,U5b, U6a1, U6b1a, U8b1) | +12308 *Hin*fI[a] | --- |
| K1c | +12308 *Hin*fI[a] | 16224-16311-146-152 |
| K2 | +12308 *Hin*fI[a] | 16224-16311-146 |
| K3 | +12308 *Hin*fI[a] | 16093-16148-16153-16224-16311-150-195-235 |
| K(xK1c, K2 ,K3) | +12308 *Hin*fI[a] | 16224-16311 |
| Others(E-G,I,M-Q,S,W-Z)[b] | | |

[a]Using a degenerate primer.

[b]For coding and control region signatures see (van Oven and Kayser, 2009).

**Table S6.** Populations with Published HVR-I Sequence Data Used for Posterior Probability of Origin and PC Analyses

| Ethnicity | Geographic origin | # of samples | Reference |
|---|---|---|---|
| Puerto Ricans | Puerto Rico | 101 | Present study |
| Armenians | Armenia | 190 | Richards et al. (2000) |
| Austrians | Austria | 272 | Brandstätter et al. (2007) |
| Azeri | Azerbaijan | 46 | Richards et al. (2000) |
| Basques | Spain | 155 | Richards et al. (2000) |
| Bedouins | Israel and Saudi Arabia | 73 | Behar et al. (2008); Richards et al. (2000) |
| Berbers | Morocco | 53 | Rando et al. (1998) |
| | Tunisia | 26 | Cherni et al. (2005) |
| Bosnians | Bosnia-Herzegovina | 142 | Malyarchuk et al. (2003) |
| Canarians | Canary Islands | 278 | Rando et al. (1999) |
| Cretans | Eastern Crete | 283 | Martínez et al. (2008) |
| Czech | Czech Republic | 177 | Malyarchuk et al. (2006) |
| Druze | Israel | 120 | Behar et al. (2008); Richards et al. (2000) |
| Dubaians | Dubai | 193 | Alshamali et al. (2008) |
| Egyptians | Egypt | 54 | Richards et al. (2000) |
| Europeans | Alps | 215 | Richards et al. (2000) |
| | Mediterranean (Central) | 296 | Richards et al. (2000) |
| | Mediterranean (East) | 165 | Richards et al. (2000) |
| | Mediterranean (West) | 209 | Richards et al. (2000) |
| | North Central | 328 | Richards et al. (2000) |
| | North East | 398 | Richards et al. (2000) |
| | North West | 453 | Richards et al. (2000) |
| | Scandinavia | 312 | Richards et al. (2000) |
| | Southeast | 229 | Richards et al. (2000) |
| Finns | Finland | 200 | Hedman et al. (2007) |
| French | France | 210 | Dubut et al. (2004) |
| Fulani, Wolofs and Serers | Senegal and Cameroon | 25 | Rando et al. (1998); Coia et al. (2005); Cerny et al. 2006) |
| Georgians | Georgia | 45 | Comas et al. (2000) |
| Hungarians | Hungary | 73 | Bogacsi-Szabo et al. (2005) |
| Iranians | Iran | 12 | Richards et al. (2000) |
| Iraqis | Iraq | 105 | Richards et al. (2000) |
| Jews | Algeria | 20 | Behar et al. (2008) |

| | | | |
|---|---|---|---|
| | Azerbaijan | 58 | Behar et al. (2008) |
| | India (Cochin) | 44 | Behar et al. (2008) |
| | India (Mumbai) | 34 | Behar et al. (2008) |
| | Iran | 75 | Behar et al. (2008) |
| | Iraq | 135 | Behar et al. (2008) |
| | Georgia | 74 | Behar et al. (2008) |
| | Libya | 80 | Behar et al. (2008) |
| | Morocco | 146 | Behar et al. (2008) |
| | Near and Middle East (Afghan, Kurdish (Iraq), Uzbek, Syrian) | 34 | Behar et al. (2008) |
| | Spanish Exiles | 213 | Behar et al. (2008) |
| | Tunisia | 36 | Behar et al. (2008) |
| | Yemen | 98 | Behar et al. (2008) |
| Kurds | Kurdistan | 78 | Comas et al. 2000; Richards et al. 2000) |
| Maure | Mauritania | 47 | Rando et al. (1998); González et al.(2006) |
| Moroccans | Morocco | 61 | Rando et al. (1998); Brakez et al. (2001) |
| North Caucasians | North Caucasus | 191 | Richards et al. (2000) |
| Northeast Germans | Mecklenburg | 212 | Telzlaff et al. 2007 |
| North Italians | North Central Italy | 388 | Turchi et al. (2008) |
| Nubians | Nile River | 34 | Richards et al. (2000) |
| Palestinians | Israel | 198 | Behar et al. (2008); Richards et al. (2000) |
| Poles | Poland (Pomerania-Kujawy) | 430 | Malyarchuk et al. (2002) |
| Portuguese | Portugal | 273 | González et al. (2003) |
| Russians | European Russia | 200 | Malyarchuk et al. (2002) |
| Sardinians | Sardinia | 42 | Fraumene et al. (2003) |
| Slovenians | Slovenia | 104 | Malyarchuk et al. (2003) |
| Spaniards | Spain | 301 | Álvarez et al. (2007) |
| Syrians | Syria | 62 | Richards et al. (2000) |
| Tunisians | Tunisia | 46 | Cherni et al. (2005) |
| Turks | Turkey | 200 | Richards et al. (2000) |
| West Saharans | West Sahara | 14 | Rando et al. (1998) |

**Table S7.** Diagnostic Motifs for Subhaplogroup Characterization Using HVR-I Sequences

| Clade | Coding region state | HVR-I signature |
|---|---|---|
| H1b | -7025 *Alu*I | 16189-16356 |
| H3c2 | -7025 *Alu*I | 16176 |
| H6 | -7025 *Alu*I | 16362 |
| H(xH1b, H3c2, H6) | -7025 *Alu*I | --- |
| HV (includes V) | -14766 *Mse*I | --- |
| R0a | +7025 *Alu*I<br>-14766 *Mse*I | 16126-16362 |
| R(xR0a) | +7025 *Alu*I<br>-14766 *Mse*I | --- |
| J1b1a1 | +4216 *Nla*III | 16069-16126-16145-16172-16261 |
| J2a1a1 | +4216 *Nla*III | 16069-16126-16145-16231-16261 |
| J(xJ1b1a1, J2a1a1) | +4216 *Nla*III | 16069-16126 |
| T2b | +4216 *Nla*III | 16126-16294-16296-16304 |
| T2c1 | +4216 *Nla*III | 16126-16292-16294-16296 |
| Pre-T2f | +4216 *Nla*III | 16126-16189-16294-16296 |
| T(xT2b, T2c1, Pre-T2f) | +4216 *Nla*III | 16126-16294 |
| U2e1'2'3 | +12308 *Hin*fI[a] | 16051-G16129C-16189-16362-152-217 |
| U5(xU5a, U5b1b1b, U5b2b3a) | +12308 *Hin*fI[a] | 16270 |
| U5a | +12308 *Hin*fI[a] | 16256-16270 |
| U5b1b1b | +12308 *Hin*fI[a] | 16189-16192-16270-16320 |
| U5b2b3a | +12308 *Hin*fI[a] | 16224-16279 |
| U6a1 | +12308 *Hin*fI[a] | 16172-16219-16278 |
| U6b1a | +12308 *Hin*fI[a] | 16163-16172-16219-16311 |
| U8b1 | +12308 *Hin*fI[a] | 16189-16234 |
| U(xU2e1'2'3, U5 U6a1, U6b1a, U8b1) | +12308 *Hin*fI[a] | --- |
| K | +12308 *Hin*fI[a] | 16224-16311 |
| Others(E-G,I,M-Q,S,W-Z)[b] | | |

[a]Using a degenerate primer.

