Genetic variation in the African mitochondrial haplogroup L3e in Puerto Rico

By

Luciann Bracero Quiñones

Thesis submitted in fulfillment for the degree of

MASTER IN SCIENCES

in

Biology

UNIVERSITY OF PUERTO RICO

MAYAGÜEZ CAMPUS
May 2012

Approved by:

_____	_____
Fernando J. Bird Picó, Ph.D. Member, Graduate Committee	Date
_____	_____
Taras Oleksyk, Ph.D. Member, Graduate Committee	Date
_____	_____
Juan C. Martínez-Cruzado, Ph.D. President, Graduate Committee	Date
_____	_____
Abigail Matos-Pagán Representative of Graduate Studies	Date
_____	_____
Nanette Diffoot-Carlo, Ph.D. Chairperson of the Department	Date

Abstract

This study aims to clarify the uncertainties of the African maternal origin of the Puerto Rico modern population that occurred during the slave trade. Haplogroup L3e was chosen for this study because it combines the properties of being one of the most frequent Sub-Saharan haplogroup (17%) within our Puerto Rican present population at the same time that it presents distinct sub-haplogroups with uneven distributions in the African continent. The mitochondrial DNA haplogroup L3e is characterized by the gain of the restriction site 2349 *MboI* or *DpnII* within the Afro-Eurasian paragroup L3 (-3592 *HpaI*). Restriction site polymorphisms and Hyper Variable Sequence-I (HVS-I) screening in haplogroup L3e define five principal sub-haplogroups (L3e1, L3e2, L3e3, L3e4 and L3e5). Thirty-five L3e samples were classified into L3e sub-haplogroups. Because Puerto Rico African heredity suggests a mainly West African origin, we expected that the most frequent sub-haplogroup would be L3e2, which is the most frequent in Western Africa. However, the most frequent L3e mitochondrial DNAs in Puerto Rico belongs to a Restriction Fragment Length Polymorphism (RFLP) type shared by L3e1 and L3e2b. Sequencing showed a higher frequency of L3e1 among our present population, leaving L3e2 as second most frequent and L3e4 and L3e3 as the least frequent sub-haplogroups. Principal Component Analysis and an equation based on a Bayesian approach of the sub-haplogroup frequency distribution suggest that the L3e origin is more similar to Mozambique than to any other African region. The study of the sub-haplogroups is important because these have not been characterized for Puerto Rico and the information provided by them may shed light on the geographic origins of African slaves sent to Puerto Rico from the 16^{th} to 19^{th} century.

Resumen

Este estudio tiene como objetivo clarificar el origen Africano de la herencia materna en la población actual de Puerto Rico como resultado del tráfico de esclavos. El haplogrupo L3e se escogió para este estudio porque combina las siguientes características: es el más común haplogrupo del Sub-Sahara (17%) entre la población presente de Puerto Rico y a la misma vez presenta una distribución distintiva de sus sub-haplgrupos entre las regiones de África continental. El haplogrupo de ADN mitocondrial L3e está caracterizado por la ganancia del sitio de restricción de la enzima *MboI* o *DpnII* en la posición 2349 dentro del paragrupo L3 (-3592 *HpaI*). Los polimorfismos de los sitios de restricción y haplotipos según el HVS-I dentro del haplogrupo L3e definen cinco sub-haplogrupos principales (L3e1, L3e2, L3e3, L3e4 y L3e5). Treinta y cinco muestras L3e fueron caracterizadas para sus sub-haplogrupos. Ya que la herencia Africana en Puerto Rico sugiere ser mayormente del Oeste de África, nosotros esperábamos que el haplogrupo L3e2 fuera el de mayor frecuencia, el cuál es el más común en el Oeste de África. Sin embargo, los mtDNAs de L3e más frecuentes en Puerto Rico pertenecen a un grupo RFLP compartido por los sub-haplogrupos L3e1 y L3e2b. Usando análisis de secuencias se llegó al resultado de que el sub-haplogrupo L3e1 fue el más frecuente en nuestra población actual, dejando L3e2 como el segundo más frecuente, L3e4 como el menos frecuente y L3e3 con tan solo una muestra. El análisis de componentes principales y la ecuación bayesiana de los sub-haplogrupos L3e sugieren que su origen es más similar a Mozambique que ninguna otra región Africana. Este estudio de los sub-haplogrupos es importante porque estos no han sido estudiados para Puerto Rico y su

estudio puede aclarar el origen demográfico del tráfico de esclavos hacia Puerto Rico durante los siglos 16 al 19.

Acknowledgments

I want to express my gratefulness first of all to God who has given me the strength and without Him I am nothing. Also, I am grateful to those who guided me through this process of writing and correcting my Masters of Science thesis. First of all, Dr. Juan C. Martínez-Cruzado to give me the opportunity to work in your laboratory, believe in me to make this investigation possible and also his support and dedication towards the academic excellence. Thank you for giving me a second chance to finish my thesis. I want to thank my faculty committee, Dr. Fernando J. Bird Pico and Dr. Taras Oleksyk, for all the help and suggestions. Also, Dr. H. J. Bandelt, who kept communication with me and gave advices to the course of my investigation and finally, I will like to thank my family for all the moral support they have provided. This success I owe to all of you!

Table of Contents

Tables and Figures

INTRODUCTION

The first African slave was brought to Puerto Rico by Juan Ponce de León in 1508. According to historical sources and through almost four centuries, most of these slaves were obtained from the west coast of Africa from Senegal to Angola, although many were captured from tribes residing very much inside the continent, and by the end of this period a strong trade was established from Mozambique. However, historical records are deficient and the contribution of the various African regions to the Puerto Rican gene pool from mtDNA is very useful to shed light on the uncertainties.

The mtDNA is a very highly used molecule in phylogeographic studies for its unique and very desirable qualities: circular DNA molecule, relatively small, maternal heritage only, does not recombine and with a high mutation rate (Cann et al., 1987). The use of specific markers found in the mtDNA can clarify the phylogeographic history of the Puerto Rican African gene pool. Martínez-Cruzado et al. (2005) suggested in their study of the phylogenetic history of the Puerto Rican population, that the African contribution to Puerto Rico tends to be mostly from West Africa. Also, they calculated that 27.2% of the mtDNA in the present Puerto Rico population is of Sub-Saharan African heritage. Other studies suggested that sub-haplogroup L3e2, a mainly West African sub-haplogroup, is predominant among the African Americans of the Caribbean (Salas et al., 2002). For these studies they used principal component analyses (PC) based on HVS-I sequences in these populations, and suggested that African American populations were more similar to those of West Africa than to those of other regions of the continent. On the other hand, studies using frequencies of different mtDNA sub-haplogroups distinctive of African populations, and analytical tools often based on

Bayesian approaches have been able to estimate the contribution of African ancestors to admixed populations (Mendizabal et al., 2008).

JUSTIFICATION

The aim of this study is to better understand and explain the origins of our African heritage. The L3e samples represent 4% of the Puerto Rican mtDNA and 17% of that of Sub-Saharan mtDNA origin. Among all African haplogroups, L3e is particularly ubiquitous. Thus, information related to geographic origin can be obtained from this haplogroup through its subdivisions. Its sub-haplogroups are L3e1, L3e2, L3e3, L3e4 and L3e5 (Behar et al., 2008). The F_{ST} value for L3e sub-haplogroups in Africa is 0.17. By contrast, most F_{ST} values for intercontinental comparisons fluctuate between 0.09 and 0.13. Therefore, being able to use these sub-haplogroups to determine our historic African demographic process during the slave trade, will help to understand better the contribution of the African slaves to our present Puerto Rican population. Thus, the objective of this study is to reach a more comprehensive estimate of the contribution of different African regions to the Puerto Rican mitochondrial gene pool by means of studying L3e sub-haplogroups frequency distribution in Puerto Rico. I have analyzed the L3e sub-haplogroups within the Puerto Rican population using a four enzyme polymorphism assay comprising the following enzymes and sites: *Ava*II (5260), *Alu*I (5584), *Hae*III (9553) and the isochizomer of *Mbo*I, *Dpn*II (14869) (Table 1). In addition, the HVS-I sequence of twenty-four L3e samples will be determined. The frequency variation of the sub-haplogroups within and among populations in Africa, using the F_{ST} overall, have helped us differentiate African samples variation distributed within *versus* among the populations studied. Then, the frequency of each of these sub-haplogroups have been used in a Principal Component Analysis (PCA) to study genetic similarity among populations. This helped us compare the distribution of our present African gene pool to the African continental regions. Finally, by using a Bayesian phylogenetic analysis of L3e we have

estimated the relative contribution of different African regions to the Puerto Rican gene pool.

LITERATURE REVIEW

Slave trade to the Americas

The Puerto Rico African heritage finds its origin in the slave trade. At least 11 million people, in their majority from the regions of West and Central Africa were forcibly moved to European and American destinations during the 15th-19th centuries (Thomas, 1997). The imported slaves that came to the Island were from the West coast of Africa, from Senegambia to Cameroon, but also from Congo, Angola, and even Mozambique (Alvarez-Nazario, 1979). Extensive trading from Sub-Saharan Africa started with the Portuguese trading slaves around 1502 (Thomas, 1997). Although Europeans did not usually adventure inside Africa, the trade included Central African people that were captured often by African slave merchants and then stored as merchandise in the west coast of Africa including the islands of Bioko and Sao Tomé before departing to the Americas. Once slaves arrived to America they interbred with Amerindians and Europeans (González et al., 2006). There are little or no records of dates, births and deaths nor geographical information on the origin of African slaves. Therefore, the use of molecular genetics, like the mitochondrial DNA, will help us understand better the geographic sources of the slave trade in relationship to our present African American heritage in the current population of the Island.

