GENETIC PRINTS OF AMERINDIAN FEMALE MIGRATIONS THROUGH THE CARIBBEAN REVEALED BY CONTROL SEQUENCES FROM DOMINICAN HAPLOGROUP A MITOCHONDRIAL DNAs

By

Arlin Feliciano Vélez

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCES

In

Biology

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS
2006

Approved by:

_____	_____
Rafael Montalvo Rodríguez, Ph.D. Member, Graduate Committee	Date
_____	_____
Rima Brusi, Ph.D. Member, Graduate Committee	Date
_____	_____
Juan C. Martínez Cruzado, Ph.D. President, Graduate Committee	Date
_____	_____
Lucy Bunkley-Williams, Ph.D. Director, Biology Department	Date
_____	_____
Carmen A. Vega. Ph.D. Graduate School Representative	Date

ABSTRACT

Both historical literature and popular opinion tend to establish that the population of the Dominican Republic descended mainly from Africans and Europeans. Amerindian ancestry was considered negligible due to Taíno extinction in the 16th century. Nevertheless, three independent mtDNA studies have recently revealed the presence of Native American haplogroups in the Dominican society. This study analyzed thirty-two Dominican mtDNA control region sequences from the Amerindian haplogroup A to determine lineages, feasible origins and their estimated time of entrance to the Dominican Republic by using median networks. Various diversity estimates, such as nucleotide diversity, haplogroup diversity, F_{ST} and S were calculated for Puerto Rico and the Dominican Republic. The results suggest the presence of 9 lineages: 6 possibly native and 3 of recent, probably post-Hispanic origin. The estimates for genetic diversity point to the existence of a diverse pre-Columbian population in Hispaniola and Puerto Rico. The estimated time of entrance for 5 native lineages was calculated using HVR-I sequences. It was estimated that 5 lineages were introduced to Hispaniola during the Archaic period. The Dominican Republic shared 34% of its sequences with Puerto Rico, all distributed between only two founder haplotypes, suggesting that both islands were colonized by these haplotypes, and that gene flow through females was reduced thereafter.

RESUMEN

Tanto la literatura histórica como la opinión popular tienden a establecer que la población de la República Dominicana desciende en su mayoría de africanos y europeos. La ascendencia indígena se ha considerado muy reducida debido a la supuesta extinción taína en el siglo XVI. Sin embargo, recientemente tres estudios independientes de ADNmt han revelado la presencia de haplogrupos nativo americanos en la sociedad dominicana. Este estudio analizó treinta y dos secuencias de la región control del ADNmt que pertenecían al haplogrupo A para determinar linajes, posibles orígenes y estimar tiempos de entrada a la República Dominicana utilizando redes medianas. Se calcularon varias medidas de diversidad tales como diversidad nucleotídica, diversidad de haplogrupo, F_{ST} y S para Puerto Rico y la República Dominicana. Los resultados demuestran que hay 9 linajes: 6 posiblemente nativos y 3 de origen reciente, probablemente post-colombino. Los estimados obtenidos apuntan a la existencia de poblaciones pre-colombinas genéticamente diversas en la República Dominicana y en Puerto Rico. Se estimó que 5 linajes se introdujeron a La Española durante el periodo Arcaico. La República Dominicana compartió el 34% de sus secuencias con Puerto Rico, todas distribuidas en sólo por dos haplotipos fundadores, sugiriendo que ambas islas fueron colonizadas por estos haplotipos y que el posterior flujo génico femenino fue limitado.

ACKNOWLEDGEMENTS

I would like to thank Dr. Lynne Guitar for allowing me to gain from her expertise in Dominican history and culture. I would also like to thank the graduate students Enid Gómez, Kathy Martínez, Aury Rodríguez and Tirtsa Porrata-Doria for their tips and support. My most sincere appreciation is for my graduate committee members, doctors Rafael Montalvo Rodríguez, Rima Brusi and Juan C. Martínez Cruzado, for their constructive criticism and advice. In addition, I would like to thank the Dominicans who volunteered for this study, and Arlene Álvarez, director of El Museo Arqueológico Regional de Altos de Chavón, who collected most of the samples used for this study. I would also like to thank Dr. Fernando Bird Picó for his assistance in population genetics statistics.

TABLE OF CONTENTS

TABLE LIST

FIGURE LIST

EQUATION LIST

INTRODUCTION

The Dominican Republic is located on the eastern side of Hispaniola, the second largest island in the Caribbean Sea, and has approximately 8 million inhabitants. It is widely accepted that its inhabitants are of European and African ancestry (Moya Pons, 1998). Although Taínos had populated Hispaniola for thousands of years, Africans and Europeans rapidly displaced them within the first fifty years of colonization. Taíno extinction has since been claimed. Nonetheless, the persistence of Taíno elements, such as words for objects and places, in addition to phenotypic traits in the population, has led some contemporary anthropologists to question the theory of Taíno extinction and suggest that the disappearance of the Taínos as a population was due to assimilation (Guitar 2002; Forte 2006).

Mitochondrial DNA (mtDNA) is a valuable molecular tool that has been used to demonstrate indigenous inheritance in other Caribbean islands such as Puerto Rico and Aruba (Martinez-Cruzado et al. 2001; Toro-Labrador et al. 2003). MtDNA facilitates the study of female migrations due to its strict maternal non-recombinant mode of inheritance and a high mutation rate (Wallace and Torroni 1992). These mtDNA studies reported the presence of Amerindian haplogroups, or sets of mtDNAs that share mutations inherited from a common female ancestor that distinguishes them from all other mtDNAs. Most haplogroups are continent-specific and some subhaplogroups specific to the indigenous populations of the North, Central and South American territories are known.

In the Dominican Republic, Martínez-Cruzado et al. (unpublished) demonstrated that 22% of samples from the Dominican Republic belonged to Amerindian haplogroups; haplogroup A obtained the highest frequencies. Nonetheless, these Amerindian samples could be from Amerindians brought as slaves throughout the Colonial Period. Therefore, it cannot be inferred that these people are descendants of Taínos.

This study's purpose is to elucidate this inquiry by constructing a reduced median network from haplogroup A control region sequences from the Dominican Republic (Bandelt et al. 1995, 2000). Median networks are phylogenetic interpretations of related haplotypes that incorporate genetic distance topology and haplotype frequency into one analytical tool. Median networks have aided the discovery of phylogenetic groups not identified by traditional phylogenetic methods (Forster et al. 1996). Moreover, they provide important information suitable for interpreting population history such as the founder haplotype identity population expansions and estimates for founding events.

In this study, I sequenced the mtDNA control region of thirty-two haplogroup A samples from the Dominican Republic. The sequences were then used to identify haplotypes, calculate a probable time of entrance, nucleotide diversity, genetic divergence and number of segregating sites (***S***). In addition, median network analysis led to the haplotypes being grouped into lineages, or families sharing diagnostic mutations. Six native lineages were identified: *ΔP* I, *ΔP* III, *ΔP* IV, *ΔP* VII, *ΔP* VIII and *ΔP* IX. Three of these lineages, *ΔP* III, *ΔP* VII and *ΔP* VIII, were not found elsewhere. Lineage *ΔP* I corresponded to the HVR-I founder haplotype of the New World and is ubiquitous throughout the Americas. In lineage *ΔP* IV, the ancestral haplotype was also observed in Puerto Rico and Cuba (Martinez-Cruzado, unpublished; Lalueza-Fox et al. 2003). Lineage *ΔP* IX was defined by a mutation found in South American sequences (Alves-Silva et al. 2000).

Subsequently, I compared the results to those of haplogroup A in Puerto Rico. The two major haplotypes found in the Dominican Republic, from lineages *ΔP* I and *ΔP* IV, were also identified in Puerto Rico. These Dominican samples pertain approximately to 34%; therefore suggesting a common early origin for haplogroup A in both islands. The estimated time of entrance was calculated for five putative native lineages. The *ΔP* III lineage was determined to be approximately 12,108 years old (σ = 9,024); perhaps representing an Archaic lineage that formed in Hispaniola given that it is unique to the island. The *ΔP* IV and *ΔP* IX lineages had an estimated time of

5,045 (σ = 5,045) years. The time of entrance for lineage *ΔP* VII was 6,727 years (σ = 6,727). Meanwhile, the *ΔP* I lineage arrived approximately 4,484 (σ = 3,171) years before present, suggesting pre- or post-ceramic migrations to Hispaniola.

LITERATURE REVIEW

I. Pre-Columbian History of the Caribbean

In 1492, Christopher Columbus stumbled upon indigenous populations, named Taínos, in Haiti (Nau 1982), currently known as Hispaniola. At the time, he was unaware of the large Amerindian populations that had occupied the Caribbean for thousands of years. The number of migrations that populated the Greater Antilles is a matter of debate; however, it is known that there were various migrations towards the Caribbean, which are mainly known from stone tools and ceramic remnants.

The first migration process introduced the Casimiroid culture to the Caribbean. The Casimiroids, recognized for their flint work and lack of pottery, were nomadic hunter-gatherers that made their way from Mesoamerica to Cuba and Hispaniola approximately 7,000 YBP (Keegan 1994). Another archaic migration from the Orinoco River Valley made its way up to Puerto Rico around 5,000 YBP. Known as the Ortoiroid culture, these people did not use flint or pottery. Pictographs in Mona Island, similar to others in Eastern Hispaniola serve as evidence that they journeyed this far, and also suggest relatively developed populations in this island (Davilá-Davilá 2003). Ceramic cultures started to migrate to the Caribbean from the Orinoco River Valley around 2,500 YBP. Identified as the Saladoid culture, they possessed various styles of pottery (Rouse 1992). The Saladoids made their way to Puerto Rico already an agricultural society. Their culture was distinctively Awarak, as the Taíno culture that thrived in the Caribbean when Columbus arrived.

Another Amerindian group was expanding up the Lesser Antilles and approaching the Greater Antilles during Columbian times. These people were the Caribs, also from the Orinoco River Valley and were raiding the Virgin Islands and Puerto Rico. Some believe the Ciguayos, who populated the Samaná Peninsula in Hispaniola, were descendants of the Caribs. They spoke a distinct language from the

other Hispaniola inhabitants, used bow and arrow, and painted their bodies red and black (Moya Pons 1998; Guitar, personal communication).