[b]For coding and control region signatures see (van Oven and Kayser, 2009).

**Table S8.** P-Values from Fisher's Exact Test Pairwise Comparisons of Subhaplogroup Frequencies

| Comparison: Puerto Rico vs X | H1b | H3c2 | H6 | H | HV | R0a | R1a | R2 | R | J1b1a1 | J1c | J1c2 | J2a1a1 | J | T2b | T2c1 | Pre.T2f | T | U2e | U5a | U5b | U5b1b1b | U5b2b3a | U6a1 | U6b1a | U8b1 | U | K1c | K2 | K3 | K | Others |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Spaniard | 0.01 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.003 | 0.001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.171 |
| North Italian | 0 | 1 | 1 | 0.003 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.086 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Cretan | 1 | 1 | 1 | 0.04 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0.002 | 1 | 1 | 1 | 0.052 | 1 | 1 | 1 | 1 | 0.002 |
| French | 0.054 | 1 | 1 | 0.066 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.002 | 0.007 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Bosnian | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.051 | 0.13 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Austrian | 1 | 1 | 1 | 0.009 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.721 | 1 | 1 | 1 | 0.001 | 0.003 | 1 | 1 | 1 | 0.244 | 1 | 1 | 1 | 1 | 1 |
| Slovenian | 0.589 | 1 | 1 | 0.007 | 1 | 1 | 1 | 1 | 1 | 0.003 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.275 | 0.589 | 1 | 1 | 1 | 0.688 | 1 | 1 | 1 | 1 | 1 |
| Czech | 1 | 1 | 1 | 0.032 | 1 | 1 | 1 | 1 | 1 | 0.001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.013 | 0.038 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.62 |
| Northeast German | 0.093 | 1 | 1 | 0.007 | 1 | 1 | 1 | 1 | 1 | 0.001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.004 | 0.013 | 1 | 1 | 1 | 0.882 | 1 | 1 | 1 | 1 | 1 |
| Pole | 1 | 1 | 1 | 0.001 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.293 |
| Finn | 0.518 | 1 | 1 | 0.018 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.043 | 0.006 | 0.018 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.436 |
| Russian | 1 | 1 | 1 | 0.074 | 1 | 1 | 1 | 1 | 1 | 0.006 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.006 | 0.018 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.254 |
| Dubaian | 0.023 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.001 | 1 | 1 | 1 | 0.003 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.007 | 0.023 | 1 | 1 | 1 | 0.001 | 1 | 1 | 1 | 1 | 0 |
| Palestinian | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.069 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.286 | 1 | 1 | 1 | 0.677 | 1 | 1 | 1 | 1 | 0.61 | 1 | 1 | 1 | 1 | 0 |
| Bedouin | 1 | 1 | 1 | 1 | 1 | 0.25 | 1 | 1 | 1 | 0.642 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Druze | 1 | 1 | 1 | 0.361 | 1 | 1 | 1 | 1 | 1 | 0.04 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.001 | 1 |
| Ethiopian Jew | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Algerian Jew | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Libyan Jew | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.031 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Moroccan Jew | 0.112 | 1 | 1 | 0.081 | 1 | 1 | 1 | 1 | 1 | 0.008 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.043 | 0.112 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.015 |
| Tunisian Jew | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.005 |
| Azeri Jew | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.259 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Cochin Jew India | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.264 | 1 | 1 | 1 | 1 | 0 |
| Mumbaikaran Jew | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Irani Jew | 1 | 1 | 0.971 | 1 | 1 | 1 | 1 | 1 | 1 | 0.04 | 1 | 1 | 1 | 0.06 | 1 | 1 | 1 | 0.378 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.023 | 1 | 1 | 1 | 1 | 0.151 |
| Iraqi Jew | 0.169 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.003 | 1 | 1 | 1 | 0.639 | 1 | 0.002 | 1 | 1 | 1 | 1 | 1 | 0.069 | 0.169 | 1 | 1 | 1 | 0.012 | 1 | 1 | 1 | 1 | 0 |
| Near-Middle Eastern Jew | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.349 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.001 |
| Yemeni Jew | 1 | 1 | 1 | 1 | 0.004 | 0.013 | 1 | 0.121 | 1 | 0.007 | 1 | 1 | 1 | 0.001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.677 | 1 | 1 | 1 | 1 | 0.286 | 1 | 1 | 1 | 1 | 1 |
| Spanish Exile Jew | 0.012 | 1 | 1 | 0.133 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.004 | 0.012 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.027 |
| | 4 | 0 | 0 | 10 | 1 | 1 | 0 | 0 | 1 | 20 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 13 | 12 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 1 | 14 |