Historical records and documentation during that time do not help us to establish the geographic origin of African Americans. Based on PC analyses of haplogroup frequencies in Puerto Rico and tribes around Africa, Martínez-Cruzado et al. (2005) suggested that the origins of African mtDNAs in Puerto Rico were mainly from Western Africa. Also the data presented by Salas et al. (2004), through their PC analysis, showed that the African American, Sao Tomé, and Caribbean samples were closer to Western

Africans. However, it is now known that one of the main haplogroups used by Martínez-Cruzado et al. (2005), L3e, is subdivided into five haplogroups that have distinct geographic distribution within Africa (Bandelt et al., 2001; Cerny et al., 2007). Therefore, I am extending this work by identifying further Puerto Rican mtDNAs of haplogroup L3e in search of a more precise estimate of African origins of Puerto Rico.

Mitochondrial DNA analysis

The mitochondrial DNA (mtDNA) has proved to be an excellent tool in the study of human population history, frequently providing geographic-specific mutations and informative data to figure out historic demographic processes. The mtDNA is a circular chromosome that contains 37 genes in 16,568 base pairs (Andrews et al., 1999). These genes codify for proteins involved in oxidative phosphorylation, and for mitochondrial rRNAs and tRNAs. The endosymbiotic theory established that this organelle arose from a prokaryote that lived inside a primitive eukaryote. This theory has being reviewed by many authors (Lang et al., 2009).

The mitochondrion is inherited from the mother only (Giles et al., 1980). It does not recombine, and thus maintains its human group identity in mixed populations such as that of Puerto Rico. In addition, its rapid mutation rate allows the accumulation of sufficient mutations to perform statistical analyses with sufficient power in the study of recent populations such as humans without requiring the analysis of a prohibitive number of samples.

For evolutionary studies researchers often utilize the control region of the mtDNA. The control region is a non-coding region that contains two hypervariable

regions known as HVS-I and HVS-II, which span from positions 16051 to 16400 and 68 to 263, respectively (Soares et al., 2009). These regions are the most variable, and thus the most effective at providing information. They represent only 7% of the mtDNA, but accumulate approximately 30% of all mutations. The variability of the mutation rate at different sites in these regions is very high, often causing parallel and reverse mutations in different branches of the evolutionary tree. These facts create a problem of large standard errors when constructing phylogenetic relationships based on genetic distances (Richards et al., 2002).

Restriction fragment length polymorphisms (RFLPs) studies are efficient to identify the different lineages because specific restriction markers have been identified for most haplogroups, which are monophyletic groups typically defined by such haplogroup's specific restriction markers. For example, the African haplogroup L0a is defined by a 9bp deletion between the genes for the second subunit of the cytochrome oxidase and the $tRNA^{Lys}$ (Salas et al., 2002). However, it is difficult to estimate the age of these haplogroups, not only because of the variability in the mutation rate among different sites, but because informative sites even for the complete mtDNA are few. Such age estimates are based on the diversity found within haplogroups. Various authors have agreed that a large sample size can better resolve age estimates (Richards et al., 1996, 2002; Saillard et al., 2000).

Other studies have relied on the comparison of mitochondrial genomes deposited in genetics databases such as GenBank. Pereira et al. (2009) analyzed the complete sequence of 5140 complete human mtDNAs and concluded that protein-coding genes have less variation. Analyses of mtDNA coding region and HVS-I and HVS-II sequences have provided us significant information to determine phylogeography of different

mtDNA haplogroups (Bandelt et al., 2001). On the other hand, it has also left us with a lot of controversies about evolutionary theories as the one discussed by Howell et al. (2004). In this study of haplogroup L sequences, the non clock-like evolution of haplogroup L2 was described. Their analysis of maximum likelihood branch lengths found that its lineages evolve at different rates. However, L1 and L3 sequences showed clock-like evolution. In order to estimate more precisely the time frame of phylogeographic changes using the molecular clock of the human mtDNA, Soares et al. (2009) used a calibration method that corrects the time dependency of the mtDNA mutation rate by estimating empirically the level of selection versus variation at any time depth within the phylogenetic tree and using this to correct the estimates from the linear clock. They concluded that human mtDNA mutations do not behave linearly, most likely because of purifying selection and saturation and that assuming a linear approach will result in an overestimation of coalescence times near the tips of phylogenetic trees.

Phylogeographic patterns of African heritage

Genetics, linguistics and anthropology have been combined to compensate for the lack of records that were kept from the centuries of slave trade (Shiver & Kittles, 2004). The study of linguistics has been key in the study of African populations. There are four main macro-families: the Niger-Kordofanian, Afro-Asiatic, Nilo-Saharan and the Khoisan (Tishkoff et al., 2009). The Niger-Kordofanian is spread across a broad portion of Africa and it includes the Bantu family with more than 500 languages and the greatest geographic distribution in Africa. Afro-Asiatic languages are spoken across North Africa, the Horn of Africa and the Middle East; Nilo-Saharan is spoken in Sudanic, Sahara, and eastern Africa, and the Khoisan family covers 30 languages spoken by about 300,000-

400,000 people most of them in the South of Africa (Tishkoff et al., 2009). The Khoisan languages, which contain click-consonants, are spoken by San and Hadza of Tanzania according to Tishkoff et al. (2007 and 2009). Also, the Khoisan languages are the most ancient, but the Bantu expansion seems to have replaced Khoisan-speaking populations in most of Sub-Saharan Africa. The Bantu expansion originated in eastern Nigeria and western Cameroon (the core of the dispersal) (Salas et al., 2002). From there, it spread east and south to cover all of Central Africa and reaching South Africa regions, where Khoisan is still spoken in addition to very small clusters that are left in Central Africa (Quintana-Murci et al., 2007). The controversy about the linguistic studies is how this dispersal took place; if it was a slow exodus of small groups or a rapid population movement.

According to Salas et al. (2002), because the population is highly homogenous with respect to their mtDNA, analysis of PCA and AMOVA suggest a common origin or high levels of gene flow between the groups, or both. They also state that because of pronounced founder effects in the networks, substantial numbers of complex processes of the Bantu expansion involved events that resulted in short-range dispersals.

The African markers identified: L0, L1, L2 and L3, start to clarify the Bantu expansions and also the Atlantic slave trade migrations. L0 is at the root of Sub-Saharan mtDNAs. It is approximately 150,000-170,000 years old (Salas et al., 2002). Starting with L0a (defined by a 9-bp deletion), this haplogroup has been identified as an eastern Bantu marker (Mishmar et al., 2003). Only two major sequences characterized L0a2, the distribution of which is mainly Central Africa. It is common in African Americans, suggesting that this haplogroup was a significant part of the Atlantic slave trade.

The L1 clade is represented by L1b found mainly in Western Africa, where it seems to have originated. L1c seems represented by carriers from Angola and its clade L1c3 in Bioko and Mozambique. Lastly, L1e is entirely of Central African provenance (Salas et al., 2002).

The haplogroup L2 is divided into L2a to L2e. Of these, L2a is the most frequent and widespread in Africa and among African Americans (Salas et al., 2002). It is characterized by its transition at nucleotide position (np) 16294 and the presence of a *Hae*III site at np 13803. The estimated age of L2 is 70,000 years.

L3 includes all Eurasian variation and is an important marker of Sub-Saharan African distribution (Salas et al., 2002). The L3 estimated origin is East African. It includes sister haplogroups L3b and L3d, which are predominantly West African. But L3e is the most widespread, frequent and ancient of all L3 clades (Salas et al., 2002). L3e originated in central or eastern Africa and according to Soares et al. (2009), its coalescent time is estimated at ~65,000 years before present.

L3f and L3g are rare and virtually restricted to East Africa. It seems that L3f1 have spread at an early date into West Africa. Meanwhile, L3g is from Tanzania but included only a small sample (Salas et al., 2002). In Puerto Rico, Viera-Vera (2006) studied all Sub-Saharan haplogroups finding L3e, L1b, and L2a as the most frequent in a sample set representative of the Puerto Rican present population.

L3e sub-haplogroup analysis

Phylogeographic analysis of mtDNA haplogroups define the composition of the diverse groups that are part of modern African populations, which can be classified into linguistic categories (explained in the previous section). Identification of RFLPs in the

coding region of the mtDNA and control region sequence haplotypes has facilitated comparisons between African populations and American populations of African origin. Bandelt et al. (2001) recommend that a combination of multiple mtDNA analyses including RFLP and HVS-I and HVS-II information should be used as a more comprehensive way to classify these haplogroups more precisely and better determine the geographic region that they represent.