Diverse populations thus formed the Taínos. They were composed of linguistically and culturally diverse tribes, as described by Peter Martyr (Swanton 1968), member of Columbus' fleet, and Bartolomé de las Casas, a Spanish priest. The Ciboneys inhabited the southwestern side of Hispaniola, in what is now Haiti. The Ciboneys may have been descendants of Archaic immigrants since they lived in caves and did not use agriculture or pottery. The Macorix were warriors found on the eastern side of Hispaniola, possibly of Carib ancestry. The Maguana, dedicated to agriculture and fishing, were in the central and northern part of the island.

Unfortunately, except for a small, admixed Carib village in Dominica, there are no contemporary Amerindian populations to study in the Caribbean. This fact complicates the efforts to confirm pre-Columbian indigenous ancestry. The Taínos of Hispaniola were claimed extinct in the middle of the 16th century due to harsh treatment, diseases, war, slavery and suicide (Moya Pons 1998; Guitar 2002). Taíno contribution to modern day society has been limited in large part to nouns and certain foods. Still, physical anthropologists and geneticists have found a unique way to decipher Amerindian migrations to the Caribbean that problematize the claim of extinction by demonstrating indigenous contribution to the current Caribbean gene pool: mtDNA.

II. Mitochondrial DNA Analysis

Mitochondrial DNA is a circular molecule that has been widely used to study female human migrations in prehistory. MtDNA has various advantages over nuclear DNA, for instance, non-recombinant maternal inheritance and a high mutation rate. These attributes have made it possible to find population-specific polymorphisms. Restriction fragment length polymorphisms (RFLPs) are differences detected by the production of DNA fragments of distinct sizes, generated by restriction endonuclease

digestion. RFLPs were used to establish haplogroups, groups of mtDNAs that share a recent common ancestor from which it inherited a mutation that distinguishes them from all other mtDNAs (Schurr et al. 1990; Ballinger et al. 1992; Torroni et al. 1993). Since haplogroups tend to be continent-specific, they are useful in admixed-population genetic studies to determine maternal ancestry (Alves-Silva et al. 2000; Green et al. 2000; Martínez-Cruzado et al. 2005).

Extensive mtDNA analysis has demonstrated that Amerindian populations belong to five haplogroups: A, B, C, D, and X (Merriweather et al. 1995; Forster et al. 1996; Silva et al. 2002). Haplogroups A, B, C and D originated in Asia while haplogroup X may be of European origin (Brown et al. 1998). Haplogroups A, C and D were identified in ancient samples from Puerto Rico, the Dominican Republic and Cuba (Sánchez-Crespo 1999; Lalueza-Fox et al. 2001, 2003). Sánchez-Crespo analyzed 4 human bones unearthed in Arecibo, Puerto Rico. HVR-I analysis concluded that all four specimens belonged to haplogroup C. Bone samples were recovered from an excavation site in La Caleta, Dominican Republic (Lalueza-Fox et al. 2001). Enzyme digestion and HVR-I sequencing demonstrated that the samples belonged to haplogroups C and D. Lalueza-Fox et al. 2003 obtained similar results from 15 Ciboney samples from 3 different sites of Northwestern Cuba: Perico Cave, Mogote La Cueva and Canimar. Three Amerindian haplogroups were represented: 1 individual belonged to haplogroup A, 9 to haplogroup C and 5 to haplogroup D. Thus, mtDNA analysis of ancient Caribbean populations may lead us to believe that the most frequent haplogroups were C and D. In addition, haplogroup C and D are the most abundant in South America where the Taíno Arawak culture emerged, and from which some Caribbean Archaic populations are thought to originate.

MtDNA analyses in modern populations indicate that haplogroup A is the most frequent in the Caribbean. In Puerto Rico, in which 61% of the population belonged to Amerindian haplogroups, 52.4% corresponded to haplogroup A (Martínez-Cruzado et al. 2005). In the Dominican Republic, where 22% were of Amerindian origin, haplogroup A obtained 59.6% frequency among Amerindian

mtDNAs (Martínez-Cruzado et al. unpublished). An explanation for the disparity observed between modern and ancient haplogroup frequencies postulates post-Columbian migrations from Mexico and Central America as the cause for the predominance of haplogroup A in modern populations, as its frequencies are higher in those regions (Torroni et al. 1994; Santos et al. 1994; Kolman et al. 1995, 1997; Batista et al. 1995). However, pre-Columbian migrations from these areas could have been the very first source of people for the Caribbean, therefore explaining the large frequency of haplogroup A in contemporary populations of Puerto Rico and the Dominican Republic. In addition, control region sequences of Puerto Rican haplogroup A have shown haplotypes not found anywhere else, indicating putative native haplotypes. No definite conclusions can be attained due to the low sample number and strong geographic partition, which could account for the desegregation between modern and ancient haplogroup frequencies.

The Spanish Chronicles narrate Taíno genocide and enslavement in the Dominican Republic (Moya Pons 1998). Thus, Dominicans are considered mainly of African and European ancestry. Nonetheless, contemporary anthropologists state that the Taínos contributed more than just words to the Dominicans. It is believed that these chronicles state the disappearance of the Taínos to justify the importation of African slaves to the island. The Spanish counted only the Amerindians working in plantations for censuses; fleeing Taínos that hid in the mountains went unaccounted for (Guitar 2002). In addition, royal documents prepared during “El Repartimiento” in 1514, report 25,303 Indians distributed to Spaniards in Hispaniola and include the names of some of the Spanish men and women who were married to Taínos (Moya Pons, 1978). In addition to the unpublished work of Martínez-Cruzado et al., two other publications identify the presence of Amerindian mtDNA. One study, conducted to examine the linkage between obese diabetes and ethnic background in the Dominican Republic, revealed an indigenous frequency of about 33% (Tajima et al. 2004). Another study, which was examining African mtDNA, found Dominican samples belonging to Amerindian haplogroups (Brehm et al. 2002). These findings raise numerous questions. For example, are these samples of Taíno or post-

Columbian origin? How much gene flow existed between the Dominican Republic and Puerto Rico? Can pre-Columbian migration routes be resolved? This study attempts to find answers to these questions.

METHODS AND MATERIALS

Samples

Hair root samples were taken with appropriate informed consent in different sites throughout the Dominican Republic. One sample was collected from a Dominican residing in Puerto Rico from a previous study. The volunteers were asked where they lived and where their oldest known maternal female ancestor was natural from. This information was used to determine the geographic origin of the samples and compare it with its latest location. The samples were tested for Native American mtDNA ancestry by RFLP. Those samples possessing the haplogroup A specific +663 *Hae*III (Appendix: Table 4) were chosen for hypervariable region amplification.

DNA Extraction

The hair roots were stored in 500 μL of 5% Chelex (Sigma Chemical Co.) in labeled 1.5 mL microtubes. DNA samples were prepared as in Martínez-Cruzado et al. (2001).

Polymerase Chain Reaction Amplification (PCR)

HVR-I and HVR-II were amplified from the mtDNA samples using PCR primers presented in Table 1. The amplification reaction mix contained the following: 1X PCR Buffer (Sigma Chemical Co. 10X PCR buffer, catalogue No. D8312), 1.5 mM $MgCl_2$, 0.4 mM dNTPs, 1 μM of each primer, 7μL of DNA sample and 1.5 U of *Taq* DNA polymerase. The total reaction volume was 25μL. Duplicates were done to obtain 50μL of each PCR product for each sample. The PCR mix was heated at 94ºC for 5 minutes and then subjected to 32 cycles of 30 seconds at 94ºC, 1 minute at 54ºC and 70 seconds at 72ºC. To complete amplification, an extension cycle of 10 minutes at 72ºC was added.

Table 1. Primers used to amplify Hypervariable Regions

Hypervariable Region	Primers
HVR-I	H34 (accaaatgcatggagagctcc)
	L15766 or L15829
	(attctaacctgaatcggagg, catccgtactatacttcacaac)
HVR-II	H501 (gtgtgtgctgggtaggatg)
	L16453 or L16491
	(ccgggcccataacacttggg, ggggtagctaaagtgaactg)

Sequencing

Amplicons were purified using the Roche Diagnostics® PCR Product Purification Kit (Roche Applied Sciences, catalogue No. 1 732 676) following the instructions provided by the manufacturer. The purified products were sent along with the corresponding sequencing primers (Table 2), to be sequenced at the Molecular Resources Facility of the University of Medicine and Dentistry of New Jersey.

Table 2. Primers used sequence Hypervariable Regions

Hypervariable Region	Primers
HVR-I	H16526 (gggaacgtgtgggctatttagg)
	L15854 (cctaatcctaataccaactatc)
HVR-II	H283 (ggtttggtggaaatttttgt)
	H462 (gattagtagtatgggagtg)
	L16504 (gtgacctgtatccgacatctgg)

DNA Analysis

Sequence analysis was done using the computer programs Omiga (GCG), Chromas (Technelysium) and MEGA version 3.0 (Kumar, Tamura and Nei 2004). Sequences were aligned with the Cambridge Reference Sequence (Anderson et al. 1981) to observe mutations. All sites referred to pertain to those of the Cambridge Reference Sequence (CRS).

The mtDNA position 64 is thought to be hypermutable in haplogroup A (Martínez-Cruzado, personal communication); nonetheless, it was used for data analysis. Another well-documented hypermutable point mutation, 16519, however, was not taken into account for analysis. The transversion in position 16183 was also disregarded due to the fact that it is artificial, created from length variations in adjacent adenine and cytosine chains.

Median Networks

The obtained sequences were employed to construct two median networks by using the program Network 4.1.0.9 (www.fluxus-engineering.com/sharenet.htm). The first median network was built using HVR-I and HVR-II sequences from the Dominican samples. This network was constructed to observe the phylogenetic relationships between the defined haplotypes in this study. The mutations were counted to identify plausible native sequences that belong to the population in Hispaniola before the beginning of the Spanish colonization. Dominican and Puerto Rican median networks were compared to ascertain similarities and establish a putative relationship. In addition, the lineages might be specific to particular regions; therefore, local distributions of the lineages were also evaluated.