**Table S9.** Power Estimates for Multiple Fisher Pairwise Comparisons of Subhaplogroup Frequencies

| Comparison: PR vs X | H1b | H3c2 | H6 | H | HV | R0a | R1a | R2 | R | J1b1a1 | J1c | J1c2 | J2a1a1 | J | T2b | T2c1 | Pre.T2f | T | U2e | U5a | U5b | U5b1b1b | U5b2b3a | U6a1 | U6b1a | U8b1 | U | K1c | K2 | K3 | K | Others |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Spaniard** | 0.9865 | 0.2157 | 0.0131 | 0.9991 | 0.3554 | 0.0777 | 0.0806 | 0.0429 | 0 | 0.9999 | 0.0411 | 0.0186 | 0.5228 | 0.0188 | 0.1445 | 0.013 | 0.0799 | 0.7546 | 0.0615 | 0.0428 | 0.1624 | 0.9928 | 0.9956 | 0.065 | 0.3353 | 0.3148 | 0.9682 | 0.0194 | 1.00E-04 | 0 | 0.0497 | 0.9999 |
| **North Italian** | 0.9988 | 0.5959 | 0.0076 | 0.9978 | 0.4206 | 0.0326 | 0.2752 | 0.2635 | 0 | 1 | 0.0503 | 0.0227 | 0.7465 | 0.2703 | 0.2836 | 0.0105 | 0.0709 | 1 | 0.3499 | 0.0687 | 0.1542 | 0.9999 | 0.9986 | 0.3573 | 0.6008 | 0.3326 | 0.8399 | 0.0304 | 0.1652 | 0 | 0.094 | 0.9944 |
| **Cretan** | 0.8313 | 0.3239 | 0.0806 | 0.985 | 0.3148 | 0.0081 | 0.0791 | 0.0201 | 0 | 1 | 0.0189 | 0.0237 | 0.6088 | 0.8234 | 0.1265 | 0.0106 | 0.0815 | 0.9833 | 0.3307 | 0.0595 | 0.2161 | 0.9978 | 0.9962 | 0.3295 | 0.3283 | 0.5878 | 0.9999 | 0.0836 | 0 | 0 | 0.0314 | 1 |
| **French** | 0.9728 | 0.0935 | 0.0392 | 0.9741 | 0.1074 | 0.0167 | 0.0804 | 0.0833 | 0 | 1 | 0.3302 | 0.0846 | 0.4001 | 0.4205 | 0.0467 | 0.0884 | 0.0821 | 0.9441 | 0.0167 | 0.2593 | 0.0921 | 0.9975 | 0.9959 | 0.0985 | 0.3377 | 0.5851 | 0.6601 | 0.0144 | 0.0379 | 0 | 0.0777 | 1 |
| **Bosnian** | 0.4921 | 0.1499 | 0.0203 | 0.9993 | 0.41 | 0.0072 | 0.0169 | 0.0206 | 0 | 0.9999 | 0.0678 | 0.0318 | 0.3169 | 0 | 0.2875 | 0.0168 | 0.0186 | 0.3815 | 0.1414 | 0.4078 | 0.0827 | 0.9921 | 0.9827 | 0.1444 | 0.1492 | 0.3663 | 0.9663 | 0.0205 | 0.0139 | 0 | 0.0538 | 0.9671 |
| **Austrian** | 0.8314 | 0.333 | 0.0626 | 0.9952 | 0.0219 | 0.081 | 0.0835 | 0.0834 | 0 | 1 | 0.0869 | 0.0101 | 0.3952 | 0.0405 | 0.0401 | 0.0857 | 0.0203 | 0.995 | 0.0231 | 0.2636 | 0.0675 | 0.9977 | 0.9954 | 0.3295 | 0.3292 | 0.5794 | 0.9393 | 0.0123 | 2.00E-04 | 0 | 0.0401 | 0.6594 |
| **Slovenian** | 0.9515 | 0.0544 | 0.0352 | 0.9918 | 0.2642 | 0.0043 | 0.0034 | 0.0031 | 0 | 1 | 0.0187 | 0.0347 | 0.3728 | 0.3855 | 0.0461 | 0.0061 | 0.003 | 0.0163 | 0.0175 | 0.3719 | 0.0806 | 0.9756 | 0.9536 | 0.0532 | 0.0546 | 0.1894 | 0.9316 | 0.0016 | 0 | 0 | 0.0425 | 0.8142 |
| **Czech** | 0.2734 | 0.3311 | 0.0067 | 0.9813 | 0.0515 | 0.0707 | 0.0803 | 0.0821 | 0 | 0.9976 | 0.0266 | 0.0105 | 0.2946 | 0.2821 | 0.0504 | 0.0304 | 0.0314 | 0.9061 | 0.1379 | 0.1404 | 0.0202 | 0.9981 | 0.9949 | 0.333 | 0.3248 | 0.5875 | 0.6667 | 0.0111 | 0 | 0 | 0.1767 | 0.9686 |
| **Northeast German** | 0.9708 | 0.3342 | 0.0303 | 0.9938 | 0.0375 | 0.0875 | 0.0799 | 0.0865 | 0 | 0.9974 | 0.1911 | 0.0186 | 0.1826 | 0 | 0.0337 | 0.0341 | 0.0057 | 0.6673 | 0.1786 | 0.1786 | 0.0317 | 0.9983 | 0.9952 | 0.3283 | 0.3331 | 0.5935 | 0.9583 | 0.0057 | 7.00E-04 | 0 | 0.0501 | 0.9365 |
| **Pole** | 0.5266 | 0.6068 | 0.006 | 0.9988 | 0.2899 | 0.2683 | 0.2717 | 0.2803 | 0 | 0.9987 | 0.0367 | 0.0043 | 0.6227 | 4.00E-04 | 0.0774 | 0.1546 | 0.0773 | 0.4601 | 0.1637 | 0.1897 | 0.0255 | 0.9996 | 0.9992 | 0.6053 | 0.6072 | 0.8103 | 0.9956 | 0.0522 | 0 | 0 | 0.1222 | 0.9999 |
| **Finn** | 0.908 | 0.32 | 0.036 | 0.9853 | 0.1071 | 0.082 | 0.0797 | 0.0359 | 0.0184 | 1 | 0.1837 | 0.0099 | 0.7754 | 0.0604 | 0.4044 | 0.0307 | 0.0169 | 0.1511 | 0.1786 | 0.3319 | 0.0858 | 0.9981 | 0.9946 | 0.3338 | 0.3283 | 0.5887 | 0.8383 | 0.0096 | 0.1477 | 0 | 0.2544 | 0.9889 |
| **Russian** | 0.6498 | 0.1803 | 0.0604 | 0.9762 | 0.4743 | 0.0391 | 0.0883 | 0.0385 | 0 | 0.9912 | 0.0321 | 0.0263 | 0.7683 | 0.0004 | 0.0409 | 0.0183 | 0.0804 | 0.4 | 0.0533 | 0.4749 | 0.0345 | 0.9985 | 0.9949 | 0.3211 | 0.3267 | 0.5864 | 0.8526 | 0.0278 | 0 | 0 | 0.5759 | 0.9994 |
| **Dubaian** | 0.9959 | 0.333 | 0.0182 | 0.7161 | 0.6868 | 0.6161 | 0.081 | 0.0261 | 0.0055 | 0.9988 | 0.3308 | 0.0788 | 0.7781 | 1 | 0.7337 | 0.0215 | 0.0823 | 0.0683 | 0.0207 | 0.3335 | 0.5903 | 0.9974 | 0.9957 | 0.3243 | 0.3289 | 0.5814 | 1 | 0.0112 | 0.0041 | 0 | 0.1471 | 1 |