Bandelt et al. (2001) studied eight RFLPs sites for L3e: 5260 *Ava*II, 5584 *Alu*I, 9253 *Hae*III, 9553 *Hae*III, 13100 *Msp*I, 13803 *Alu*I, 14869 *Mbo*I and 15812 *Rsa*I. They divided L3e into 4 sub-haplogroups: L3e1, L3e2, L3e3, L3e4 and sometimes into further sub-haplogroups. The loss of *Mbo*I at np 14869 was unique for L3e2a. The gain of *Ava*I restriction site at 5260 separated L3e3 and L3e4 from the rest of the sub-haplogroups, plus the loss of *Alu*I restriction site at np 5584 was distinct for L3e4. Finally, the loss of the *Hae*III restriction site at np 9553 was only found in L3e3. The only two groups that this RFLP analysis could not separate were L3e1 and its sub-haplogroup L3e1a from L3e2 and its sub-haplogroup L3e2b. Due to that reason, we are using HVS-I haplotypes to differentiate among samples that RFLPs fails to identify.

Using HVS-I sequencing the groups were divided into seven categories: L3e1 (characterized by a transition at np 16327), L3e1a (16185), L3e2 (16320), L3e2a (16320 and 16286), L3e2b (16172, 16189 and 16320), L3e3 (16265T), and L3e4 (16264). Also, more recently, L3e5 has been described by Cerny et al. (2007). It is distributed across North Africa and it is estimated to be 7100 years old. It is absent in Sub-Saharan African and African-American samples, which indicates very low contribution to the Atlantic Slave trade. Therefore, it is not included in this study for further analysis. This sub-haplogroup is characterized by two transitions in the HVS-I at np 16041 and 16223

(Torroni et al., 2006). Finally, Bandelt et al. (2001) used HVS-II sequences to confirm the HVS-I categories.

In this study I have assessed the L3e samples for identification of the four sub-haplogroups: L3e1, L3e2, L3e3 and L3e4 using RFLP for the distinctive np 5260 *Ava*II, 5584 *Alu*I, 9553 *Hae*III and the isochizomer of *Mbo*I, *Dpn*II (14869) (Table 1). I have also sequenced HVS-I, and for confirmation, some samples of the HVS-II were sequenced, as well.

The previous studies on African populations used for comparison in this study are the following (Table2): from West Africa, including Senegambia (Bandelt et al., 2001; Hünemeier et al., 2007; Chen et al., 2000), Senegal (Batini et al. 2007), Mali and Mauritana (Gonzalez et al., 2006; Ely et al. 2002); from West Central Africa, including Cameroon, Nigeria, Mbuti Pigmies and Biaka Pygmies (Bandelt et al., 2001; Silva et al. 2006; Chen et al. 2000; Veeramah et al. 2010); Sao Tomé (Mateu et al., 1997; Batini et al., 2007); Mozambique (Pereira et al., 2001; Salas et al., 2002), and Angola (Plaza et al., 2004). L3e1, L3e2 and L3e3 are very common among southeastern Bantu speakers. L3e1 is established throughout Sub-Saharan Africa and Salas et al. (2002) suggested a West Central African origin, it is rare among West Africans from this sub-haplogroup are derived two sub-clades: L3e1a and L3e1b. L3e1a is the result of Bantu family gene flow and has a high frequency among Khwe populations. Its distribution is southeastern Africa; meanwhile L3e1b is extensively widespread and its corresponding HVS-I sequence motifs is 2223-325D-327 (Salas et al., 2004). L3e2 is the most common sub-haplogroup in Central Africa and West Africa. The *Mbo*I site at np 14869 seems to partition L3e2 into two sub-clades: the loss characterizes L3e2a, while the complementary clade, L3e2b, appears to be defined by transitions at np 16172 and 16189

in its HVS-I haplotype. The sub-clade L3e2a is found mainly in Central Africa, and has a particular transition in HVS-II at np 198 (Salas et al., 2002). On the other hand, the sub-clade L3e2b is found at a high frequency in West-Central African regions, and at small frequencies in southeastern Africa, although it is widespread through the African continent and also is found in high frequency in a Caribbean sample (Santo Domingo) and Brazilian samples (Bandelt et al., 2001). L3e3 is primarily Angolan, but it has a high frequency in Southeast Africa. L3e3 is found in Brazil, West-Central African regions and North Eastern Africa. L3e4 is a small clade present in East, and Central Africa, with only no samples found so far in Angola. However, it accounts for approximately 20% of all L3e mtDNAs in West Africa.

We focused on the Puerto Rican L3e geographic origin from Africa (Figure 1). For this study I divided Africa into the following regions and populations from previous studies (Table 2): West Africa, including Bantu-speaking, from Senegal, Senegambia, Mandenka, Wolof, Mali and Mauritana; West Central Africa including samples that are from Mbuti Pygmies, Biaka Pygmies, Nigeria, Cameroon and other Non-Atlantic West Central Africa; Sao Tomé, Angola and Mozambique. We compared the frequency of the sub-haplogroup L3e2 in Puerto Rico, because it is found in high frequency among other Caribbean populations (Salas et al., 2002).

Region Profiles

L3e mtDNA sub-haplogroups have been studied in depth in the literature, to the extent that variation shows strong phylogeographic structure among African continental regions (Bandelt et al., 2001; Salas et al., 2002; Soares et al., 2009). The use of L3e sub-haplogroups can explain the migration events that occurred during the slave trade from

Africa to America. Therefore, using phylogeographic structure analysis of these sub-haplogroups we can better understand the specific events that occurred during this time, to the point of being able to assign continent origin of our present African mtDNA heritage in Puerto Rico.

The use of principal component analysis (PCA) in our study can help us achieve that objective. PCA gives us the opportunity to compare the L3e sub-haplogroup distribution in the modern Puerto Rican population with that of various African populations, suggesting the populations that most contributed to the slave trade to Puerto Rico. PCA is a fast and effective way to diagnose population structure (Zhu and Yu et al. 2009). PCA identifies polymorphisms representing strong variations and cluster samples inside stable haplogroup clusters (Alexe et al., 2008). Moreover, PCA summarizes variation observed across all markers into a number of underlying component variables. It organizes individual coordinates along axes of variation, describes the weight of each sub-haplogroup in each individual component and is able to make computational analyses of a large number of markers. On the other hand, it needs large sample sizes to capture most of the variation and it has a very low dimensional manifold (Zhu and Yu 2009).

Another approach employed that can be useful for phylogeographic structure analysis was used by Mendizabal et al. (2008). In his study of Cuban genetic admixture, they had the objective to determine the contribution of different continents to the Cuban mtDNA and Y chromosome gene pool. For this, they devised a Bayesian approach-based equation. The advantage of using the Bayesian computation is that it will give us a relative numeric estimate of the contributions of different African continental regions to Puerto Rico, based on the L3e sub-haplogroup frequencies in Puerto Rico and in African regions. It has a disadvantage, in which it can overestimate the contribution of regions of

very low contribution, and similarly underestimate the contribution of regions with very high contributions.

Finally, I will use the Overall F_{ST} value from the PCA calculations to estimate the level of variation between L3e African populations relative to the variation within populations. A very high F_{ST}, suggesting large differences between populations, will suggest that this approach should be effective in distinguishing the possible African origins of the L3e mtDNAs in Puerto Rico.

OBJECTIVE

Considering the advantages and disadvantages discussed above, I am confident that by using these methods the following questions about the events that took place during the Atlantic Slave trade can be answered: what is the origin of the L3e sub-haplogroup in Puerto Rico and which combination of contributions of different African regions to Puerto Rico may such geographic origin? Thus, the objective of this study is to reach a more comprehensive estimate of the contribution of different African regions to the Puerto Rican mitochondrial gene pool by means of studying L3e sub-haplogroups frequency distribution in Puerto Rico.

MATERIALS AND METHODS

Population

Two-hundred and twenty Puerto Rican mtDNA individual samples were characterized for an African origin (Martínez-Cruzado et al., 2005). Restriction site assays provided 37 of these samples as belonging to mtDNA haplogroup L3e, defined by the gain of the restriction site 2349 np *MboI* within Afro-Eurasian super haplogroup L3 (-3592 *HpaI*). The 37 samples, together with other 3 L3e samples collected later in Vieques and directly sequenced by a group of undergraduate students under Dr. Martínez Cruzado, were further analyzed in this study (Table 3).

RFLP screening

We employed the classification of the L3e sub-haplogroups established in Bandelt et al. (2001), which combines HVS-I and RFLP information. Four enzyme restriction site assays were performed using the following enzymes and sites: *AvaII* (5260), *AluI* (5584), *HaeIII* (9253) and the isochizomer of *MboI*, *DpnII* (14869) (Table1). The primers and amplification reaction conditions used by this study were described by Bandelt et al. (2001) (Table 4).

The restriction digests were performed according to the manufacturer's conditions, and the DNA fractioned in a 3% agarose gel with PhiX174DNA-*Hae III* Digest (New England Bio Labs N3026L) as molecular weight marker. The agarose gels were stained with ethidium bromide and photo-documented under UV light.