A second median network was constructed using only positions 16090-16365 to estimate the time of particular migrations. It was necessary to circumscribe the network to these positions because only for this region has the nucleotide substitution rate been calibrated (Forster et al. 1996). This network included our Dominican sequences and additional sequences that were obtained from Tajima et al. (2004). The Amerindian HVR-I sequences were downloaded from Genbank and eight belonged to haplogroup A. Founding haplotypes were determined, and time of entrance estimates were calculated using ρ, with standard deviations calculated as in Saillard et al. (2000) with the aid of the program Network 4.1.0.9. The accession numbers for the Tajima sequences were the following: AB174910, AB174911, AB174914, AB174917, AB174923, AB174932, AB174936, and AB174964.

Identification of Haplotype Lineages

In this study, haplotypes were categorized into lineages depending on the outcome of the median network. Two criteria were established to define the haplotypes lineages. The haplotypes in the lineage must share informative mutations in common. In addition, there must be no more than two mutations between nearest haplotypes in a lineage.

One disadvantage of this study is that it is difficult to identify with certainty whether samples are pre- or post-Columbian in origin because there are no Native American populations in the Caribbean to compare them to. In addition, there are few mtDNA studies from Taíno remains. The Dominican sequences were compared to continental HVR-I and HVR-II sequences obtained from previous publications (Santos et al. 1994; Kolman et al. 1995, 1997; Rickards et al. 1999; Alves-Silva et al. 2000; Moraga et al. 2000) in an effort to identify post-Columbian arrivals and putative migratory histories for native lineages. Other publications were also used for sequence comparison (Bonilla et al. 2004; Green et al. 2000; Lalueza-Fox et al. 2003 and Bolnick and Smith 2003), however they only display HVR-I sequences. As a criterion to separate haplotypes representing pre-Columbian migrations from those representing post-Columbian migrations, those with one or two mutations separating them from the founder haplotypes and not found in the mainland as shown by the literature search were regarded as having a putative pre-Columbian origin. This criterion was used under the assumption that native sequences may form cohesive lineages; not differing by many mutations from ancestral haplotypes. For example, Martínez-Cruzado (unpublished) has found for Puerto Rico networks that derived haplotypes differing from ancestral haplotype by not more than one mutation do not have a close relative in the continent whereas all others do.

Analyses of Haplotype Diversity

After sequence analysis, various population genetics statistics were employed to determine diversity and time parameters. Haplotype diversity was calculated using the formula for heterozygosity (Bolnick and Smith 2003) for the Dominican and Puerto Rican populations (Equation 1), where x_i is the frequency of the i^{th} haplotype, k is the number of haplotypes in the population and n is the sample size. The np 16519 was excluded from this analysis due to its hypermutable state.

$$h = \frac{n\left(1 - \sum_{i=1}^{k} x_i^2\right)}{n - 1}$$

Equation 1. Heterozygosity, a measure of haplotype diversity (Bolnick and Smith 2003).

F_{ST} is a measure of genetic diversity apportionment that compares diversity within populations *versus* between populations. It was calculated using Equation 2, where $\boldsymbol{H_T}$ is the heterozygosity calculated grouping together Dominican and Puerto Rican samples, while $\boldsymbol{H_S}$ was calculated by averaging the heterozygosity of Dominican and Puerto Rican calculated separately.

$$F_{ST} = (H_T - H_S) / H_T$$

Equation 2. Genetic Divergence within *vs.* between populations

Nucleotide diversity ($\boldsymbol{\pi}$) and number of segregating sites ($\boldsymbol{S}$) were calculated using HVR-I and HVR-II for Dominican and Puerto Rican samples. Additionally, $\boldsymbol{\pi}$ and $\boldsymbol{S}$ were calculated for 5 pre-Columbian lineages using nps 16090-16365. Nucleotide diversity is a measure of the average pairwise number of differences between samples while the number of segregating sites takes into account the number of variable nucleotide positions in the samples. Nucleotide diversity was calculated using equation 3, where k is the number of haplotypes in the population and p_i and p_j are the frequencies of the i^{th} and j^{th} haplotypes, respectively. d_{ij} is the number of different nucleotides between haplotypes i and j, and n is the number of samples. $\boldsymbol{S}$

was calculated as in Forster et al. (1996). Nucleotide diversity and number of segregating sites can be compared. The Neutral Theory of Molecular Evolution predicts that under conditions of constant population size and no selection $\boldsymbol{S}$ equals $\boldsymbol{\pi}$. $\boldsymbol{S} > \boldsymbol{\pi}$ is indicative of a demographic expansion in the studied population or of positive selection. On the other hand, if $\boldsymbol{S} < \boldsymbol{\pi}$, then either population subdivision or balanced selection occurred.

$$\pi = \frac{n\sum_{i=1}^{k}\sum_{j<i} p_i p_j d_{ij}}{n-1}$$

Equation 3. Nucleotide diversity

Unique haplotype frequency was measured for the Dominican Republic and Puerto Rico (Equation 4). A large frequency of unique haplotypes may signify a large population size or substantial gene flow.

$$R = \frac{\text{No. of unique haplotypes}}{\sum_{i=1}^{n-1} 1/i}$$

Equation 4. Unique haplotype frequency

RESULTS

I. Geographic Distribution of Haplogroup A in the Dominican Republic

The northern, central and eastern regions of the country attained the highest percentage of Amerindian ancestry, with haplogroup A as the most frequent in the northern and central regions (Figure 1). Monte Llano (ML), an extensively sampled small village located in the north coast, illustrated a low Amerindian heritage (11%); which is much lower than the remainder of the northern region. Therefore the samples acquired from this location were analyzed separately from other northern samples when performing analyses on their current location. The south, the capital city's location, and the western region revealed the lowest percentages of Amerindian ancestry. Analysis of the other Amerindian samples from the Dominican Republic indicated that the Amerindian haplogroups B and C obtained higher frequencies in the eastern side (Martínez-Cruzado et al., unpublished). Frequencies for current geographic distribution (Fig. 1b) showed that the western region apparently obtained a high percentage of haplogroup A samples; however, this represented only 1 of 2 samples. Haplogroup A demonstrated a pattern similar to the contemporary distribution; the majority of the haplogroup A samples were accumulated in the northern and central areas. The percentage of haplogroup A samples increased notably by 14% on the eastern side. The southern region showed the lowest haplogroup A frequency (0%).

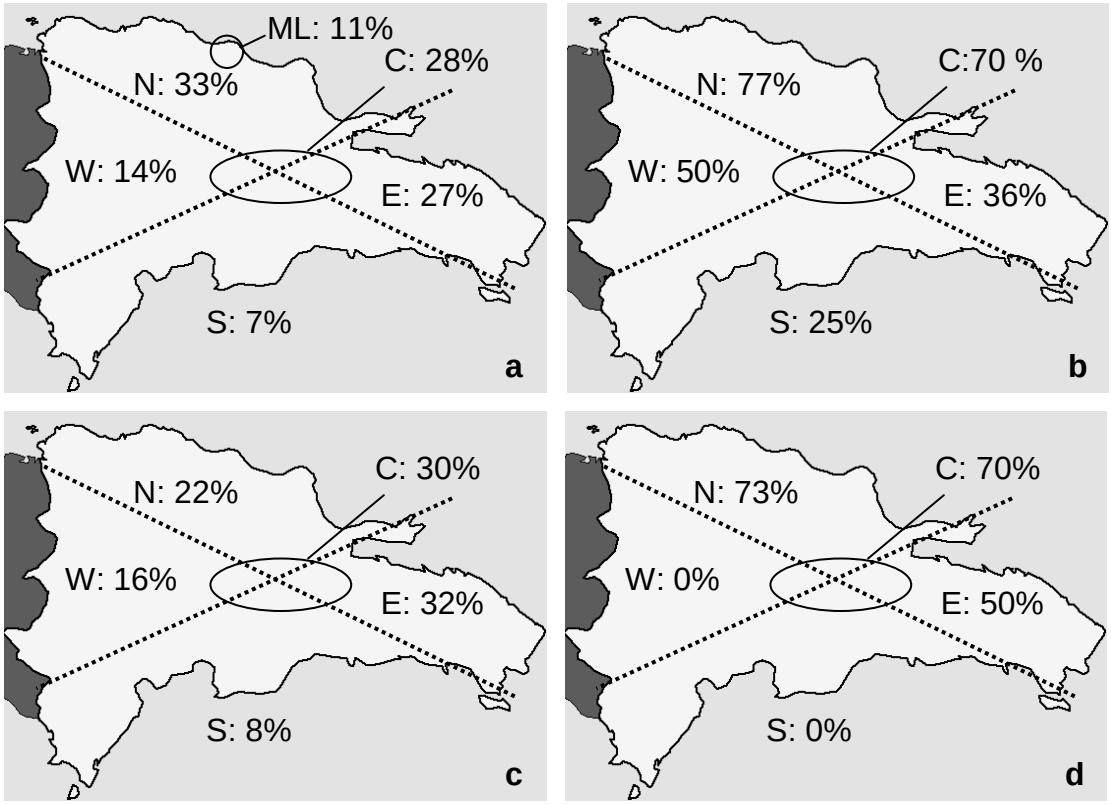


Figure 1. Current and ancestral distribution frequencies of indigenous samples. (a) Current frequency distribution of indigenous haplogroups, *n* = 235. (b) Current frequency distribution of haplogroup A mtDNAs among Amerindian samples, *n* = 52. (c) Ancestral frequency distribution of Amerindian haplogroups, *n* = 235. (d) Ancestral frequency distribution of haplogroup A mtDNAs among Amerindian samples, *n* = 52. ML = Monte Llano.

II. MtDNA Sequence Analysis

Thirty-two haplogroup A samples from the Dominican Republic were sequenced for HVR-I and HVR-II (Table 5). The most abundant haplotype, with a 25% frequency, had the haplogroup A HRV-I founder haplotype of the New World (Torroni et al. 1993). The founder haplotype is widely distributed in North, Central and South America, and is of Asian origin. The HVR-I founder haplotype was also found in Puerto Rico with a frequency of 31% among haplogroup A samples, but showed remarkable variation within HVR-II. It cannot be determined with certainty which of the samples having this haplotype is of native or post-Columbian origin, since it subsists in most New World populations.