| **Palestinian** | 0.9007 | 0.0155 | 0.0013 | 0.0627 | 0.7623 | 0.6878 | 1.00E-04 | 7.00E-04 | 0 | 0.9848 | 0.0145 | 2.00E-04 | 0.2198 | 0.5691 | 0.2063 | 8.00E-04 | 4.00E-04 | 0.9743 | 0.0151 | 0.0851 | 0.0193 | 0.9434 | 0.8977 | 0.0082 | 0.0141 | 0.0346 | 0.9527 | 8.00E-04 | 0.0479 | 0 | 0.1399 | 1 |
| **Bedouin** | 0.5472 | 2.00E-04 | 0 | 0.0562 | 0.2204 | 0.9426 | 0 | 0 | 0 | 0.9843 | 0.0163 | 0 | 0.0106 | 0.0718 | 0.0266 | 0.3745 | 0 | 0.5834 | 2.00E-04 | 4.00E-04 | 0.004 | 0.678 | 0.5536 | 1.00E-04 | 2.00E-04 | 0.0022 | 0.7723 | 0 | 0.0751 | 0 | 0.0252 | 1 |
| **Druze** | 0.062 | 0.0189 | 0.1612 | 0.9197 | 0.0275 | 8.00E-04 | 5.00E-04 | 6.00E-04 | 0.0174 | 0.9993 | 0.0103 | 0.0011 | 0.2196 | 0.1395 | 0.3905 | 9.00E-04 | 4.00E-04 | 0.0172 | 0.0172 | 0.0164 | 0.0813 | 0.9432 | 0.8979 | 0.0168 | 0.0175 | 0.0817 | 0.7517 | 9.00E-04 | 0.0159 | 0 | 0.9982 | 0.3496 |
| **Ethiopian Jew** | 0.1152 | 0 | 0 | 0.8303 | 0 | 0.598 | 0 | 0 | 0 | 0.828 | 0 | 0 | 2.00E-04 | 0 | 0.0021 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1935 | 0.113 | 0.0584 | 0 | 0 | 0.6028 | 0 | 0 | 0 | 2.00E-04 | 1 |
| **Algerian Jew** | 0.0035 | 0 | 0.1314 | 0.754 | 0.0771 | 0 | 0 | 0 | 0 | 0.3254 | 0.0834 | 0 | 0 | 0 | 0 | 0 | 0 | 0.2667 | 0 | 0 | 0.0488 | 0.0096 | 0.0035 | 0 | 0 | 0.0459 | 0 | 0.6683 | 0 | 0 | 0 | 0.6054 |
| **Libyan Jew** | 0.8944 | 0.0142 | 0.0011 | 0.7856 | 0.0818 | 0.3115 | 9.00E-04 | 9.00E-04 | 0 | 0.9995 | 0.0079 | 8.00E-04 | 0.0411 | 0.0152 | 0.3966 | 4.00E-04 | 0.0054 | 0.0159 | 0.0177 | 0.0087 | 0.0262 | 0.9381 | 0.8931 | 0.0184 | 0.0157 | 0.0876 | 0.1391 | 2.00E-04 | 0 | 0 | 0.0866 | 1 |
| **Moroccan Jew** | 0.982 | 0.1488 | 0.1743 | 0.9688 | 0.7421 | 0.0074 | 0.0175 | 0.0072 | 0 | 0.9913 | 0.0336 | 0.0203 | 0.3655 | 0.3936 | 0.5192 | 0.0068 | 0.0183 | 0 | 0.1442 | 0.0359 | 0.1999 | 0.9923 | 0.9833 | 0.0372 | 0.1437 | 0.3532 | 0.2255 | 0.3838 | 0.0053 | 0 | 0.2657 | 0.9998 |
| **Tunisian Jew** | 0.1873 | 0 | 0.1941 | 0.6831 | 0.3022 | 0.4177 | 0 | 0 | 0 | 0.8941 | 0 | 0 | 8.00E-04 | 0 | 0.0046 | 0 | 0 | 0 | 0 | 0 | 0.0111 | 0.3069 | 0.1808 | 0.0212 | 0 | 0 | 0.5842 | 0 | 0 | 0 | 0.0047 | 0.9927 |
| **Azeri Jew** | 0.3803 | 0.0042 | 1.00E-04 | 0.0728 | 0.0064 | 2.00E-04 | 0 | 1.00E-04 | 0 | 0.9972 | 0.0057 | 0 | 0.1177 | 1 | 0.2419 | 0.0273 | 1.00E-04 | 0.7753 | 0.0049 | 0.0041 | 0.0334 | 0.8832 | 0.8134 | 0.0038 | 0.0053 | 0.0332 | 0.8851 | 0.0287 | 0 | 0 | 0.1091 | 0.7719 |
| **Cochin Jew India** | 0.4074 | 1.00E-04 | 0 | 0.9292 | 1.00E-04 | 0 | 0 | 0.0352 | 0.9996 | 0.9666 | 0 | 0 | 0.0068 | 0 | 0.0323 | 0 | 0 | 0 | 0 | 0 | 0.0011 | 0.5435 | 0.4153 | 1.00E-04 | 0 | 7.00E-04 | 0.9468 | 0 | 0 | 0 | 0.0059 | 1 |
| **Mumbaikaran Jew** | 0.1897 | 0 | 0 | 0.1178 | 0 | 0 | 0 | 0 | 0.5732 | 0.8878 | 0 | 0 | 0.0013 | 0 | 0.0044 | 0 | 0 | 0 | 0 | 0 | 0 | 0.3009 | 0.189 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7.00E-04 | 1 |
| **Irani Jew** | 0.8944 | 0.0177 | 0.0835 | 0.1298 | 0.0164 | 0.0071 | 2.00E-04 | 3.00E-04 | 0 | 0.9991 | 0.0763 | 4.00E-04 | 0.2245 | 0.9916 | 0.1786 | 0.5974 | 5.00E-04 | 0.9628 | 0.0085 | 0.0162 | 0.0783 | 0.9408 | 0.8966 | 0.0183 | 0.0157 | 0.0884 | 0.9977 | 2.00E-04 | 0 | 0 | 0.2166 | 0.9864 |
| **Iraqi Jew** | 0.9837 | 0.1468 | 0.0214 | 0.1146 | 0.121 | 0.0192 | 0.0179 | 0.0299 | 0 | 0.9986 | 0.5252 | 0.0169 | 0.0409 | 0.9859 | 0.7515 | 0.9992 | 0.0213 | 0.2183 | 0.0622 | 0.1509 | 0.3613 | 0.9931 | 0.9816 | 0.1419 | 0.1481 | 0.0799 | 0.9998 | 0.0221 | 8.00E-04 | 0 | 0.1687 | 1 |
| **Near-Middle Eastern Jew** | 0.1904 | 0 | 0 | 0.1646 | 0 | 0 | 0 | 0 | 0 | 0.8948 | 0.1808 | 0 | 7.00E-04 | 0.8929 | 0.0046 | 0.2631 | 0 | 0.3192 | 0 | 0 | 0 | 0.3002 | 0.1833 | 0 | 0 | 1.00E-04 | 0.3252 | 0 | 0 | 0 | 0.3442 | 0.9976 |
| **Yemeni Jew** | 0.8904 | 0.0138 | 8.00E-04 | 0.7715 | 0.9976 | 0.9957 | 9.00E-04 | 0.9769 | 0.0031 | 0.9996 | 0.0166 | 2.00E-04 | 0.2282 | 0.9999 | 0.3991 | 0.0017 | 4.00E-04 | 0.0032 | 0.0147 | 0.0148 | 0.0819 | 0.9479 | 0.8949 | 0.0179 | 0.0151 | 0.0871 | 0.9739 | 7.00E-04 | 0.0456 | 0 | 0.0282 | 0.5713 |
| **Spanish Exile Jew** | 0.9959 | 0.3351 | 0.0308 | 0.9666 | 0.0843 | 0.0191 | 0.0745 | 0.0785 | 0 | 0.9998 | 0.1066 | 0.0378 | 0.1268 | 0.9347 | 0.2999 | 0.0091 | 0.0184 | 0.9329 | 0.3266 | 0.0584 | 0.2504 | 0.9972 | 0.9949 | 0.0568 | 0.3261 | 0.1491 | 0.7622 | 0.0343 | 0 | 0 | 0.2933 | 0.9999 |