HVS-I screening

The HVS-I region was amplified through PCR using primers L15829 (CATCCGTACTATACTTCACAAC) and H16345 (GGGACGAGAAGGGATTTGAC), also L16219 (TGCTTACAAGCAAGTACAGCA) and H34 (ACCAAATGCATGGAGAGCTCC). The PCR reaction contained 1X Buffer without $MgCl_2$ (Sigma P-2317), 2mM $MgCl_2$, 4 mM dNTP, 1.3µM each primer, 8uL of DNA sample from hair root extract (Martínez-Cruzado et al., 2005), 0.45 U of *Taq* DNA polymerase (Sigma), 0.03 mg BSA (New England Bio Labs) in a total volume of 25 µL. The amplification conditions included one cycle of 5 minutes at 94°C followed by 35 cycles of 30 seconds at 94°C, one minute at 55°C, and 1.5 minutes at 72°C. One extension cycle of 10 minutes at 72°C completed the amplification. After confirming appropriate amplification through agarose gel electrophoresis, the amplicons were purified using the High Pure PCR Product Purification Kit (Roche Diagnostics Corp.) as instructed by the manufacturer.

Cycle sequencing was performed using the Big Dye 3.1 Terminator Kit (Applied Biosystems) as instructed by the manufacturer. Reaction conditions were as follows: 0.5µl Buffer 5x, 0.5µl Big Dye, primer at 1µM and 3 µl of 5ng/µL of Purified PCR product in a total volume of 5 µl. Amplification conditions were 96º C for one minute, 35 cycles of 96ºC for 15 seconds, 50ºC for 15 seconds and 60ºC for 4 minutes. Finally, samples were kept at 4ºC. Primers used for sequencing were L15854 (CCTAATCCTAATACCAACTATC) and H16526

(GGGAACGTGTGGGCTATTTAGG). Resulting fragments were purified through ethanol precipitation. Automated sequencing was performed using an ABI 3130 Genetic Analyzer. Obtained sequences were aligned to the revised CRS (Andrews et al., 1999) using the OMIGA 2.0 (GCG) program.

Phylogeographic analysis

The L3e sub-haplogroups were analyzed for phylogeographic specificity using only sequences with confirmed L3e status as described before by Bandelt et al. (2001) from previously published data (see Table 2). The African geographical regions were West Africa (Senegambia, Senegal, Mandenka, Wolof, Mali and Mauritana), West-Central Africa (Mbuti and Biaka Pygmies, Nigeria, Cameroon and other Non-Atlantic West Central samples), Angola, Sao Tomé, and Mozambique.

The resulting L3e sub-haplogroup frequencies were compared with those of other populations in the African regions. Using F_{ST} overall measurements we describe the distribution of the variation within and between populations in Africa. F_{ST} and Principal Component Analyses (PCA) were calculated using the POPSTR program of Henry Harpending (University of Utah).) The probability of the L3e origin from each African sub-continental region was estimated using an equation based on a Bayesian approach, previously explained by Mendizabal et al. (2008).

RESULTS

RFLP

In our RFLP, we used four enzymes that let us differentiate three L3e sub-haplogroups: L3e2, L3e3 and L3e4. The only ambiguity of this test was the incapacity to differentiate between L3e1 and L3e2b. Of the 40 samples, three never amplified and three others (from Vieques) were sequenced directly and never subjected to RFLP. The RFLP analysis let us identify 20 of the remaining 34 samples that amplified, about 54% of the samples. Table 5 contains all the samples and their resulting RFLP data. Each sample is codified by letters as explained above (see Table 3). Fourteen samples were classified as L3e1 or L3e2b (38%), four as L3e4 (11%), and only two samples were L3e2a (5%). Samples for which not enough data was collected from the RFLP assay (14 samples, 38%) were classified as not determined (ND).

HVS-I screening

Using 22 sequences of the hypervariable region, we were able to confirm and clarify many ambiguities of the RFLP. Eleven of them belonged to the group of 14 for which not enough RFLP data had been collected and had thus been classified as ND. Of these, eight belonged to the sub-haplogroup L3e1, one to L3e2a, one to L3e3, and one to L3e4. In addition, we were able to sequence six of the 14 mtDNA samples that were previously classified as to either L3e1 or L3e2b. Of these, three were confirmed to be L3e1 and three L3e2b. We were also able to sequence one of the samples that we had not been able to amplify for RFLP (FE2) and classified to L3e1. Finally, we sequenced four samples for confirmatory purposes. Two belonged to L3e2a and two to L3e4, and all

were confirmed correspondingly. Table 6 contains the 22 HVS-I sequences, listed as a motif relative to the CRS.

Adding to all this information were the samples from Vieques, two of which were shown to belong to L3e2b and one to L3e4, we finished with 12 L3e1, three L3e2a, five L3e2b, one L3e3, and six L3e4. These are summarized in Table 7. In addition, we had eight samples that we knew from the RFLP analysis that they belonged to either L3e1 or L3e2. Because ignoring those samples would bias our results toward an underestimation of L3e1 and L3e2 in Puerto Rico, we used the ratio of L3e1:L3e2a already obtained (12:5) to predict that the 8 samples would be distributed into six belonging to L3e1 and two to L3e2. Therefore, the total of L3e1 and L3e2 samples were raised to 18 and 10, respectively.

In conclusion, the distribution of L3e sub-haplogroups in the Puerto Rico present African derived population is as follows: 51.4% (SD: 4.5%) to L3e1, 28.6% (SD: 2.8%) to L3e2, 17.1% (SD: 2.9%) to L3e4 and 3.7% (SD: 0.5%) to L3e3.

Phylogeographic analysis

The F_{ST} value for our African populations overall is 0.17, which represents a significant difference among populations. This difference should allow us to distinguish among African populations using L3e sub-haplogroups (Table 2). Puerto Rican L3e distribution has more L3e1 and is similar to Sao Tomé, Mozambique and Angola (Figure 1), followed by L3e2 which is more frequent in the West and West-Central Africa. There is also a high frequency of L3e4 which again confirms the West African origin.

Table 8 pools the samples used from previous studies (Table 2) into African geographic regions, and it includes the results of this study. These data were used to

make the statistical analyses for Principal Component Analysis (PCA), the Bayesian approach equation, and to calculate F_{ST}.

Using the analysis of Principal Components (PC) we compared the frequency among the L3e sub-haplogroups identified in the representative sample of our present Puerto Rican population. The first PC captured 78% of the variance and the second one 18%, which placed Puerto Rico closer to Mozambique than to any other region of Africa (Figure 3A). The PC1 vs PC3 captured 82% of the variance and it also grouped closer Puerto Rico and Mozambique (Figure 3B). The sub-haplogroups that explain this PCA are L3e1 and L3e2 (Figure 3C). The next analysis included Sao Tomé, an island that became inhabited only after the start of the slave trade, when it was used to store slaves before their long journey to the Americas. The analysis resulted with PC1 capturing 76% of the variation, PC2 20% and PC3 4% (Figure 4A). Therefore, PC1 vs. PC2 (Figure 4A) represents 96% of the variance and PC1 vs. PC3 80% (Figure 3B). The comparison of PC1 vs. PC2 places Puerto Rico almost equidistant between Sao Tomé and Mozambique, at the center of an apparent cluster. Figure 4C shows which haplogroups are mostly responsible for the variation present in PC1 and PC2. PC1 is determined mostly by L3e1 and L3e2. Therefore, PC1 separates better the Mozambique and West Africa populations, which are the ones clustered by their high frequency of one of these two sub-haplogroups and low frequency of the other. Puerto Rico is placed closer to Mozambique because of its higher L3e1 frequency.

We next explored the effects of adding all the African haplogroups to this analysis (Table 9). When this is done, the variation contained in the first PC lowers to 54% and in the second to 18% (Figure 5 A). Also, the F_{ST} value is lower, 0.031. On the other hand, it reveals a weaker association between the Puerto Rican African population and West

Africa than with other African continental region. As expected, in this analysis it was found close to Sao Tomé. Interestingly, Mozambique and Angola populations are similar to each other than to any other continental region. Most of the L3 haplogroups clustered near the center of the graph accounting for the central positions of West Central Africa and Sao Tomé (Figure 5B and 5C). L0a and L2a are contributing to Mozambique and Angola cluster, while Puerto Rico's position to the top and right of the graph seems to be determined mostly by its high frequency of L1b and L3d.

Finally, if we compare L3e distributions with the Africa sub-continental regions using the Bayesian approach equation we can infer the putative origin of these samples within the African continent. Thus, the putative origin of the Puerto Rican African L3e samples sequences could be distributed as follows: 29.14 % (SD: 7.7%) to Mozambique, 26.1% (SD: 7.4%) to West Africa, 23.1% (SD: 7.1%) to West Central Africa and 21.7% (SD: 6.7%) to Angola (Figure 2).

DISCUSSION

The overall F_{ST} for L3e is 0.17, allowing us to explore the phylogeographic origin of Puerto Rican L3 samples. L3e3 is represented with only one sample in Puerto Rico, and although, L3e2 and L3e4 are also present in the Puerto Rican population, in this study the highest in frequency is L3e1. It can be inferred that L3e Africans from the slave trade to Puerto Rico were concentrated in these groups and that this explains its distribution in Puerto Rico, where the majority of slaves were brought to San Juan first and then disseminated through the Island (Alvarez-Nazario, 1979). Therefore, we determined phylogeographic difference across African continent of these sub-haplogroups, just as it has been reported by other authors (González et al., 20063 and Veerahmah et al. 2010)

The presence of a higher proportion of L3e1 is present in other American populations that also experienced dramatic demographic shift during the African slave trade, similar to that in Brazil (Hünemeier et al., 2007). This should not be surprising, because Brazil's own slave trade companies had their biggest slave factories in Angola and Mozambique, which are the African regions with the highest L3e1 frequencies. Given the low number of samples and high standard deviations, my results are not inconsistent with previous studies that suggest a meaningful West African origin (Martínez-Cruzado et al. 2005). However, it suggests a stronger affinity to Mozambique than previous studies. According to historical records the slave trade took place mainly in West Africa and Central Africa, where slaves were taken from Senegal to Cameroon in the West, Angola in the South and Mozambique in the Southeast. Sao Tomé was an intermediate point during the slave trade and as expected it is closer in L3e frequency

with Puerto Rico. On the other hand, Sao Tomé was not included in the Bayesian approach equation because our interest was to determine the continental African origin.