The mutation 16129, which defines lineage *ΔP* IV, was present in five samples from the Dominican Republic; it was also present in 6 out of 49 Puerto Rican samples and in a HVR-I sequence from Cuban Ciboneys (Martinez-Cruzado et al. unpublished; Lalueza-Fox et al. 2003). This specific mutation has also been detected in other North and South American tribes (Alves-Silva et al. 2000; Moraga et al. 2000; Budowle et al. 2002). However, the nucleotide position (np) 16129 is hypermutable (Bandelt et al. 2002) and due to the large number of differences between the continental and Dominican haplotypes that carry this mutation, as well as its presence in both Cuban Ciboneys and modern Puerto Ricans, it was considered as of native origin.

Four other samples shared mutations with continental sequences. One sequence from lineage *ΔP* II revealed various mutations. The most important were the backward mutation in np 73, which is uncommon in haplogroup A samples, and a deletion in np 106-111. These two mutations were identified in mtDNAs of Costa Rica and Panama (Santos et al. 1994; Kolman et al. 1995). Two other sequences matched Brazilian sequences, having in common np 16266 (Alves-Silva et al. 2000). Another Dominican sequence from lineage *ΔP* I shared np 16092 with a Uruguayan HVR-I haplotype (Bonilla et al. 2004).

III. Median Networks

Two median networks were constructed for further analysis. The first median network, assembled with 32 Dominican HVR-I and HVR-II sequences (Figure 2), displayed 18 haplotypes and a high degree of complexity due to 3 reticulations. Furthermore, 7 hypothetical haplotypes were also observed. These hypothetical haplotypes may be present in the population and were not sampled or were present but have become extinct. Another possibility is that the hypothetical haplotypes may represent continental ancestors.

Nine lineages were identified in Dominican median network. The family *ΔP* I contained the New World founder, in addition to 2 haplotypes containing mutations at np 16189 and 1 haplotype with np 16092. The 2 haplotypes with the mutation at np 16189 were grouped in *ΔP* I because, although this mutation has also been identified in the Panama, Brazil and Ecuador (Kolman et al. 1995, 1997; Alves-Silva et al. 2000 and Rickards at al. 1999), this position has been shown to be hypermutable (Stoneking 2000), and thus its presence in the continent provides little evidence for a recent arrival. The mutation at np 16092 was also found in a HVR-I sequence from haplogroup A in Uruguay (Bonilla et al. 2004). However, because the geographic distance of Uruguay and due to the fact that the HVR-II sequence of the Uruguay sample was unavailable, this haplotype was grouped into *ΔP* I. Its existence provided little evidence that the 16092 Dominican haplotype was of post-Columbian origin. The Dominican haplotype with the mutation at np 16266 was included as a separate lineage because np 16266 is not a mutational hotspot. In addition, its 2 differences with the haplotype with mutations at np 16293 and np 153 are of little weight (Bandelt et al. 2003). Thus, the two haplotypes with the 16266 mutation combine to form lineage *ΔP* IX. Furthermore, the mutation at np 16266 has been found among Brazilian sequences (Alves-Silva et al. 2000).

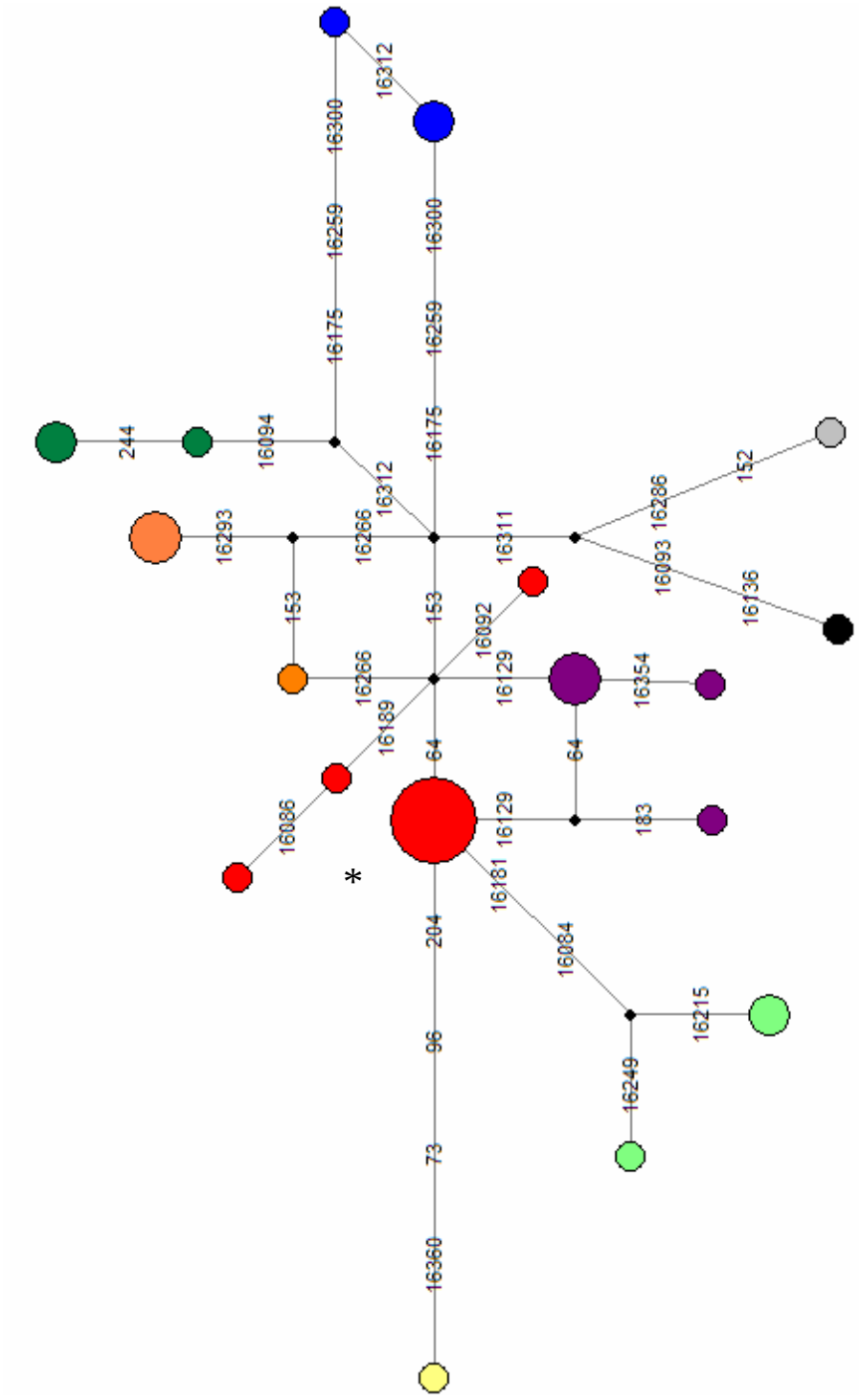
Three lineages, *ΔP* II, *ΔP* V and *ΔP* VI, were represented by 1 haplotype. These lineages are probably of post-Columbian origin since each haplotype

corresponded to one sample and contained 4 or more mutation differences from its nearest neighbor. Lineage *ΔP* II, which possesses the np 106-111 deletion and the lack of the np mutation 73, was linked to Central American tribes such as the Huetar and Ngöbé cluster (Santos et al. 1994; Kolman et al. 1995). Lineages *ΔP* V and *ΔP* VI share the np 16311 mutation, which was also found in a Brazilian haplotype (Alves-Silva et al. 2000); although they differ by 4 mutations from each other.

The remaining five lineages were considered to be of native origin. Lineages *ΔP* VII and *ΔP* VIII are more than 4 mutations differences from their closest neighbors, suggesting they represent distinct lineages. Both have haplotypes that are represented by 2 or more samples, and the derived haplotype within each lineage differs by only one mutation from the putative founder. Furthermore, these lineages contained sets of mutations that have not been identified elsewhere. *ΔP* III, represented by 2 haplotypes and totaling 3 samples, is a unique lineage since the set of mutations that define it (16084 and 16181) have not been identified elsewhere. On the contrary, *ΔP* IV and *ΔP* IX showed mutations that have been identified in other haplogroup A sequences. The *ΔP* IV lineage has np 16129 mutation which has been identified in the New World (Bolnick and Smith 2003 and Moraga et al. 2000). However, it was categorized as native given the hypermutable nature of this site (Stoneking 2002), and because this mutation was identified in Cuba (Lalueza-Fox et al. 2003) and Puerto Rico (Martinez-Cruzado, unpublished). Lineage *ΔP* IX shared the mutation np 16266 with a Brazilian sequence (see above); and the 2 haplotypes that compose it differ between themselves by only 2 unweighty mutations.

In addition, the local distribution of lineages was analyzed. The following significant associations were established:

- Approximately 63% of lineage *ΔP* I was located in the northern region.
- On lineage *ΔP* III, 66% were from Cotuí, situated in the central area.
- All of the lineage *ΔP* IV sequences were sampled in the north and central regions.