**Table S10.** Unique Coding Region Mutations in the Two Main J1b1a1 Lineages in Puerto Rico

| | Position | Gene | Mutation type |
|---|---|---|---|
| J1b1a1 with 152 (*n*=11) | 13943 | ND5 | ACA->ATA (Thr->Met) |
| J1b1a1 without 152 (n=4) | 3324 | ND1 | CTC->CTT (Leu->Leu) |
| | 9438 | COX3 | GGC->AGC (Gly->Ser) |
| | 14560 | ND6 | GTC->GTT (Val->Val) |
| | 15740 | CYTB | CTC->TTC (Leu->Phe) |

**Table S11.** HVR-I-Based Sample Categorization Used for Posterior Probability of Origin and PCA Analyses

| Subhaplogroup | H1b | H3c2 | H6 | H(xH1b,H3c2,H6) | HV (includesV) | R0a | R(xR0a, R1a,R2) | J1b1a1 | J2a1a1 | J(xJ1b1a1, J1c,J1c2,J2a1a1) | T2b | T2c1 | Pre-T2f | T(xT2b,T2c1,Pre-T2f) | U2e1'2'3 | U5a | U5(xU5a) | U5b1b1b | U5b2b3a | U6a1 | U6b1a | U8b1 | U(xU2e1'2'3,U5a,U5b,U6a1, U6b1a,U8b1) | K(xK2,K3) | Others(E-G,I,M-Q,S,W-Z) | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Puerto Rican | 9 | 2 | 1 | 16 | 2 | 1 | 2 | 16 | 4 | 3 | 5 | 1 | 1 | 0 | 2 | 2 | 3 | 10 | 9 | 2 | 2 | 3 | 0 | 5 | 0 | 101 |
| Middle Eastern | 5 | 1 | 20 | 169 | 60 | 57 | 15 | 3 | 4 | 84 | 9 | 10 | 3 | 59 | 8 | 6 | 4 | 0 | 0 | 8 | 0 | 1 | 97 | 73 | 155 | 851 |
| Turk | 0 | 0 | 5 | 53 | 14 | 2 | 3 | 4 | 0 | 18 | 2 | 1 | 1 | 22 | 3 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 36 | 12 | 23 | 202 |
| Caucasian | 2 | 0 | 8 | 124 | 30 | 1 | 12 | 2 | 2 | 32 | 8 | 3 | 3 | 47 | 5 | 26 | 3 | 0 | 0 | 0 | 0 | 5 | 68 | 28 | 63 | 472 |
| East Mediterranean | 11 | 0 | 3 | 235 | 36 | 10 | 2 | 0 | 2 | 53 | 13 | 4 | 0 | 32 | 1 | 24 | 15 | 0 | 0 | 0 | 0 | 0 | 52 | 27 | 70 | 590 |
| South Eastern European | 2 | 0 | 8 | 155 | 29 | 4 | 1 | 0 | 1 | 33 | 10 | 2 | 1 | 26 | 2 | 24 | 15 | 0 | 0 | 0 | 0 | 0 | 35 | 22 | 35 | 407 |
| Central Mediterranean | 2 | 2 | 12 | 289 | 50 | 3 | 2 | 1 | 2 | 56 | 18 | 11 | 3 | 58 | 3 | 20 | 25 | 0 | 0 | 1 | 0 | 4 | 49 | 61 | 54 | 726 |
| Alpine | 8 | 1 | 8 | 203 | 16 | 1 | 0 | 0 | 11 | 37 | 20 | 0 | 4 | 30 | 8 | 28 | 12 | 0 | 0 | 0 | 0 | 0 | 43 | 35 | 22 | 487 |
| North Central European | 29 | 2 | 27 | 470 | 59 | 0 | 3 | 12 | 17 | 77 | 58 | 7 | 7 | 60 | 6 | 58 | 37 | 0 | 0 | 0 | 0 | 0 | 73 | 56 | 89 | 1147 |
| West Mediterranean | 4 | 3 | 5 | 378 | 41 | 3 | 1 | 4 | 2 | 50 | 24 | 6 | 0 | 38 | 5 | 18 | 23 | 1 | 0 | 11 | 2 | 2 | 60 | 46 | 56 | 783 |
| North Western European | 8 | 4 | 14 | 283 | 39 | 3 | 0 | 7 | 6 | 61 | 23 | 0 | 1 | 34 | 7 | 29 | 20 | 0 | 0 | 2 | 0 | 1 | 23 | 59 | 39 | 661 |
| Scandinavian | 1 | 0 | 7 | 155 | 13 | 0 | 0 | 5 | 3 | 19 | 14 | 1 | 0 | 11 | 0 | 17 | 22 | 0 | 0 | 0 | 0 | 0 | 16 | 15 | 13 | 312 |
| North Eastern European | 17 | 1 | 16 | 301 | 43 | 1 | 7 | 6 | 3 | 48 | 31 | 5 | 3 | 26 | 10 | 58 | 65 | 0 | 0 | 0 | 0 | 0 | 49 | 38 | 70 | 798 |
| Basque | 1 | 6 | 2 | 83 | 17 | 0 | 3 | 0 | 1 | 3 | 1 | 1 | 0 | 6 | 0 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 3 | 7 | 2 | 155 |
| North Western African | 0 | 0 | 4 | 106 | 26 | 2 | 0 | 0 | 1 | 16 | 5 | 5 | 2 | 5 | 0 | 3 | 11 | 0 | 0 | 19 | 0 | 8 | 9 | 12 | 9 | 243 |
| Canarian | 8 | 0 | 0 | 105 | 6 | 0 | 0 | 0 | 3 | 18 | 9 | 1 | 0 | 28 | 2 | 5 | 14 | 0 | 1 | 3 | 39 | 0 | 9 | 12 | 15 | 278 |
| Fulani + Wolof + Serer | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 15 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 27 |
| Kurd | 0 | 1 | 0 | 23 | 4 | 1 | 1 | 0 | 2 | 3 | 0 | 0 | 0 | 6 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 10 | 13 | 78 |
| Algerian Jew | 0 | 0 | 1 | 9 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 2 | 20 |
| Libyan Jew | 0 | 0 | 0 | 27 | 3 | 4 | 0 | 0 | 2 | 2 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 35 | 80 |
| Moroccan Jew | 0 | 0 | 6 | 55 | 15 | 1 | 1 | 2 | 1 | 11 | 1 | 1 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 5 | 23 | 19 | 146 |
| Tunisian Jew | 0 | 0 | 2 | 13 | 3 | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 3 | 1 | 8 | 36 |
| Azeri Jew | 1 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 35 | 0 | 1 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 1 | 4 | 58 |
| Georgian Jew | 0 | 0 | 0 | 8 | 48 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 2 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 1 | 3 | 74 |
| Cochin Jew | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 26 | 44 |
| Mumbaikar Jew | 0 | 0 | 0 | 3 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 28 | 34 |
| Iranian Jew | 0 | 0 | 9 | 16 | 0 | 1 | 0 | 0 | 0 | 13 | 1 | 6 | 0 | 8 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 9 | 75 |
| Iraqi Jew | 0 | 0 | 0 | 16 | 6 | 0 | 3 | 1 | 6 | 23 | 0 | 25 | 0 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 18 | 3 | 27 | 135 |
| Near & Mid Eastern Jew | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 9 | 34 |
| Yemeni Jew | 0 | 0 | 0 | 5 | 29 | 28 | 16 | 0 | 0 | 30 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 6 | 9 | 139 |
| Spanish Exile Jew | 0 | 0 | 6 | 76 | 9 | 2 | 2 | 1 | 5 | 18 | 4 | 3 | 2 | 15 | 0 | 3 | 2 | 0 | 0 | 3 | 0 | 3 | 12 | 22 | 25 | 213 |

## Figure Captions

**Supplementary Figure S1**: J1b1a1 median-joining network constructed with haplogroup J control region sequences containing a transition at 16172. Only the sequences from Dubai covered nps 462 and 489 and the lack of these transitions in all other samples is thus artifactual. Node size is proportional to number of samples sharing the haplotype.

**Supplementary Figure S2**: U5b median-joining network constructed with haplogroup U control region sequences with transitions at 16270 and 150, or otherwise characterized by the authors as belonging to subhaplogroup U5b through coding region genotyping (Malyarchuk et al. 2002, 2003, Turchi et al. 2008). Node size is proportional to number of samples sharing the haplotype.