Recent publications have raised the number of samples for which data is available, especially in West Central Africa, and it was therefore appropriate for us to reconstruct a new PCA plot using all haplogroups, together with the L3e sub-haplogroups that we found in Puerto Rico. Figure 3 shows that Puerto Rico is not clustering any more with West Africa and West Central Africa, although it is still closer to West Africa. According to PCA, Puerto Rican population L3e samples distribution is more similar to Mozambique and Sao Tomé than to any other geographic zones. Only in Mozambique and Angola L3e1 mtDNAs are more frequent than L3e2, and this is exactly what is found in Puerto Rico. This is critical because our Principal Component (PC) analysis comprises 96% of the variance (PC1 vs. PC2), and it is evident that most of the variation can be defined in terms of L3e1 and L3e2 frequencies (Figures 1C and 2C), although the expectations were that L3e2 should be the most frequent, according to Salas et al. (2001 and 2004), because of its highest frequency in the Caribbean regions explored by these authors, in Puerto Rico is second in frequency. Also, samples from Santo Domingo (Martínez-Cruzado, personal communication) that also show a far higher frequency of L3e1 than L3e2. Therefore, it is becoming evident that at least in those Spanish Antilles were the slave trade ended early (such as in Santo Domingo) or where it was weak and controlled by foreign companies (Puerto Rico), L3e1 outnumbers L3e2. By contrast, in Cuba, a survey covering all provinces identified 11 L3e2 mtDNAs and only two belonging to L3e1 (Mendizabal et al., 2008). This study shows that Puerto Rico's L3e haplogroup is closer to Mozambique (Figure 2A).

One important caveat to keep in mind is that the samples used from previous studies reflect small L3e numbers, and these might not be representative of such large populations. This is especially true for Angola, for which only 9 L3e mtDNAs have been published. With so few samples and its higher frequency of L3e1, its L3e profile is statistically undistinguishable from that of Mozambique. Furthermore, because of its small effective population size, mtDNA haplogroups are highly susceptible to drift and may show large frequency differences between closely related tribes.

It is known that the sample sets of the previous studies were not systematically selected to be representative of their respective populations. Moreover, the slave trade itself surely was highly biased toward the capture of certain peoples within their geographic zones. Therefore, only the use of very high numbers of samples obtained from various places within and across different regions of Africa could provide strong confidence to conclusions on African origins through mtDNA studies. Because the Puerto Rico sample set has been randomly and systematically selected to be representative of the Puerto Rico population, my main contribution to our knowledge on this subject is the partition of Puerto Rico's L3e haplogroup into its sub-haplogroups. As further data is obtained, especially from Africa, their origin in that continent will be determined with more confidence.

However, the data available suggests a strong Mozambique-Angola contribution to L3e in Puerto Rico. For instance, the frequency of L3e in Puerto Rico, 18%, is much higher than its 11% frequency in West Africa. Only West Central Africa (22%) and Angola (21%) show higher frequencies than Puerto Rico. West Central Africa is the only continental region were sampling has been well represented, with 1806 samples. Among these samples, L3e2 is twice as frequent as L3e1. Therefore, it is reasonable to conclude

that Puerto Rico's higher L3e frequency may be due to significant contributions from Angola and Mozambique, where the L3e frequency is 17%.

Because these regions have higher L3e frequencies, it should not be surprising that Puerto Rico's L3e sub-haplogroup distribution is more similar to Mozambique than to West Africa, even when West Africa may have been the major contributor to Puerto Rico overall. Sao Tomé is off the coast of northern Gabon, and thus it was used to store mostly African slaves from West Central Africa, and from Angola in a lesser extent. This fact is confirmed with the PCA plot constructed with all haplogroups (Figure 4A), where Sao Tomé is found far closer to West Central Africa, and then to Angola, than to any other region. The similarity of Sao Tomé to West Central Africa is repeated in the PCA plot of Figure 3, where only L3e samples are accounted for. In that plot, Puerto Rico falls equidistant between Sao Tomé and Mozambique, suggesting its L3e mtDNAs represent a combination from both groups. As Sao Tomé may possess highly significant contributions from Angola, it will be reasonable to expect that the origin of most L3e mtDNAs in Puerto Rico is in Angola and Mozambique.

As PCA, the Bayesian approach equation assigns the same weight to populations with very dissimilar sample sizes. In addition, it may overestimate the contribution of regions with low contributions, and may underestimate that of regions of high contribution. The latter seems not to be a problem in this study, as all four continental regions were estimated to contribute from 21.7 to 29.1% of the Puerto Rico L3e pool. This is further evidence that, unlike L1b, the most common African haplogroup in Puerto Rico, which is far more frequent in West Africa than in any other region of the continent (Table 9), L3e in Puerto Rico is the product of significant contributions from all African regions.

Now that ancestry-informative markers have become available, it may be possible to make a stronger association between mtDNA haplogroup and African region of origin. Because the African Slave Trade started in West Africa, and moved later to West Central Africa, and later to Angola and Mozambique, which were the regions most exploited after the first one or two decades of the 19th century, it may be expected that Puerto Ricans with African ancestries centered in Angola and Mozambique may exhibit a higher African ancestry than those whose African origins are rooted in West Africa. Accordingly, chromosomal segments of African ancestry may be larger, on average, among Puerto Ricans with L3e mtDNAs than those with L1b mtDNAs.

The distribution of L3e sub-haplogroups clearly represents 17% of our present Puerto Rican population and it suggests that the origin is more closely to Mozambique according to PC analyses, but the estimation of Bayesian approach establishes the West Coast as an equally important region, according to our slave trade traditional history. The representation of these data is limited by the fact that our samples size is small, and so it is in populations across Africa, although the Puerto Rican sample was randomly selected to be represented of Puerto Rican population.

In fact, the PCA that considered all haplogroups showed a stronger affinity of Puerto Rico for West Central Africa than for any other region, and the least affinity to Mozambique. This is in big contrast relative to the results using only L3e mtDNAs that showed that Mozambique and West Africa were nearly equally contributors to our L3e in Puerto Rico. This could be explained if Mozambique had a far higher frequency of L3e mtDNA that West and West Central Africa. In that case, our selection of L3e mtDNAs would have biased our sample set toward Mozambique. However, L3e frequencies in West Central Africa, Angola, and Mozambique are 22.0%, 20.9%, and 16.6%,

respectively. Thus, if the data we have from those regions, and if the slave trade as well would have selected samples truly representative of their respective regions, we would have expected our L3e samples set to be biased toward West Central Africa and Angola, and clearly they are not. Rather, these regions show higher affinity to Puerto Rico only when all haplogroups are considered. It is tempting to suggest that women belonging to the L3e haplogroup may have been targeted for the slave trade in Mozambique, and spared in West Central Africa.

CONCLUSION

We conclude that the use of mtDNA variation is a very efficient tool to trace back the maternal heritage of African origin but it may need large sample size or representative sample sets. The F_{ST} overall value established the differences among L3e African population helping us in the process to compare the variation of them with our present Puerto Rican population. The distribution of L3e1 and L3e2 through the African continent is around the Southeast and West African regions, respectively. Sequencing analysis showed that our samples have a higher frequency of L3e1, which can be explained because the Atlantic slave trade included populations that were not only from Western Africa. It is important to consider that from all four sub-haplogroups, L3e2 followed in frequency in Puerto Rico, then L3e4 with six samples and finally, L3e3 was represented by only one sample.

The variation closeness of Mozambique to Puerto Rico observed in our PCA shed light on how this Southeast African region contributed to the slave trade to Puerto Rico. The common region of Sao Tomé, where many Africans from the slave trade stopped before their trip to the Americas, was also confirmed through our PCA analysis because it is the closest group from all African geographic regions to Puerto Rico. Finally, the results from the Bayesian approach equation confirmed historical and previously published data (Martinez et al., 2005) that our slave trade main mtDNA gene pool contribution is from West Africa. It also established Mozambique as an equally contributor of our mtDNA gene pool and it let us interpret better the results from PCA.

These two analyses, although different, in our case complement each other giving us a better understanding on historical records that are very poor in details such as geographic and communities contributing to the slave trade to Puerto Rico.

RECOMMENDATIONS

Our analysis of mtDNA suggests that populations of southeastern Africa, like Mozambique, have played an important and persistent role in the slave trade to Puerto Rico. It also gave us new insights into how the slave trade took place. It is important to explore the contribution of other haplogroups within paragroup L to see if the same correlation could be established, studies of other haplogroups and their sub-haplogroups can help in this objective. Other haplogroups with high F_{ST}, such as L2a (F_{ST} = 0.30) and L3d (F_{ST} = 0.34), which are also in high frequency in Puerto Rico, could also contribute to explain the demographic changes that took place during the African slave trade.