Legend:

●*ΔP* I ●*ΔP* II ●*ΔP* III ●*ΔP* IV ●*ΔP* V ●*ΔP* VI ●*ΔP* VII ●*ΔP* VIII ●*ΔP* IX

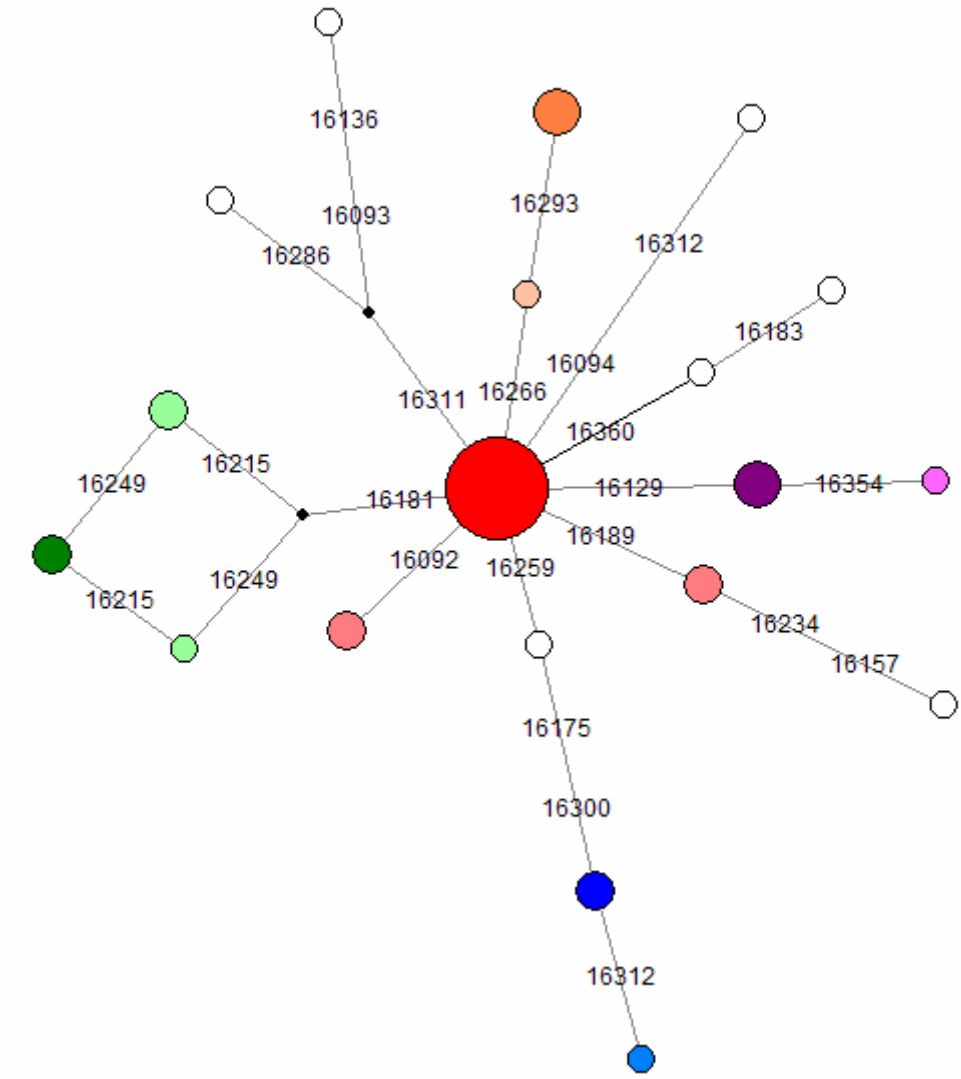
Figure 2. Median Network constructed with Dominican HVR-I and HVR-II sequences. Haplotype size is proportional to the frequency.

*np mutations relative to the CRS: 16111, 16223, 16290, 16319, 16362, 73, 146, 153, 235, 263

- All of Lineage *ΔP* VII sequences were also identified in the north and central areas.
- Lineage *ΔP* VIII was exclusive to the eastern region. One of the samples had a current location in the west; however its ancestral distribution was in the east.
- Lineage *ΔP* IX was exclusive to the northern region, specifically Puerto Plata.

Another median network (Figure 3) was constructed using only sequences from nps 16090-16365 (Table 6), incorporating 8 Dominican sequences from Tajima et al. (2004). It displayed a star-like topology, consistent with an early demographic expansion following a founder event. All haplotypes derive by few mutations from the founder haplotype, which was located at the center of the network. The Tajima et al. (2004) samples shared some of the mutations observed in this study. Three of the samples belonged to lineage *ΔP* I, which contains the HVR-I founder haplotype of the New World and of the whole network, excluding lineage *ΔP* III. Another 2 samples from Tajima et al. (2004) formed an ancestral haplotype for lineage *ΔP* III. One, with a mutation at np 16259, seems to be ancestral to *ΔP* VII. Another may belong to the putatively Central American lineage ΔP II, and the remainder is probably an independent offshoot from the founder haplotype.

The time estimates for 5 lineages, *ΔP* I, *ΔP* III, *ΔP* IV, *ΔP* VII and *ΔP* IX, were calculated with Network 4.9.0.1, using the putative native founder haplotypes of these lineages (Table 3). Calculations for $\boldsymbol{S}$ and $\boldsymbol{\pi}$ for these lineages were also included Table 3. Lineage *ΔP* VIII was not included since it did not display derived haplotypes in nps 16090-16365. Lineage *ΔP* III was the oldest lineage, probably pertaining to Archaic migrations which may have been exclusive to Hispaniola. The most recent was *ΔP* I, also most likely of Archaic origin.



Legend:

●Hypothetical ●● *ΔP* I ●● *ΔP* III ●● *ΔP* IV ●● *ΔP* VII ●● *ΔP* IX

Figure 3. Median network constructed using Dominican and Tajima et al. (2004) sequences using positions 16090-16365.
*np mutations with respect to the CRS: 16111, 16223, 16290, 16319, 16362

Table 3. Time estimates with standard deviation, S and π for Dominican lineages.

Lineage	Time estimate (ρ) Years	Standard deviation (σ) Years	S	π
ΔP I	4,484	3,171	0.629	0.264
ΔP III	12,108	9,025	0.960	0.500
ΔP IV	5,045	5,045	0.545	0.250
ΔP VII	6,727	6,727	0.667	0.333
ΔP IX	5,045	5,045	0.545	0.250

Migration time estimates are calculated under the assumption that a strong population expansion occurred shortly after arrival. The putative population expansion should leave a signature in the population such that within the 16090 to 16365 region, $\boldsymbol{S}$ must be larger than $\boldsymbol{\pi}$ and $\boldsymbol{\pi}$ larger than $1.5\boldsymbol{\rho}$ (Forster et al. 1996)**.** Table 3 shows $\boldsymbol{S}$, $\boldsymbol{\pi}$ and $1.5\boldsymbol{\rho}$ for lineages *ΔP*I, *ΔP*III, *ΔP*IV, *ΔP*VII and *ΔP*IX. The condition $\boldsymbol{S} > \boldsymbol{\pi} > 1.5\boldsymbol{\rho}$ is not met for any of the lineages. This could have been caused by different factors.

- Significant population expansions did not occur in prehistory.
- Population bottleneck: Due to historical events during the colonial period, the Amerindian population of the Dominican Republic diminished causing an excessive reduction of the founder haplotype frequency, in that way increasing $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$ statistics.
- Sampling error: Small sampling sizes increases sampling error and have the same effect as population bottlenecks.
- Putative inclusion of imported haplotypes: The inclusion of imported haplotypes of low frequency could artificially increase $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$ statistics.

IV. Comparison between the Dominican Republic and Puerto Rico

The median network constructed with Dominican sequences was compared to the Puerto Rican network (Martinez-Cruzado et al., unpublished). The Dominican median network shared approximately 34% of its samples with Puerto Rico in only 2 haplotypes. The most common haplotype between the two countries was the New World founder haplotype, which accounted for 25% of the Dominican samples and 14% of the Puerto Rican samples. The other contained the np 16129 mutation from

lineage *ΔP* IV and accounted for 9% of the Dominican samples and 12% of the Puerto Rican samples. Derived haplotypes identified in the Dominican Republic were not found in Puerto Rico, and vice versa.

V. Analyses of Haplotype Diversity

Haplotype diversity, measured by heterozygosity, was slightly higher in Puerto Rico, $h = 0.918$, than in the Dominican Republic, $h = 0.896$. The value of F_{st} was only 0.0233; indicating most of the variation exists within and not between populations. The variation between populations does not add significantly to the variation existing within populations.

Nucleotide diversity measured with HVR-I and HRV-II for Puerto Rico and the Dominican Republic demonstrate that the nucleotide diversity in the Dominican Republic, $\pi = 2.137$, was larger than that for Puerto Rico, $\pi = 1.945$. In contrast, $\boldsymbol{S}$ was smaller for the Dominican Republic, $\boldsymbol{S} = 5.753$, than for Puerto Rico, $\boldsymbol{S} = 6.728$. A larger S than π scenario is a signal of expanded populations when there is no positive selection (Tajima, 1989). Hence, this points towards an expansion in the native populations of both islands, especially in Puerto Rico.

Unique haplotype frequencies for Dominican Republic and Puerto Rico were 2.731 and 1.794, respectively, probably indicating a larger proportion of post-Columbian arrivals. Alternatively, a larger population size and more complex pre-Columbian haplotype network in combination with a smaller sample size in the Dominican Republic (32 to 49) may have lead to an artificially larger unique haplotype frequency in this country.

DISCUSSION

I. Geographic Distribution of Haplogroup A in the Dominican Republic

The highest frequencies of Amerindian ancestry were concentrated in the north, central and eastern regions of the country. This can be due to the establishment of the Spanish capital, Santo Domingo, in the south. Taínos fled to the central region, where the number of Spanish colonists was lower. Spanish Chronicles narrate that Indian and African slaves fled together to the mountains in the central mountains, hiding to avoid recapture. A similar effect occurred in Puerto Rico, where higher frequencies of indigenous samples were in the central, southern and western regions of the country, away from its capital in the east, San Juan (Martínez-Cruzado et al. 2005).

Haplogroup A frequency in the eastern side was much lower in comparison to other Native American haplogroups. It has been documented that the Taínos in Hispaniola were a heterogeneous population, consisting of various tribes that spoke distinctive languages. It is thus plausible that these tribes had different mtDNA haplogroup frequencies. The eastern side, where haplogroup C predominated, was inhabited by the Macorix in the north, and the Ciguayos, who settled in the northeastern side of Samaná. In addition, the southeast may have been inhabited by tribes closely related to the Taínos of Puerto Rico. The Maguana occupied the north and central regions, while the Ciboneys lived in the western area. These tribes may have a Central American origin, where haplogroup A predominates, thus explaining the higher haplogroup A frequency in the north and central regions of the Dominican Republic. The higher haplogroup A frequency in the north could also have been influenced by the Indian slave trade that reportedly stripped the Bahamas of their inhabitants. Slave ships departed and arrived to harbors in the northern coast of the Dominican Republic, and it is thus expected that any influence of this slave trade will be expressed in the northern region (Anderson-Córdova 1995).

II. MtDNA Sequence Haplotype Analysis

The New World founder haplotype obtained the highest frequency in the Dominican sequences. Some sequences were perhaps introduced during pre-Columbian colonization of the Caribbean islands, while others were imported as slaves during the Spanish colonization. It is difficult to identify which are pre- and post-Columbian in origin. However, it may be expected that pre-Columbian haplogroups are more frequent and may form cohesive families, whereas post-Columbian migrants may be represented by isolated haplogroups of very low frequency.