**Supplementary Figure S3**: H1b median-joining network constructed with haplogroup H samples containing a transition at 16356. Node size is proportional to number of samples sharing the haplotype. HVR-II sequences in the French samples in this network spanned up to position 222, and their lack of the transition at np 263 is thus artifactual.

**Supplementary Figure S4**: J2a median-joining network constructed with control region sequences containing the transition at np 195, at least two of transitions 16069, 16126 and 295, and at least one of transitions 150 and 152 relative to the rCRS. All differences between the three major haplotypes within J2a1a1, 319 and 489, are artifacts created by the lack of sequencing of the region containing those sites in the studies from which the sequences were obtained.

**Supplementary Figure S5**: Median-joining network of haplogroup T control region sequences containing a transition at 16304 that defines subhaplogroup T2b. Node size is proportional to number of samples sharing the haplotype.

**Supplementary Figure S6.** Geographic representation of posterior probability values (%) distribution across Western Eurasia, North Africa, Senegal and Cameroon of WE-NA mtDNAs in Puerto Rico using the HVR-I-based classification scheme in Table S7 and the resulting data in Table S11. Posterior probability representations for dispersed Jewish populations are shown separately in squares. Probability values were represented for: A) global mtDNA sequences; B) haplogroup J sequences; C) haplogroup U sequences.

**Supplementary Figure S7:** PC plot of WE-NA populations and Puerto Rico based on mtDNA subhaplogroup frequencies (Table S11). PC1 vs PC2 was performed for: A) global mtDNA subhaplogroups; B) subhaplogroups H; C) subhaplogroups J; and D) subhaplogroups U.

**Supplementary Figure S8:** PC plot of WE-NA populations and Puerto Rico based on mtDNA subhaplogroup frequencies (Tables S10). PC1 vs PC3 was performed for: A) global mtDNA subhaplogroups; B) subhaplogroups H; C) subhaplogroups J; and D) subhaplogroups U.