Comprehensive studies of the human mtDNA genome analyzing single nucleotide polymorphisms (SNPs) using RFLPs have been proven to be an excellent tool, as well as sequences of HVS-I and HVS-II. Next, to take advantage of new technologies such like Next Generation Sequencing (NGS), complete mtDNA sequences can be made available, thus, allowing to measure diversity across African populations and also the average number of pairwise nucleotide differences to determine similarity.

Analyses derived from NGS, such as Admixture mapping can also help in this endeavor of understanding admixed populations like Puerto Ricans. Africa is by far the continent with the highest diversity. Ancestry-informative-markers (AIMs) that could distinguish unambiguously between different zones of the continent must exist, and these could be used for a better understanding of the African Slave Trade. It is important to undertake such studies because mtDNA only traces the maternal lineage. Thus, this study only registers female African migrations to Puerto Rico. Male African migrations could be addressed through Y-chromosome studies. However, the Y-chromosome has

particularly little diversity in Africa, and thus autosomal studies based on AIMs may hold more promise.

.

REFERENCES

Alexe G., Satya R. V., Seiler M., Platt D., Bhanot T., Hui S., Tanaka M., Levine A.J., Bhanot G. 2008. **PCA and clustering reveal alternate mtDNA phylogeny of N and M clades.** J. Mol. Evol. 67:465-487.

Álvarez-Nazario M. 1974. **El elemento afronegroide en el español de Puerto Rico**. San Juan, Puerto Rico: Instituto de Cultura Puertorriqueña

Atkinson D.Q., Gray R.D., and Drummond A.J. 2009. **Bayesian coalescent inference of major human mitochondrial DNA haplogroup expansions in Africa.** Proc. Biol. Sci. 276:367-373.

Andrews RM, Kubacka I, Chinnery PF, Lightowelers RN, Turnbull DM and Howell N. 1999. **Reanalysis and revision of the Cambridge reference for human mitochondrial DNA.** Nat. Genet. 23:147

Bamshad M., Wooding S., Salisbury B. and Stephens C. 2004. **Deconstructing the relationship between genetics and race.** Nat. Rev. Genet. 5:598-609.

Bandelt H.J., Silva J. A., Guimaraes P.E., Santos M.S., Brehm A., Coppa A., Larruga J. M., Rengo C., Scozzari R., Torroni A., Trata M. J., Amorim A., Prado V. F. and Pena D.J. 2001. **Phylogeography of human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade**. Am. J. Hum. Genet. 65:549-563.

Bandelt H.J., Herrnstadt C., Yao Y.G., Kong Q.P., Kivisidild T., Rengo C., Scozzari R., Richards M., Villems R., Macaulay V., Howell N., Torroni A. and Zhang Y.P. 2003. **Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats.** Ann. Hum. Genet. 67:512-524.

Batini C., Coia V., Battaggia C., Rocha J., Pilkington M.M., Spedini G., Comas D., Destro-Bisol G., Clafell F. 2007. **Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa.** Mol. Phylogenet. Evol. 43:635-644.

Behar D., Villems R., Soodyall H., Blue-Smith J., Pereira L., Metspalu E., Scozzari R., Makkan H., Tzur S., Comas D., Bertranpetit J., Quintana-Murci L., Tyler-Smith C., Wells R.S., Rosset S., and the Genographic Consortium. 2008. **The dawn of human matrilineal diversity.** Am. J. Hum. Genet. 82:1130-1140.

Bortolini M.C., Da Silva W.A. Jr., Zago M.A., Elion J., Krishnamoorthy R., Goncalves V.F., and Pena S.D. 2004. **The phylogeography of mitochondrial DNA haplogroup L3g in Africa and Atlantic slave trade.** Am. J.Hum. Genet. 75:522-524.

Brehm A., Pereira L., Bandelt H.J., Prata M.J., and Amorim A. 2002. **Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade**. Ann. Hum. Gent. 66:49-60.

Brehm A., Pereira L., Kivisild T., and Amorim A. 2003. **Mitochondrial portraits of the Madeira and Acores archipelagos witness different pools of its settlers.** Hum. Genet. 124:77-86.

Campbel MC and Tishkoff SA. 2010. **The evolution of human genetic and phenotypic variation in Africa.** Curr. Biol. 20:R166-173.

Cann R.L., Stoneking M. and Wilson A.C. 1987. **Mitochondrial DNA and human evolution**. Nature. 325:31-36.

Cerny V., Hájek M., Bromová M., Cmejla R., Diallo I., and Brdicka R. 2006. **mtDNA of Fulani nomads and their genetic relationships to neighboring sedentary populations.** Hum. Biol. 78:9-27.

Cerny V., Salas A., Hajek M., Saloudkova M., and Brdicka R. 2007. **A bidirectional corridor in the Sahel-Sudan Belt and the distinctive features of the Chad Basin populations: a history revealed by the mitochondrial DNA genome.** Ann. Hum. Genet. 71-433-452.

CoiaV., Destro-Bisol G., Verginelli F., Battaggia C., Boschi I., Cruciani F., Spedini G., Comas D., and Calafell F. 2005. **Brief communication: 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.

Chen Y., Ockers A., Schurr T.G., Kogenik A.M., Houponene K., and Wallace D.C. 2000. **mtDNA variation in the South African Kung and Khwe and their genetic relationships to other African populations.** Am. J. Hum. Genet. 66.1362-1383.

Destro-Bisol G., Coia V., Boschi I., Verginelli F., Cagliá A., Pascali V., Spedini G., Calafell F., 2004. **The analysis of variation of mtDNA hypervariable region 1 suggests that Eastern and Western Pygmies diverged before the Bantu expansion.** Am. Nat. 163:212-226.

Ely B., Wilson J.L., Jackson F., and Jackson B.A. 2006. **African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups.** BMC Biol. 4:34.

Excoffier L and Lischer HEL (2010) **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows**. Molecular Ecology Resources 10:564-567.

Fadhlaoui-Zid K., Plaza S., Calafell F., Ben Amor M., Comas D. and Bennamar El gaaied A. 2004 **Mitocondrial DNA heterogeneity in Tunisian Berbers**. Annals of Human Genetics 68:222-233.

Giles RE, Blanc H, Cann HM, Wallace DC**.** 1980. **Maternal inheritance of human mitochondrial DNA.** Proc. Natl. Acad. Sci. USA 77:6715-6719

Goncalves V.F., Carvalho C.M., Bortolini M.C., Bydlowski S.P., Pena S.D. 2008. **The phylogeography of African Brazilians.** Hum. Hered. 65:23-32.

Gonder M.K. Mortensen H.M., Reed F.A., de Sousa A., Tishkoff S.A., 2007. **Whole-mtDNA genome sequence analysis of ancient African lineages.** Mol. Biol. Evol. 24;757-768.

González A.M., Cabrera V.M., Larruga J.M., Tounkara A., Noumsi G., Thomas B.N. and Moulds J.M. 2006. **Mitochondrial DNA variation in Mauritania and Mali and their genetic relationship to other Western Africa populations.** Ann. Hum. Genet. 70:631-657.

Hooker D. 1996. **The African Diaspora.** Publisher: Pullman Wash.

Howell N.., Elson J.L., Turnbull D.M. and Herrnstadt C. 2004. **African Haplogroup L mtDNA sequences show violations of clock-like evolution.** Mol. Biol. Evol. 21:1843-1854.

Hünemeier T., Carvalho C., Marrero AR., Salzano FM., Junho-Pena SD., Bortolini MC. 2007. **Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data.** Am. J. Phys. Anthropol. 133:854-867

Kitchen A., Ehret C., Assefa S., and Mulligan CJ. 2009. **Bayesian phylogenetic analysis of Semitic languages identifies an early bronze age origin of Semitic in the Near East.** Proc. R. Soc. B 276:2703-2710

Kivisild T., Reidla M., Metspalu E., Rosa A., Brehm A., Pennarum E., Parik J., Geberhiwot T., Usanga E., and Villems R. 2004. **Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears.** Am. J. Hum. Genet. 75:752-770.

Kivisild T., Shen P., Wall D.P., Do B., Sung R., Davis K., Passarino G., Underhill P.A., Scharfe C., Torroni A., Scozzari R., Modiano D., Coppa A., Knijff P., Feldman M., Cavalli-Sforza L.L. and Oefner P.J. 2006. **The role of selection in the evolution of human mitochondrial genomes.** Genetics 172:373-387.

Knight A., Underhill P.A., Mortensen H.M., Shivotovsky L.A., Lin A.A., Henn B.M., Louis D., Ruhlen M., and Mountain J.L. 2003. **African Y chromosome and mtDNA divergence provides insight into the history of click languages.** Current Biology. 13:464-473

Martínez-Cruzado JC, Toro G, Viera J, Rivera MY, Startek J, Latorre M, Román A, Rivera R, Navarro I, Gómez E, Caro H and Valencia P. 2005. **Reconstructing the population history of Puerto Rico by means of mtDNA phylogeographic analysis**. *Am. J. Physical Anthropology. 128:131-155.*

Mateu E., Comas D., Calfell F., Pérez A., Abade A. and Bertranpetit J. 1997. **A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and Sao Tomé, Gulf of Guinea**. Ann. Hum. Genet. 61:507-518.