In the HVR-I and HVR-II median network containing the 32 Dominican sequences (Figure 2), 3 reticulations were formed. This is true particularly around the hypothetical haplotype with a mutation at np 64, related to the HVR-I founder haplotype of the New World. This hypothetical haplotype was found in Puerto Rico. It was identified as the HVR-I, -II New World founder (Malhi et al. 2001), and is thus likely present in the Dominican Republic.

Three putative post-Columbian haplotypes that demonstrated low frequencies were identified. They formed lineages, *ΔP* II, *ΔP* V and *ΔP* VI, that were represented each by only one haplotype and sample. In addition, they revealed mutations identified in continental sequences and differed by various mutations from neighboring haplotypes. These samples possibly originated by post-Columbian colonization events where Europeans introduced Amerindians from continental destinies into the Dominican Republic, as wives, servants or slaves.

Six lineages are probably pre-Columbian in origin. These lineages, *ΔP* I, *ΔP* III, *ΔP* IV, *ΔP* VII, *ΔP* VIII and *ΔP* IX, had unique mutations, sets of mutations found in previous Caribbean mtDNA studies or, as *ΔP* I, included the HVR-I haplotype founder of the New World. There were three lineages, *ΔP* I, *ΔP* IV and *ΔP* IX, that

shared continental mutations while *ΔP* III, *ΔP* VII, and *ΔP* VIII revealed unique mutations.

The median network constructed with only nps 16090 to 16365 exhibited a star-like morphology, as did the Amerindian networks constructed by Forster et al. 1996. The estimated time of entrance of the New World founder haplotype to the Dominican Republic was 4,484 YBP and the standard deviation was 3,171 years, ranging between 7,655 and 1,313 years. This timing overlaps with calculations for Puerto Rico and may correlate to Archaic migrations to the Greater Antilles.

The time of arrival of the *ΔP* III lineage was approximately 12,108 YBP, with a standard deviation of 9,025 years, clearly indicating a Pre-Ceramic origin. The mutations from the *ΔP* III lineage were not depicted in any other publications, making it a suitable candidate for an authentic Dominican lineage. The *ΔP* IV lineage was introduced 5,045 ± 5,045 YBP and thus could have arrived at Pre- or Post-Ceramic times. However, lineage *ΔP* IV is characterized by the np 16129 mutation, which has been identified in modern Puerto Ricans and Cuban Ciboneys, and thus is very probably native to the Caribbean.

The condition $\boldsymbol{S} > \boldsymbol{\pi} > 1.5\boldsymbol{\rho}$ was not met for any of the lineages for which variation was found within the 16090 to 16365 region, and thus did not show the population expansion required to apply the $\boldsymbol{\rho}$ statistics for estimating time of arrival. It is possible that population expansions were never significant. Due to its highest frequency in both Puerto Rico and Hispaniola, it is feasible that haplogroup A arrived to the Greater Antilles during the Archaic Period. It is well-known that hunter-gatherer populations lack the expanding power of agricultural societies because of their nomadic nature and late weaning times. Thus, haplogroup A may represent Archaic migrations that never experienced significant population expansions.

Alternatively, population bottlenecks and too small sample sizes may alter the relative haplotype frequencies that determine $\boldsymbol{\pi}$ and $\boldsymbol{\rho}$ statistics. Traditional history

strongly suggests a drastic reduction in the indigenous population size during the early colonial period. Just as importantly, the fact that the haplogroup may contain six native lineages is a strong factor reducing lineage sample size. The six native lineages could have arrived simultaneously or at different times in prehistory. However, the diversity found within them in spite of the small number of samples from each of them suggests their presence in the Dominican Republic for a considerable amount of time. If so, their apparent geographic partition could be the signature of separate arrivals or of genetic drift in a small population dispersing across this large island.

Another explanation to the failure to meet the conditions signaling a significant population expansion may be the incorporation of haplotypes arriving to Hispaniola in historic times. Anderson-Cordova (1990) described the development of an Amerindian Slave Trade that emptied the population of the Bahamas into the Dominican Republic and that brought thousands of Amerindians from Mexico, Central America, Margarita Island and the northern coast of Venezuela to Hispaniola. Although I eliminated from the time estimate analysis the unique haplotypes that were separated from their nearest neighbors in the median network by more than two mutations, the inclusion in the analysis of imported haplotypes closely related to native haplotypes cannot be discarded.

Further analyses of haplogroup A sequences from the Dominican Republic, the Caribbean and the Americas could discern if the hypothetical haplotypes are present and aid in determining origins. Thus, this information can help resolve if $\boldsymbol{\rho}$ statistics can accurately be determined for the Dominican Republic.

III. Comparison between the Dominican Republic and Puerto Rico

The presence of the New World founder haplotype and the ΔP IV lineage suggests a common origin for Archaics in Puerto Rico and the Dominican Republic. The presence of derived haplotypes in the ΔP IV lineage that were not identified in

Puerto Rico and the presence of this lineage in Cuban Ciboneys may indicate that this is an ancient family, innate to the Caribbean.

Although the Dominican Republic shares 34% of the sequences with Puerto Rico, this represents only 2 haplotypes. Hence, although both populations seem to share a common origin, the expansions of the populations in the islands seem to have occurred rather independently from one another. The larger number of hypothetical ancestors in the Dominican Republic network could suggest the arrival of several, at least 6, independent lineages in pre-Columbian times. It could also be due to sampling error or to the drastic reduction in population size of the native population during the Spanish colonization. In Puerto Rico, the network of which shows no hypothetical ancestors (Figure 5), the impact of Spanish colonization was less severe than in the Dominican Republic, the first established New World colony. A 1777 census reported pure Indians in Puerto Rico (Brau 1904) while in the Dominican Republic Taíno extinction was assumed in the beginning of the 16th century.

It is also essential to remember that mtDNA analysis only examines female ancestry. It is possible that a minimum number of women were migrating between Puerto Rico and the Dominican Republic. This inter-island voyage is precarious and it is likely that more men were performing this journey; thus explaining the difference between haplotypes. The observed limited gene flow may refer only to females; it is still likely that substantial male-mediated gene flow occurred between the islands.

IV. Analyses of Haplotype Diversity

The slightly higher haplotype diversity obtained for Puerto Rico is possibly due to a stable indigenous population that was maintained on the island after colonization. The observed difference between nucleotide diversity of the Dominican Republic and Puerto Rican may be due to a large original population in the Dominican Republic possessing more mtDNA variability than Puerto Rico and later

suffering a more dramatic demise. This study cannot clearly ascertain if this interpretation is correct.

The elevated diversity also may be caused by pre-Columbian migrations of genetically diverse populations. Both countries also revealed ***S*** higher values when compared to nucleotide diversity. It can thus be inferred that the indigenous populations in Puerto Rico and Hispaniola experienced substantial growth.

Unique haplotype frequency demonstrated that the Dominican Republic had double the amount of unique haplotypes. This could be due to 3 things:

- The Dominican Republic had fewer samples. If more samples are taken, the number of unique haplotype could decrease.
- The imported haplotypes are present. These haplotypes were introduced to Hispaniola, probably throughout the European colonization.
- There was a large population expansion on Hispaniola that recently suffered a drastic reduction in size. This reduction could have been after European colonization.

Although the diversity estimates revealed significant variance, the Dominican median network (Figure 3) displays hypothetical haplotypes and 2 or more mutation differences between lineages. This could be due to a bottleneck event that occurred after the start of the European colonization. The alleles that survive a bottleneck event subsequently obtain the highest frequencies. When the median network is observed, unique haplotypes are generally peripherally located while more common haplotypes have a central position (Crandall and Templeton, 1993). When the bottleneck event is recent, fewer peripheral haplotypes and reduced levels of diversity are detected. If peripheral haplotypes become widespread, the time estimates cannot be made, as S may no longer be larger that π, as required (Forster et al. 1996, Bolnick and Smith 2003). Perhaps this was the case of the Dominican Republic. The varying analyses sustain the presence of genetically diverse population that underwent bottleneck event which influenced mtDNA variation in the Dominican Republic.

CONCLUSIONS

Haplogroup A control sequences from the Dominican Republic suggested the existence of 6 pre-Columbian and 3 post-Columbian lineages. The *ΔP* I lineage revealed the New World founder haplotype. The *ΔP* IV lineage, characterized by a mutation at np position 16129, was present in Puerto Rico and Cuba, serving as evidence that it represents a very early arrival to the Caribbean. Conversely, lineage *ΔP* IX may be South American in origin and has not been found in Puerto Rico. Thus, it could represent a direct migration from South America to Hispaniola as suggested by Veloz-Maggiolo (1991) to explain in part the arrival of a Ceramic culture in Hispaniola by 600 A.D. The *ΔP* III lineage was not identified elsewhere and may have originated in Archaic populations in Hispaniola. Lineages *ΔP* VII and VIII have not been found elsewhere. Three lineages, *ΔP* II, *ΔP* V and *ΔP* VI may represent post-Columbian migrations due to their low frequencies, large number of mutations from their nearest neighbors and shared mutations with continental sequences. Approximately 34%, representing 2 haplotypes, of the Dominican samples were shared with Puerto Rico; one of these also observed in Cuba. Therefore, genetic exchange occurring between the islands was at a reduced rate.

The time of entrance was estimated for the 5 putative native lineages, along with their respective standard deviations. The earliest migrations probably introduced the *ΔP* III lineage around 12,108 YBP. The *ΔP* IV and *ΔP* IX lineages were established approximately 5,045 YBP, during the pre-Ceramic period while *ΔP* VII lineage arrive around 6,727 YBP. The *ΔP* I lineage arrived to Hispaniola roughly 4,484 YBP, earlier in the pre-Ceramic period and similar to the time of arrival in Puerto Rico. All these estimates have typically very large standard deviations (Saillard et al. 2000). In addition, the population demise that followed the Spanish colonization may have artificially lowered these estimates.

This study reflects limited female indigenous history of the Dominican Republic. It demonstrates:

- Reduced gene flow between Puerto Rico and Hispaniola.
- A genetically diverse pre-Columbian population that suffered the likely extinction of native haplotypes.
- Introduction of continental haplotypes upon the European colonization.