**Figure S1.**

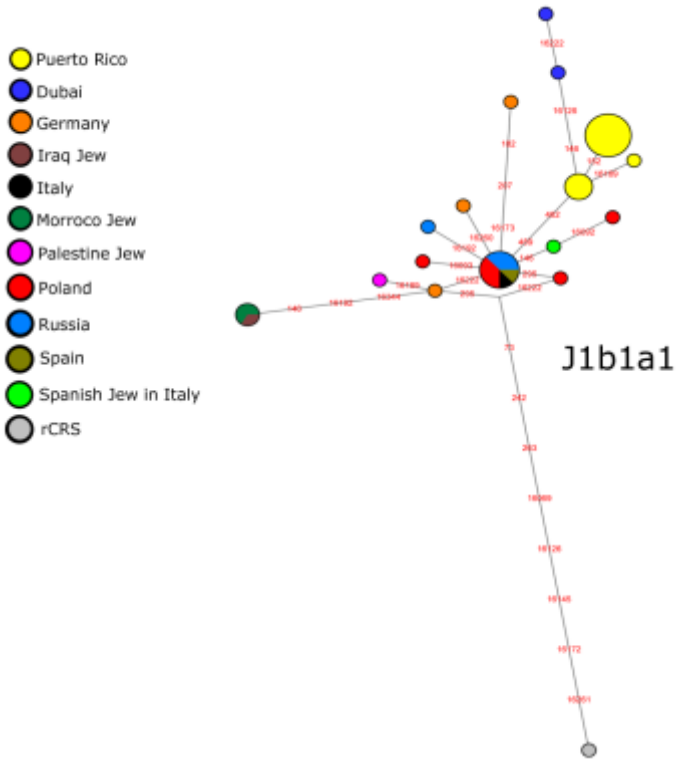

**Figure S2.**

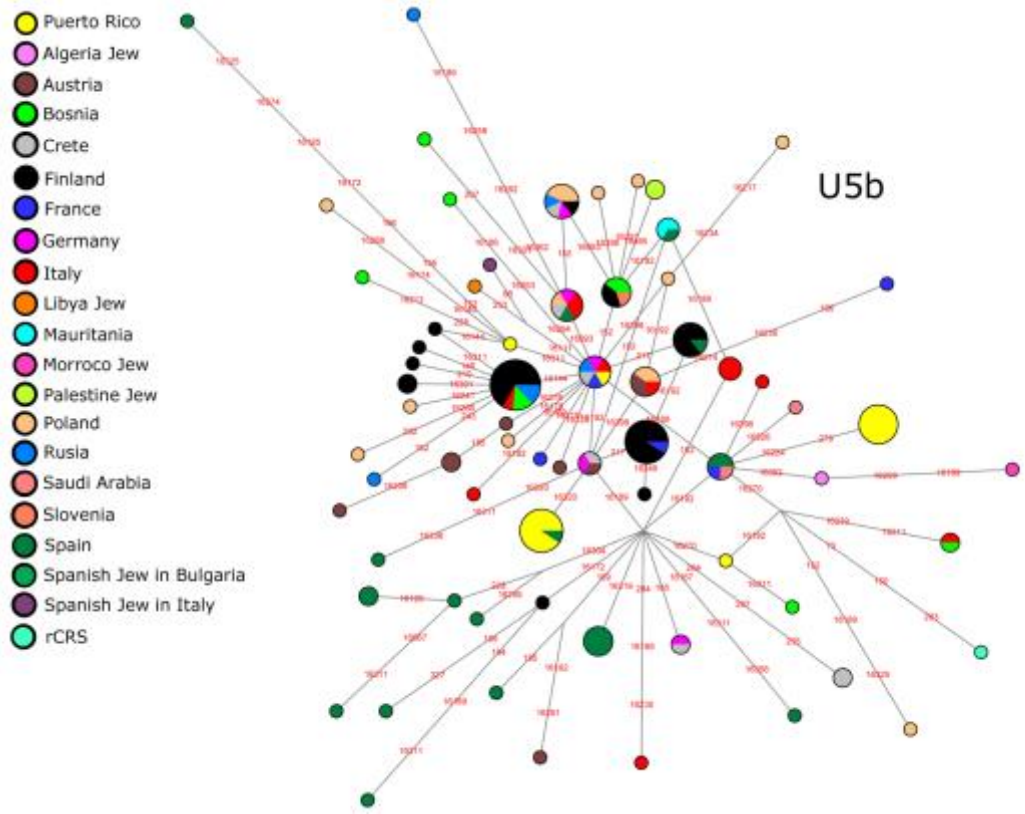

**Figure S3.**

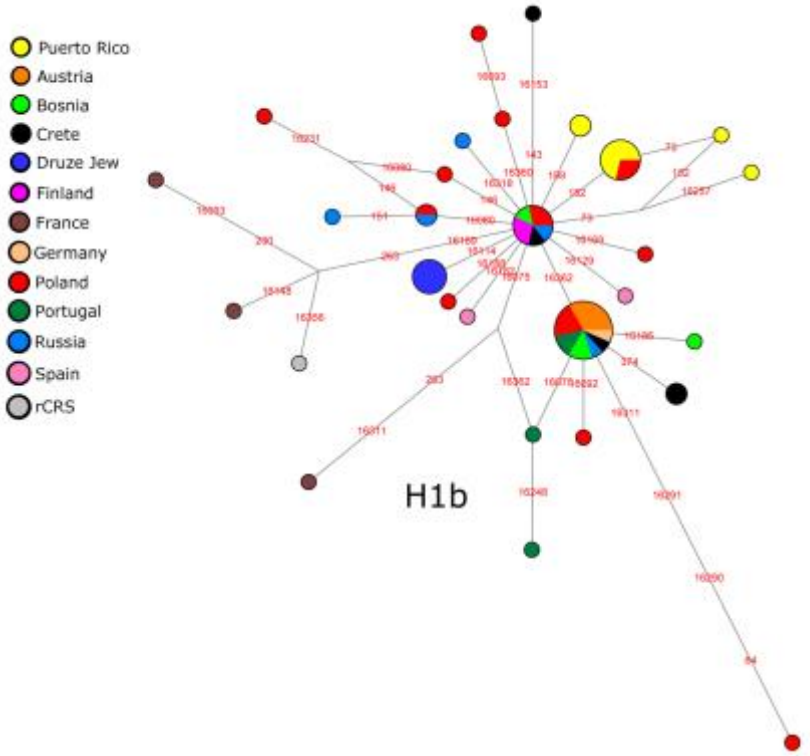

**Figure S4.**

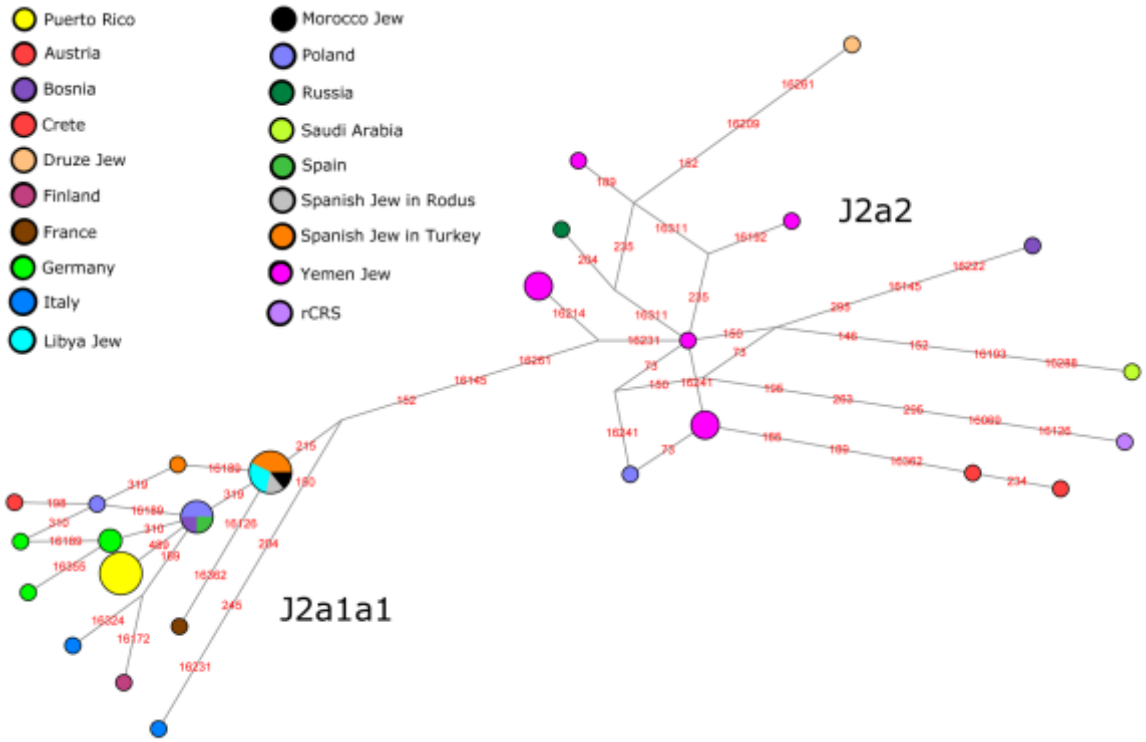