Mendizabal I., Sandoval K., Berniell-Lee G., Calafell F., Salas A., Martínez-Fuentes A. And Comas D. 2008. **Genetic origin, amixture and asymmetry in maternal and paternal human lineages in Cuba.** BMC Evol. Biol. 8:213-223

Merriweather D.A., Clark A.G., Ballinger S.W., Churr T.G., Soodyall H., Jenkins T., Sherry S.T. 1991. **The structure of human mitochondrial DNA variation.** J. Mol. Evol. 33:543-555.

MishmarD, Ruiz-Pesini E. Glik P, Macaulay V, Clark AG, Hosseini S., Brandon M. , Easley K., Chen E., Brown MD, Sukernik RI, Okkers A, and Wallace DC. 2003 **Natural selection shaped regional mtDNA variation in humans.** PNAS. 100:171-176.

Mishmar D., Ruiz-Pesini E., Brandon M., and Wallace DC. 2004. **Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration.** Hum. Mut. 23:125-133.

National Human Genome Research Institute.2005. **The use of racial, ethnic, and ancestral categories in human genetics**. Am. J. Hum. Gent. 77:000-000.

Pereira L., Macaulay V., Torroni A., Scozzari R., Prata M. J., Amorim A. 2001. **Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade.** Ann. Hum. Genet. 65:439-458.

Pereira L., Freitas F., Fernandes V., Pereira J.B., Costa M.D. Costa S., Máximo V., Macaulay V., Rocha R. Samuels D.C. 2009. **The diversity present in 5140 human mitochondrial genomes.** Am. J. Hum. Genet. 84:628-640.

Plaza S., Salas A., Calefell F., Corte-Real F., Bertranpetit J, Carracedo A., and Comas D. 2004. **Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola.** Hum. Genet. 115:439-447

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA and Reich D. 2006. **Principal components analysis corrects for stratification in genome-wide association studies.** Nat.Genet. 12:e190.

Quintana-Murci L., Quach H., Harmant C., Luca F., Massonnet B., Patin E., Sica L., Mouguiama-Daouda P., Comas D., Tzur S., Balanovsky O. Kidd K.K., Kidd J.R., van der Veen L., Hombert J.M., Gessain A., Verdu P., Froment A., Bahuchet S., Heyer E., Dausset J., Salas A., Behar D.M. 2008. **Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers.** Proc. Natl. Acad. Sci. USA. 105:1596-1601.

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ and Sykes B. 1996. **Paleolithic and Neolithic lineages in Eruropean mitochondrial gene Pool**. Am. J. Hum. Genet. 59:185-203.

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

Rosa A., Brehm A., Kivisild T., Metspalu E., and Villems R. 2004. **mtDNA profile of West Africa Guineans: towards a better understanding go the Senegambia regions.** Ann. Hum. Genet. 79:340-352.

Saillard J, Magalhaes PJ, Schwarz , Rosenberg T, Norby S. 2000. **Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV.** Hum Biol. 72:1065-1068

Salas A., Richards M., De la Fe T., Lareu M., Sobrino B., Sánches-Diz P., Macaulay V., and Cariacedo A. 2002. **The making of the African mtDNA landscape**. Am. J. Hum. Genet. 71:1082-1111.

Salas A., Richards M., Lareu M., Scozzari R., Coppa A., Torroni A., Macaulay V. and Carracedo A. 2004. **The African Diaspora: Mitochondrial DNA and the Atlantic slave trade**. Am .J. Hum. Genet. 74:454-465.

Salas A., Carracedo A., Richards M. and Macaulay V. 2005. **Charting the ancestry of African Americans**. Am. J. Hum. Genet. 77:676-680.

Salas A., Richards M., Lareu MV., Sobrino B., Silva S., Matamoros M., Macaulay V. and Carrecedo A. 2005. **Shipwrecks and founder effects: divergent demographic histories reflected in Caribbean mtDNA.** Am. J. Phys Anthropol. 000:000-000.

Silva W.A., Bololinin M.C., Schneider M.P., Marrero A., Kirishnamoorthy R., and Zago M.A. 2006. **MtDNA haplogroup analysis of black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade.** Hum. Biol. 78:29-41.

Singh T. R. 2008. **Mitochondrial gene rearrangements: new paradigm in the evolutionary biology and systematics.** Bioinformation 3:95-97.

Shriver M.D., and Kittles R.A. 2004. **Genetic ancestry and the search for personalized genetic histories.** Nat. Rev. Genet. 5:611-618.

Schurr T.G. and Sherry S.T. 2004. **Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence.** Am. J. Hum. Biol. 16:420-439.

Soares P., Ermini L., Thomson N., Mormina M., Rito T., Röhl A., Salas A., Oppenheimer S., Macaulay V., and Richards M.B. 2009. **Correcting for purifying selection: an improved human mitochondrial molecular clock.** Am. J. Hum. Genet. 84:740-759.

Tishkoff SA., Goder MK., Henn BM., Mortensen H., Knight A., Gignoux C., Fernandopulle N., Lema G., Nyambo TB., Ramakrishnan U., Reed FA., and Mountain JL. 2007. **History of Click-Speaking population of Africa inferred from mtDNA and Y chromosome genetic variation.** Mol. Biol. Evol. 24:2180-2195.

Tishkoff SA., Reed FA., Friedlaender FR., Ehret C., Ranciaro A., Froment A., Hirbo JB, Awomoyi AA., Bodo JM., Boumbo O., Ibrahim M., Juma AT., Kotze MJ., Lema G., Moore JH., Mortensen H., Nyambo TB., Omar SA., Powell K., Pretorius GS., Smith MW, Thera MA., Wambebe C., Weber JL>, Williams SM. 2009. **The genetic structure and history of Africans and African Americans.** Science 324:1035-1044

Tsigenopoulos CS, Ráb P, Naran D and Berrebi P. 2002. **Multiple origins of polyploidy in the phylogeny of southern African barbs cyprinidae) as inferred from mtDNA markers.** Heredity. 88:466-473.

Torroni A., Achilli A., Macaulay V., Richards M. and Bandelt H.J. 2006. **Harvesting the fruit of the human mtDNA tree.** Trends Genet. 22:339-345.

Torroni A., Rengo C. Guida V., Criciani F., Sellitto D., Coppa A. Luna F., Calderon L., Simionati B., Valle G., Richards M., Macaulay V., and Scozzari R. 2001. **Do the four clades of the mtDNA haplogroup L2 evolve at different rates?** Am. J. Hum. Genet. 69:1348-1356.

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

Turchi C., Bucemi L., Giacchino E., Onofri V., Fendt L., Parson W., Tagliabracci A. 2009. **Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enrichment of forensic mtDNA databased with Northern Africa data.** Forensic Sci. Int. Genet. 8:166-172.

Veeeramah KR., Connell BA., Pour NA., Powell A., Plaster CA., Zeitlyn D., Mendell NR., Weale ME., Bradman N., Thomas MG. 2010. **Little genetic differentiation as assessed by a uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria.** BMC Evol. Biol. 10:92

Xing J., Watkins W.S., Witherspoon D.J., Zhang Y., Guthery S.L., Thara R., Mowry B.J., Bulayeva K., Weiss R.B., and Jorde L.B. 2009. **Fine-scaled human genetic structure revealed by SNP microarrays.** Genome Res. 19:815-825.

Zhu C. and Yu J. 2009. **Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types.** Genetics 182:875-888.

Table 1. RFLP 4 enzyme analysis for L3e in Puerto Rican samples
This table shows the unresolved groups L3e1, L3e1a and L3e2b and the characterized sub-haplogroups L3e2a, L3e3 and L3e4.

Sub-haplogroups	*Hae III (9553)*	*Ava II (5260)*	*Alu I (5584)*	*DpnII (14869)*
L3e1	+	-	+	+
L3e1a	+	-	+	+
L3e2	+	-	+	+
L3e2a	+	-	+	-
L3e2b	+	-	+	+
L3e3	-	+	+	+
L3e4	+	+	-	+

Table 2. Mitochondrial DNA sequences used in the present study including their geographic origin, sample size, and reference

Geographical origin (Code)	*n*	*Reference*
Puerto Rico (PR)		
Puerto Rico	35	Present study
Puerto Rico	164	Viera-Vera, 2006
West Central Africa (WCA)		
Non-Atlantic WCA	19	Bandelt et al. 2001
Cameroon	42	Batini et al. 2007
Mbuty Pygmies and Biaka Pygmies	98	Chen et al. 2000
Cameroon	3	Silva et al. 2006
West Africa (WA)		
Mali	122	González et al. 2006
Mali	78	Ely et al. 2002
Senegambia	12	Bandelt et al. 2001
Senegal	5	Batini et al. 2007
Mandenka and Wolof	171	Chen et al. 2000
Mauritana	28	González et al. 2006
Angola (A)		
Angola	43	Plaza et al. 2004
Mozambique (M)		
Mozambique	112	Pereira et al. 2001
Mozambique	280	Salas et al. 2002
Sao Tomé (ST)		
Sao Tomé	36	Batini et al. 2007
Sao Tomé	48	Mateu et al. 1997