More so, this study exhibits an incomplete panorama, giving way to other unanswered questions. For example, can the proposed hypothetical haplotypes or other Puerto Rican haplotypes be identified in the Dominican Republic from a larger sample? Are other Dominican haplotypes found in other parts of the Caribbean? Do the other Native American haplogroups from the Dominican Republic display the same behavior in their median networks? Further mtDNA analysis in the Dominican Republic and the Caribbean is recommended to unmask pre-Columbian history that was silenced for over 500 years.

RECOMMENDATIONS

1. An extensive study of the Dominican population should take place to determine their mtDNA ancestry and compare the results with the present study and Puerto Rico. Haplogroup A samples from this study should be sequenced to identify their haplotypes and consider other migration routes.

2. Sequences from other Amerindian haplogroups found in the Dominican population, such as B, C and D should also be analyzed to determine their putative origins. These results, in addition to those from haplogroup A, can be used to construct a migratory hypothesis.

3. MtDNA studies in other Caribbean islands, for example Cuba and Jamaica, should take place to develop a better image of the historic background, determine probable continental relationships and identify native haplotypes for the Caribbean.

4. Ancient mtDNA studies of Amerindian remains should be performed to identify putative origins and/or missing theoretical haplotypes.

5. Confirm the lineages proposed here using control region sequences by sequencing complete genomes of the distinct putative lineages.

LITERATURE CITED

Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.* 94: 444-461

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG. 1981. Sequence and organization of the human mitochondrial genome. *Nature.* 290: 457-465

Anderson-Córdova K 1990. Hispaniola and Puerto Rico: Indian acculturation and heterogeneity, 1492-1550. Ph.D. thesis, Yale University, New Haven, Connecticut

Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V. 2002 The fingerprint of phantom mutations in mitochondrial DNA data. *Am. J. Hum. Genet.* 71:1150-1160

Batista O, Kolman CJ, Bermingham E. 1995. Mitochondrial DNA diversity in the Kuna Amerinds of Panama. *Hum. Mol. Genet.* 4(5): 921-9

Ballinger SW, Schurr TG, Torroni A, Gan YY, Hodge JA, Hassan K, Chen KH, Wallace DC. 1992. Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics.* 130(1): 139-52

Bolnick DA, Smith DG. 2003. Unexpected patterns of mitochondrial DNA variation among Native Americans from the southeastern United States. *Am. J. Phys. Anthropol.* 122(4): 336-54

Bonilla C, Bertoni B, Gonzalez S, Cardoso H, Brum-Zorilla N, Sans M. 2004. Substantial Native American Female Contribution to the Population of Tacuarembó, Uruguay, Reveals Past Episodes of Sex-Biased Gene Flow. *Am. J. Hum. Biol*. 16: 289-297

Brau S. 1904. *Historia de Puerto Rico*. San Juan, Puerto Rico. Editorial Caqui.

Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A. 2002. Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann. Hum. Genet*. 66:49-66

Brown MB, Hosseini SH, Torroni A, Bandelt HJ, Allen JC, Schurr TG, Scozzari R, Cruciani F, Wallace DC. 1998. *Am. J. Hum. Genet*. 63(6): 1852-61

Budowle B, Allard MW, Fisher CL, Isenberg AR, Monson KL, Stewart JEB, Wilson MR, Miller KWP. 2002. HVI and HVII Mitochondrial DNA data in Apaches and Navajos. *Int. J. Legal Med.* 116: 212–215

Crandall KA, Templeton AR. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics.* 134: 959–969

Davilá-Davilá O. 2003. *Arqueología de la isla de la Mona*. Instituto de Cultura Puertorriqueña. San Juan, Puerto Rico

Forster P, Harding R, Torroni A, Bandelt HJ. 1996. Origin and evolution of Native American mtDNA variation: a reappraisal. *Am. J. Hum. Genet.* 59(4): 935-45

Forte M. 2006. *Indigenous resurgence in the Contemporary Caribbean: Amerindian Survival and Revival*. Peter Lang Publishing Group. New York, New York

Green LD, Derr JN, Knight A. 2000. mtDNA Affinities of the Peoples of North-Central Mexico. *Am. J. Hum. Genet.* 66: 989-998

Guitar L. 2002. Documenting the Myth of Taíno Extinction. *KACIKE: The Journal of Caribbean Amerindian History and Anthropology* [On-line Journal], Special Issue, Lynne Guitar, Ed. Available at: http://www.kacike.org/GuitarEnglish.html [Date of access: January 26, 2005]

Keegan WF. 1994. West Indian Archaeology I. Overview and Foragers. *J. Archaeological Research.* 2: 255-284

Kolman CJ, Bermingham E, Cooke R, Ward RH, Arias TD, Guionneau-Sinclair F. 1995. Reduced mtDNA diversity in the Ngöbe Amerinds of Panama. *Genetics.* 140(1): 275-83

Kolman CJ, Bermingham E. 1997. Mitochondrial and nuclear DNA diversity in the Choco and Chibcha Amerinds of Panama. *Genetics.* 147(3): 1289-302

Kumar S, Tamura K, and Nei M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics.* 5:2 (In press)

Lalueza-Fox, C, Luna-Calderon F, Calafell F *et al.* (2001). MtDNA from extinct Taínos and the peopling of the Caribbean. *Ann. Hum. Genet.* 65: 137-151

Lalueza-Fox, C, Gilbert, MTP, Martínez-Fuentes, AJ, Calafell, F, Bertran-petit J. 2003. Mitochondrial DNA from precolumbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean. *Am. J. Phys. Anthrop.* 121(2): 97-108

Malhi RS, Schultz BA, Smith DG. 2001. Distribution of mitochondrial DNA lineages among Native American tribes of northeastern North America. *Hum. Biol.*73: 17-55

Martinez-Cruzado JC, Toro-Labrador G, Ho-Fung V, Estevez-Montero MA, Lobaina-Manzanet A, Padovani-Claudio DA, Sanchez-Cruz H, Ortiz-Bermudez P, Sanchez-Crespo A. 2001. Mitochondrial DNA analysis reveals substantial Native American ancestry in Puerto Rico. *Hum. Biol.* 73(4): 491-511

Martinez-Cruzado JC, Toro-Labrador G, Viera-Vera J, Rivera-Vega MY, Startek J, Latorre-Esteves M, Román-Colón A, Rivera-Torres R, Navarro-Millán IY, Gomez-Sanchez E,Caro-González HY, Valencia-Rivera P. 2005. Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis. *Am. J. Phys. Anthropol.* 218: 131-155

Merriweather DA, Ferrell RE. 1996. The four founding lineage hypothesis for the New World: a critical reevaluation. *Mol. Phylogenet. Evol.* 5: 241-246

Merriweather DA, Rothhammer F, Ferrell RE. 1995. Distribution of the four founding lineage haplotypes in Native Americans suggests a single wave of migration for the New World. *Am. J. Phys. Anthropol.* 98: 411-430

Moraga ML, Rocco P, Miquel JF, Nervi F, Llop E, Chakraborty R, Rothhammer F, Carvallo P. 2000. Mitochondrial DNA Polymorphisms in the Chilean Aboriginal Populations: Implications for the Peopling of the Southern Cone of the Continent. *Am. J. Phys. Anthropol*. 113:19-29

Moya Pons F. 1998. *The Dominican Republic: A National History.* 2nd ed. Princeton, New Jersey: Markus Wiener Publishers.

Moya Pons F. 1978. La Española en el Siglo XVI 1493-1520: *Trabajo, Socoedad y Política en la Economía de Oro.* 3ra ed. Santiago, República Dominicana: Universidad Católica Madre y Maestra

Nau E. 1982. *Historia de los Caciques de Haití.* Santo Domingo, República Dominicana: Sociedad Dominicana de Bibliófilos Inc.

Rickards O, Martínez-Labarga C, Lum JK, De Stefano GF, Cann RL. 1999. mtDNA History of the Capaya Amerinds of Ecuador: Detection of Additional founding Lineages for the Native American Populations. *Am J Hum Genet.* 65: 519-530

Rouse I. 1992. *The Taínos: rise and decline of the people who greeted Columbus.* New Haven, Connecticut: Yale University Press.

Saillard J, Forster P, Lynnerup, Bandelt HJ, Norby S. 2000. MtDNA Variation among Green and Eskimos: The Edge of the Beringian Expansion. *Am J Hum Genet.*67: 718-726

Sánchez-Crespo A. 1999. Variación genética en una población indígena en Puerto Rico. MSc thesis. Mayagüez, Puerto Rico: University of Puerto Rico.

Santos M, Ward RH, Barrantes R. 1994. mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. *Hum. Biol.* 66:963-977

Schurr TG, Ballinger SW, Gan YY, Hodge JA, Merriweather DA, Lawrence DN, Knowler WC, Weiss KM, Wallace DC. 1990. Amerindian Mitochondrial DNAs have Rare Asian Mutations at High Frequencies, Suggesting they Derived from Four Primary Maternal Lineages. *Am J Hum Genet.* 46(3): 613-23

Silva WA, Bonatto SL, Holanda AJ et al. 2002. Mitochondrial genome diversity of Native Americans supports a single early entry of founder populations into America *Am J Hum Genet.* 71: 187-192

Stoneking M. 2002. Hypervariable Sites in the mtDNA Contol Region Are Mutational Hotspots. *Am J Hum Genet.* 67: 1029-1032

Swanton JR. 1968. *The Indian Tribes of North America*. Bureau of American Ethnology Bulletin 145-1953. Smithsonian Institute. 608-611

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

Tajima A, Hamaguchi K, Terao H, Oribe A, Perrotta VM, Baez CA, Arias HR, Yoshimatsu H, Sakata T, Horai S. 2004. Genetic background of people in the Dominican Republic with or without obese type 2 diabetes revealed by mitochondrial DNA polymorphism. *J Hum Genet.* 49: 495-499

Toro-Labrador G, Wever OR, Martínez-Cruzado JC. 2003. Mitochondrial DNA Analysis in Aruba: Strong Maternal Ancestry of Closely Related Amerindians and Implications for the Peopling of Northwestern Venezuela. *Caribbean J of Science.* 39.1: 11-22

Torroni A, Schurr, TG, Cabell, MF et al. (1993a). Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am. J. Hum. Gene*t. 53: 563-590

Torroni A, Sukernik, RI, Schurr, TG et al. (1993b). MtDNA Variation of Aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Gene*t. 53: 591-608

Torroni A, Chen YS, Semino O, Santachiara-Benerecetti AS, Scott CR, Lott MT, Winter M, Wallace DC. 1994. mtDNA and Y-chromosome polymorphisms in four Native American populations from Southern Mexico. *Am. J. Hum. Genet.* 54:303-318

Torroni A, Rengo C, Guida V, Cruciani F, Sellitto D, Coppa A, Luna-Calderon F, Simionati B, Valle G, Richards M, Macaulay V, Scozzari R. 2001. Do the Four

Clades of the mtDNA Haplogroup L2 Evolve at Different Rates? *Am. J. Hum. Genet.* 69:1348–1356

Wallace DC, Torroni A. 1992. American Indian Prehistory as griten in the Mitochondrial DNA: A Review. *Hum. Biol.* 64(3): 403-417

Veloz Maggiolo, Marcio. 1991. *Panorama Historico del Caribe Precolombino*. Banco Central de la Republic Dominicana, Santo Domingo

APPENDIX

Table 4. Dominican haplogroup A samples with their current locations, location of oldest known maternal ancestor, and identified by lineage.