**Figure S5.**

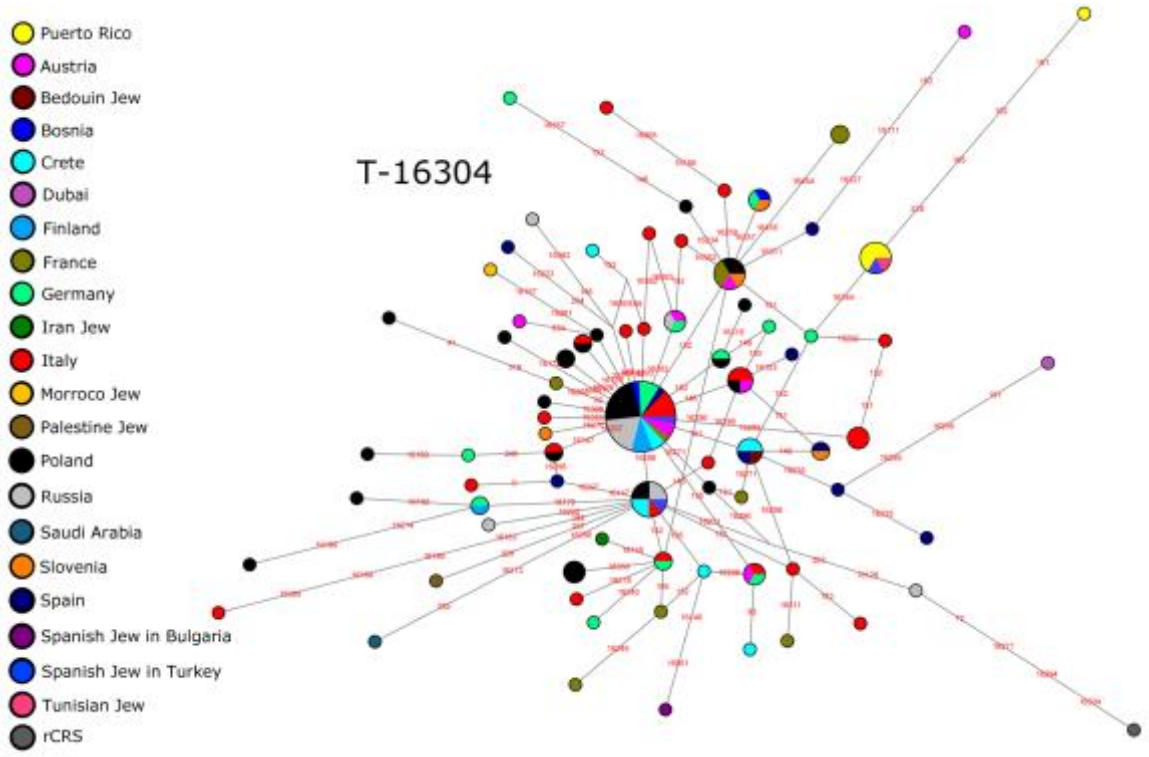

**Figure S6.**

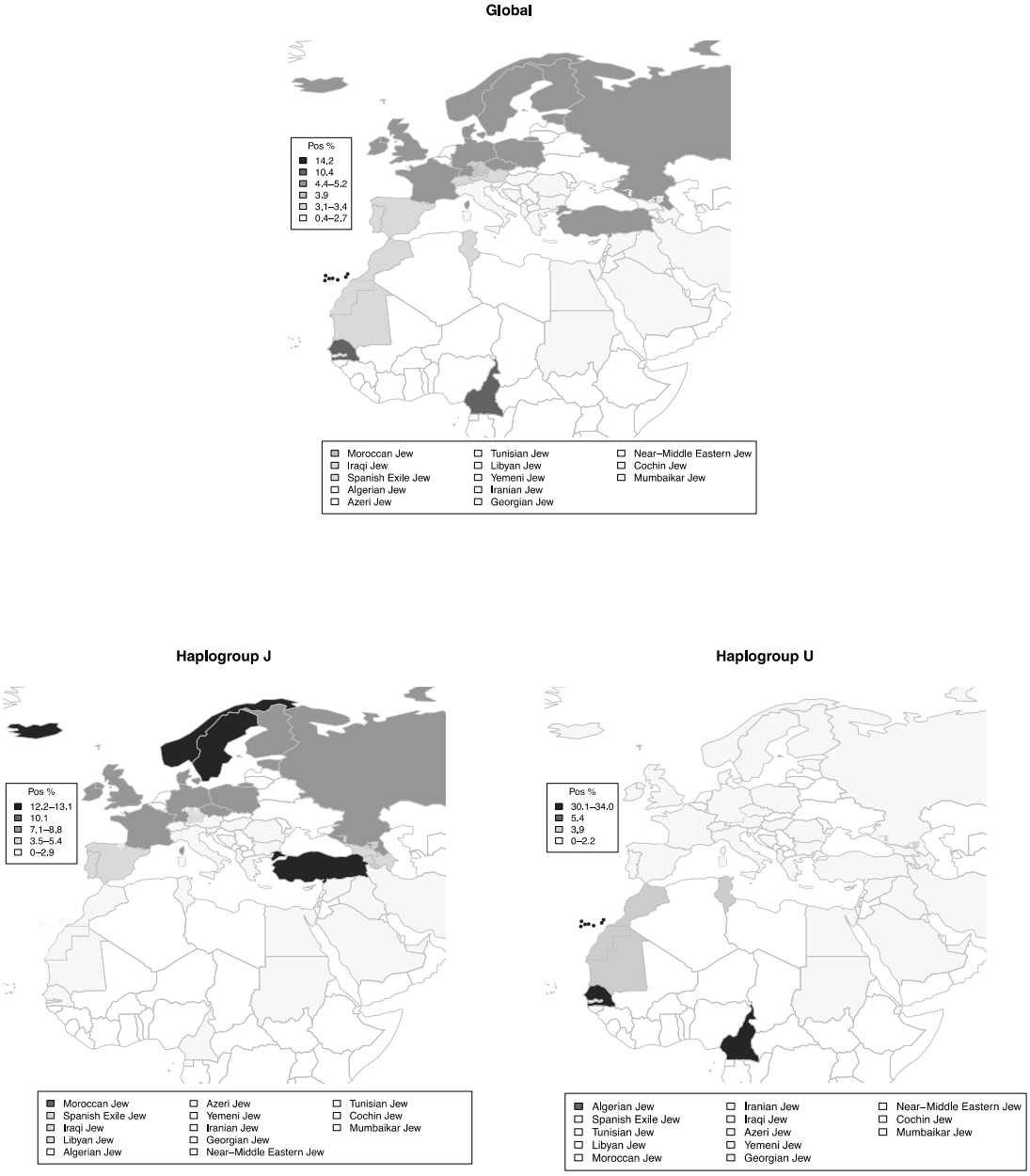

**Figure S7.**

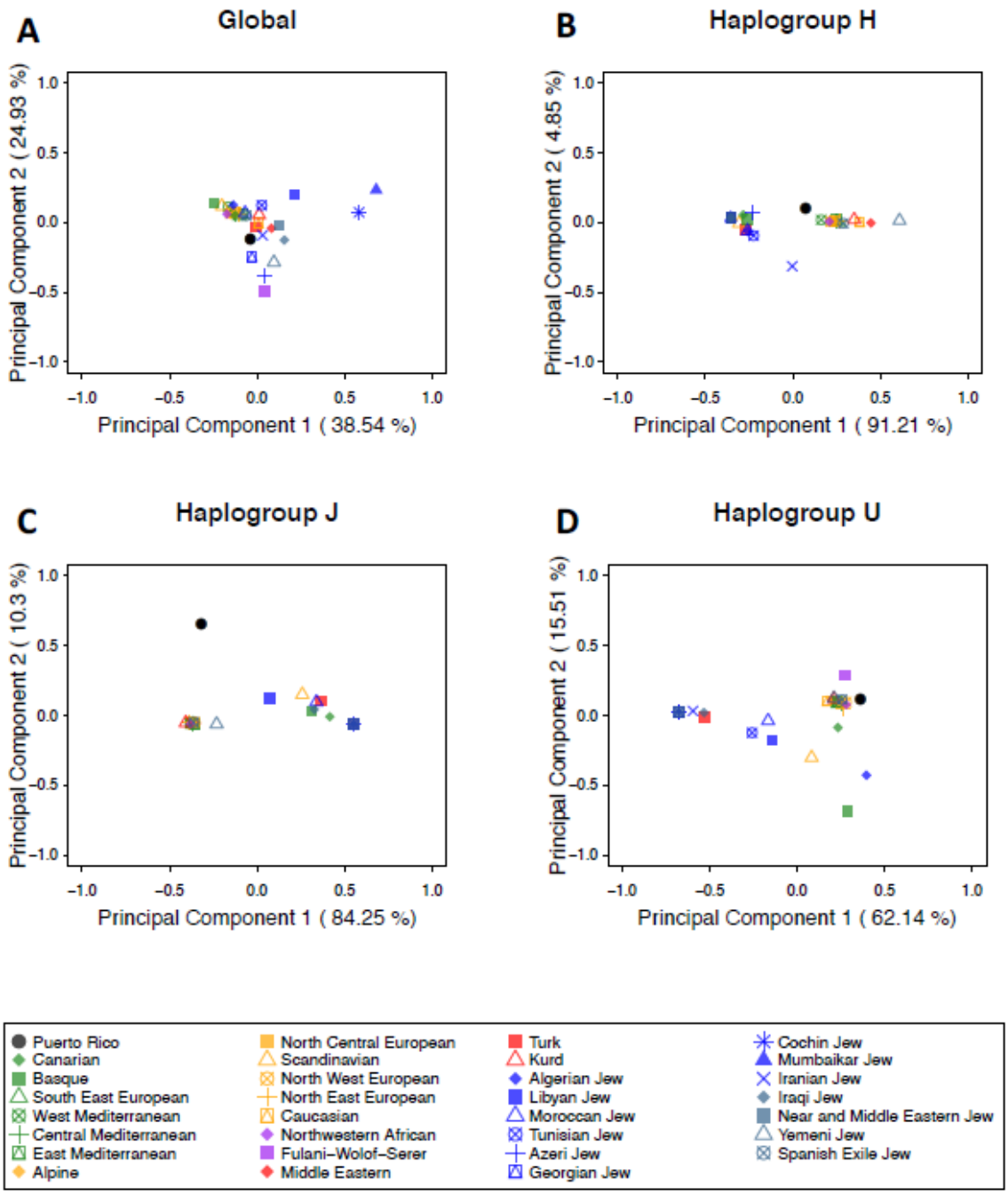

**Figure S8.**