Sample Identification	Current location	Ancestral location	Comments
D2	Cacique Monción	San José de las Matas, Santiago	*ΔP* VII
E1	Cacique Monción	Monción, Hato Viejo	*ΔP* I, NWF
E2	Cacique Monción	Valverde, Mao	*ΔP* IV
F9	Los Cocos, Monte Llano	Jagua, Santiago Rodríguez	*ΔP* I, NWF
G3	Tubagua, Puerto Plata	Pedro García, Santiago	*ΔP* IX
G7	Tubagua, Puerto Plata	Parque Yásira, Puerto Plata	*ΔP* IX
G11	Yásira, Puerto Plata	Yásira Arriba, Puerto Plata	*ΔP* I, NWF
G13	Yásira, Puerto Plata	Puerto Plata, Puerto Plata	*ΔP* I, NWF
I2	San José de las Matas, Santiago	San José de las Matas, Santiago	*ΔP* IV
J1	Piedras Blancas, El Seibo	Magarín, Hato Mayor	*ΔP* VIII
J6	Piedras Blancas, El Seibo	El Llano, El,Seibo	*ΔP* I
J7	Piedras Blancas, El Seibo	Los Cuamos,El Seibo	*ΔP* I
J12	El Seibo	El Seibo	*ΔP* VIII
K7	Santo Domingo	Jánico, Santiago	*ΔP* I, NWF
L3	Cotuí	Cotuí	*ΔP* IV
L7	Cotuí	Los Corojos, Hato Mayor	*ΔP* I
L22	Cotuí	La Vega	*ΔP* VII
L23	Cotuí	Fantino, Cotuí	*ΔP* III
L24	Cotuí	Fantino, Cotuí	*ΔP* VII
L25	Cotuí	La Vega	*ΔP* IV
L27	Cotuí	San Francisco de Macoris	*ΔP* I, NWF
L29	Cotuí	Cotuí	*ΔP* III
M5	San Juan de la Maguana	El Seibo	*ΔP* VIII
P6	Santiago	La Vega	*ΔP* V
P7	Santiago	Jávico, Santiago	*ΔP* I, NWF
P9	Santiago	Santiago	*ΔP* III
P10	Santiago	El Mamey, La Vega	*ΔP* I, NWF
R6	Puerto Plata	Altamira, Puerto Plata	*ΔP* II
R10	Puerto Plata	Luperón, Puerto Plata	*ΔP* IV
R12	Puerto Plata	La Isabela, Puerto Plata	*ΔP* IX
R15	Puerto Plata	Villa Isabela, Puerto Plata	*ΔP* IX
Y24	Cotuí	Not Available	*ΔP* VI, Obtained in PR

Table 5. Sequence mutations from HVRI and HVRII used to construct the median network (Figure 2).

NWF = New World Founder (E1, F9, G11, G13, K7, L27, P7 and P10)

Sample	16084	16086	16092	16093	16094	16111	16129	16136	16175	16181	16189	16215	16223	16249	16259	16266	16286	16290	16293	16300	16311	16312	16319	16354	16360	16362	16519	64	73	96	146	152	153	183	204	235	244	263
CRS	G	T	T	T	T	C	G	T	A	A	T	A	C	T	C	C	C	C	A	A	T	A	G	C	T	T	T	C	A	C	T	T	A	A	T	A	A	A
D2	•	•	•	•	•	T	•	•	G	•	•	•	T	•	T	•	•	T	•	G	•	G	A	•	•	C	•	T	G	•	C	•	•	•	•	G	•	G
E2	•	•	•	•	•	T	A	•	•	•	•	•	T	•	•	•	•	T	•	•	•	•	A	•	•	C	C	•	G	•	C	•	G	T	•	G	•	G
G3, R12, R15	•	•	•	•	•	T	•	•	•	•	•	•	T	•	•	T	•	T	G	•	•	•	A	•	•	C	•	T	G	•	C	•	G	•	•	G	•	G
L22, L24	•	•	•	•	•	T	•	•	G	•	•	•	T	•	T	•	•	T		G	•	•	A	•	•	C	•	T	G	•	C	•	•	•	•	G	•	G
R6	•	•	•	•	•	T	•	•	•	•	•	•	T	•	•	•	•	T	•	•	•	•	A	•	C	C	•	•	•	T	C	•	G	•	C	G	•	G
L23, P9	A	•	•	•	•	T	•	•	•	G	•	G	T	•	•	•	•	T	•	•	•	•	A	•	•	C	•	•	G	•	C	•	G	•	•	G	•	G
M5	•	•	•	•	C	T	•	•	•	•	•	•	T	•	•	•	•	T	•	•	•	G	A	•	•	C	C	T	G	•	C	•	•	•	•	G	•	G
I 2	•	•	•	•	•	T	A	•	•	•	•	•	T	•	•	•	•	T	•	•	•	•	A	T	•	C	•	T	G	•	C	•	G	•	•	G	•	G
R10, L3, L25	•	•	•	•	•	T	A	•	•	•	•	•	T	•	•	•	•	T	•	•	•	•	A	•	•	C	•	T	G	•	C	•	G	•	•	G	•	G
J6	•	C	•	•	•	T	•	•	•	•	C	•	T	•	•	•	•	T	•	•	•	•	A	•	•	C	•	T	G	•	C	•	G	•	•	G	•	G
J7	•	•	•	•	•	T	•	•	•	•	C	•	T	•	•	•	•	T	•	•	•	•	A	•	•	C	C	•	G	•	C	•	G	•	•	G	•	G
L29	A	•	•	•	•	T	•	•	•	G	•	•	T	C	•	•	•	T	•	•	•	•	A	•	•	C	•	T	G	•	C	•	G	•	•	G	•	G
G7	•	•	•	•	•	T	•	•	•	•	•	•	T	•	•	T	•	T	•	•	•	•	A	•	•	C	•	T	G	•	C	•	G	•	•	G	•	G
P6	•	•	•	C	•	T	•	C	•	•	•	•	T	•	•	•	•	T	•	•	C	•	A	•	•	C	C	T	G	•	C	•	G	•	•	G	•	G
Y24	•	•	•	•	•	T	•	•	•	•	•	•	T	•	•	•	T	T	•	•	C	•	A	•	•	C	C	T	G	•	C	A	•	•	•	G	•	G
L7	•	•	C	•	•	T	•	•	•	•	•	•	T	•	•	•	•	T	•	•	•	•	A	•	•	C	C	T	G	•	C	•	G	•	•	G	•	G
J1, J12	•	•	•	•	C	T	•	•	•	•	•	•	T	•	•	•	•	T	•	•	•	•	A	•	•	C	C	T	G	•	C	•	•	•	•	G	G	G
NWF	•	•	•	•	•	T	•	•	•	•	•	•	T	•	•	•	•	T	•	•	•	•	A	•	•	C	•	•	G	•	C	•	G	•	•	G	•	G

Table 6. Sequence mutations used to construct the median network (16090-16365)
NWF = New World Founder

Sample	16092	16093	16094	16129	16136	16157	16175	16181	16183	16189	16215	16234	16249	16259	16266	16286	16293	16300	16311	16312	16354	16360
CRS	T	T	T	G	T	T	A	A	A	T	A	C	T	C	C	C	A	A	T	A	C	T
D2	•	•	•	•	•	•	G	•	•	•	•	•	•	T	•	•	•	G	•	G	•	•
E2	•	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
G3, R12, R15	•	•	•	•	•	•	•	•	•	•	•	•	•	•	T	•	G	•	•	•	•	•
L22, L24	•	•	•	•	•	•	G	•	•	•	•	•	•	T	•	•		G	•	•	•	•
R6, AB174932	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	C
P9, L23	•	•	•	•	•	•	•	G	•	•	G	•	•	•	•	•	•	•	•	•	•	T
M5, Jl, J12	•	•	C	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	G	•	•
I 2, L3, R10, L25	•	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	T	•
J6, J7	•	•	•	•	•	•	•	•	•	C	•	•	•	•	•	•	•	•	•	•	•	•
L29	•	•	•	•	•	•	•	•	•	•	•	•	C	•	•	•	•	•	•	•	•	•
G7	•	•	•	•	•	•	•	•	•	•	•	•	•	•	T	•	•	•	•	•	•	•
P6	•	C	•	•	C	•	•	•	•	•	•	•	•	•	•	•	•	•	C	•	•	•
Y24	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	T	•	•	C	•	•	•
L7, AB174964	C	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
NWF, AB174923, AB174936	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
AB174932	•	•	•	•	•	•	•	•	•	•	•	•	•	T	•	•	•	•	•	•	•	•
AB174914	•	•	•	•	•	C	•	•	C	C	•	G	•	•	•	•	•	•	•	•	•	•
AB174910, AB174911	•	•	•	•	•	•	•	G	•	•	G	•	C	•	•	•	•	•	•	•	•	•

Figure 4. Median network constructed using Puerto Rican HVR-I and HVR-II sequences from Haplogroup A (Constructed by Dr. J.C. Martínez- Cruzado)
*np mutations: 16111, 16223, 16290, 16319, 16362, 73, 146, 153, 235, 263