Table 3. Puerto Rican samples used in the present study

Geographical Origin	*Geographic Area*	*Code*	*n*
Mayagüez	West	A	2
Moca	West	C	1
Caguas	Center	F	4
San Lorenzo	Center	G	1
Peñuelas	South	K	1
Toa Baja	North	L	1
Bayamón	North	M	2
Carolina	North	N	3
Humacao	East	O	1
Cayey	Center	P	4
Yauco	South	Q	4
Guaynabo	Center	S	1
Barranquitas	Center	U	2
San Sebastián	West	V	1
San Juan	North	W	5
Patillas	South	Z	1
Vega Alta	North	α	3
Vieques	East	δ	3
Total			40

Table 4. Primers sequences used in this study

Primers	**Test**	**Sequences**
L29	SEQUENCING	GGTCTATCACCCTATTAACCAC
H580	SEQUENCING	TTGAGGAGGTAAGCTACATA
L15829	SEQUENCING	CATCCGTACTATACTTCACAAC
H34	SEQUENCING	ACCAAATGCATGGAGAGCTCC
L16219	SEQUENCING	TGCTTACAAGCAAGTACAGCA
H16345	SEQUENCING	GGG ACG AGA AGG GAT TTG AC
L15854	SEQUENCING	CCTAATCCTAATACCAACTATC
H16526	SEQUENCING	GGGAACGTGTGGGCTATTTAGG
L9445	HAE III	ACCTGTCCAAAAAGGCCTTC
H9587	HAE III	GTGTTTAGGAGTGGGACTTC
L5120	AVA II AND ALU I	TAACTACTACCGCATTCCTA
H5652	AVA II AND ALU I	AGTCCCATTGGTCTAGTAAG
L14885	DPN II	ATGGCTAGGAATAGTCCTG
H14711	DPN II	CAACCACGACCAATGATATG

Table 5. RFLP tested for L3e sub-haplogroups in Pueto Rican samples

Samples/Enymes	*HaeIII*	*AvaII*	*AluI*	*DpnII*	*Sub-haplogroup*
AE4	+	+	-	+	L3e4
AF3	+	-	+	+	L3e1 or L3e2b
CB3	+	0	+	+	L3e1 or L3e2b
FB3	+	-	+	+	ND
FD1	+	-	+	+	L3e1 or L3e2b
FF4	+	+	-	+	L3e4
GD7	+	-	+	+	L3e1 or L3e2b
KA2	+	+	+	+	ND
LD10	+	-	+	+	L3e1 or L3e2b
MF4	+	0	+	+	ND
MH2	+	-	+	-	L3e2a
NB7	+	-	+	+	L3e1 or L3e2b
NF2	+	-	+	+	L3e1 or L3e2b
NF4	+	-	-	+	L3e1 or L3e2b
OB3	+	-	+	+	L3e1 or L3e2b
PB6	+	-	+	+	L3e1 or L3e2b
PC1	+	-	+	+	L3e1 or L3e2b
PC2	+	-	+	+	L3e1 or L3e2b
QA9	+	-	+	+	L3e1 or L3e2b
QA17	+	+	+	+	ND
QB3	+	-	+	+	L3e1 or L3e2b
SA1	+	-	+	+	ND
UB14	+	0	+	+	ND

Samples/Enymes	*HaeIII*	*AvaII*	*AluI*	*DpnII*	*Sub-haplogroup*
UC5	+	0	+	+	ND
VB7	+	0	+	+	ND
WD1	+	0	+	+	ND
WE4	+	0	+	+	ND
WI1	+	0	+	+	ND
WI5	+	0	+	+	ND
WO1	+	+	+	-	L3e2a
ZB3	+	+	-	+	L3e4
αB4	+	+	-	+	L3e4
αB6	+	+	+	+	ND
αB15	+	0	+	+	ND
δ1[0]	No	No	No	No	L3e2b
δ2[0]	No	No	No	No	L3e2b
δ3[0]	No	No	No	No	L3e4

[0]This samples are from Vieques and they were sequenced directly without RFLP testing. ND=no data.

Table 6. HVS-I sequences from L3e sub-haplogroups, listed by motif relative to CRS

Clade/ Population Code	*HVS-I*	*Sub-haplogroup*
L3e1		
FD1	16223, 16327	L3e1
FE2	15942, 16223, 16327	L3e1
GD7	16223, 16327	L3e1
KA2	16223, 16327	L3e1
MF4	16185, 16223, 16327	L3e1a
NF2	15942, 16093, 16223, 16327	L3e1
SA1	15942, 16223, 16327	L3e1
UB14	15942, 16189	L3e1
UC5	16223, 16327	L3e1
VB7	16223, 16325d, 16327	L3e1b
WD1	16185, 16223, 16327	L3e1a
WE4	16223	L3e1
L3e2		
CB3	16223, 16320	L3e2a
MH2	16223, 16320	L3e2a
QA9	16172, 16189, 16223, 16320	L3e2b
QB3	16223, 16320, 16324	L3e2a
WI1	16223, 16320	L3e2a
WO1	16223, 16320	L3e2a
L3e3		
QA17	16265T	L3e3
L3e4		
AE4	16051, 16223, 16264, 16519	L3e4
FF4	16223, 16264, 16291	L3e4
αB6	16223, 16264, 16291	L3e4

Table 7. Summary of Results from RFLP assay and Sequencing analysis in Puerto Rico

Populations	***Sub-Haplogroups***			
	L3e1	**L3e2**	**L3e3**	**L3e4**
Puerto Rico (PR)	12	8	1	6

Table 8. L3e haplogroup distribution in Puerto Rican and African geographic regions from samples of the present study and previous studies.

Populations	***Sub-Haplogroups***			
	L3e1	**L3e2**	**L3e3**	**L3e4**
Puerto Rico (PR)	18	10	1	6
West-Africa (W)	3	27	3	9
West Central Africa (WCA)	118	221	45	13
Angola (A)	5	1	3	0
Mozambique (M)	43	6	9	7
Sao Tomé (ST)	5	6	0	2

Table 9. Major sub-Saharan African mtDNA haplogroups and their distributions in the Puerto Rican populations and 5 African geographic regions from samples of the present study and previous studies.

Populations	Sub-Haplogroups																			
	L3e1	**L3e2**	**L3e3**	**L3e4**	**L0**	**L0a**	**L1b**	**L1c**	**L2***	**L2a**	**L2b**	**L2c**	**L2d**	**L3***	**L3b**	**L3d**	**L3f**	**L3g**	**L3h**	**Totals**
Puerto Rico (PR)	18	10	1	6	2	1	49	22	1	30	6	7	1	1	15	21	3	1	4	199
West-Africa (W)	3	27	3	9	21	5	78	14	37	45	19	51	3	22	18	14	5	0	0	374
West Central Africa (WC)	118	221	45	13	88	102	155	202	16	353	63	55	16	78	83	78	120	0	0	1806
Angola (A)	5	1	3	0	0	6	2	7	0	11	3	0	0	0	0	2	1	2	0	43
Mozambique (M)	43	6	9	7	0	103	5	22	0	136	6	3	3	6	12	29	2	0	0	392
Sao Tomé (ST)	5	6	0	2	0	4	3	43	0	12	3	2	0	1	1	1	1	0	0	84

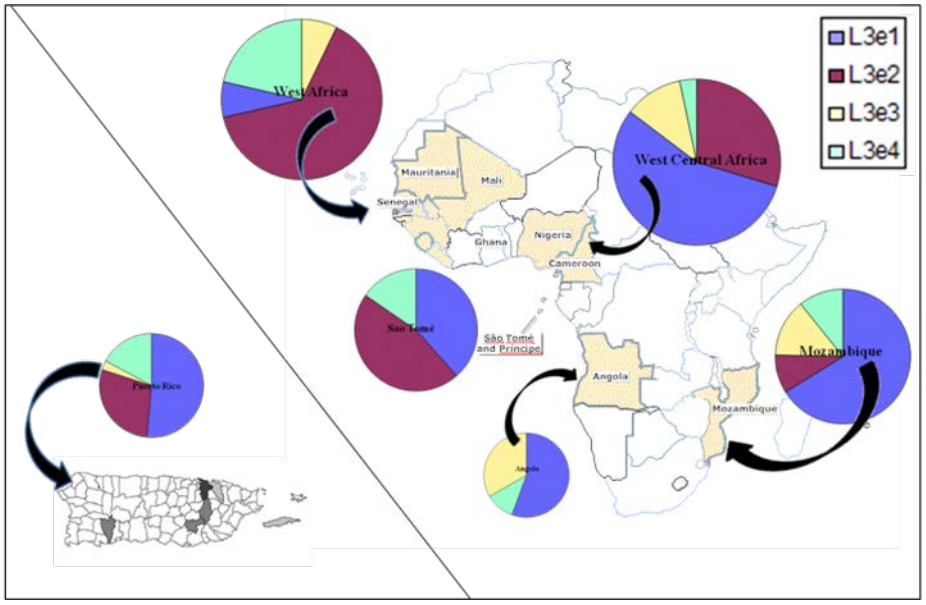


Figure 1. Map of Africa and Puerto Rico. L3e sub-haplogroups distributions acrooss Africa and in Puerto Rico are represented by the pie charts. Puerto Rico is similar to Mozambique and Sao Tomé. The pie charts area are proportional to sample size. The shaded countries represent the populations used for comparison in the present study.

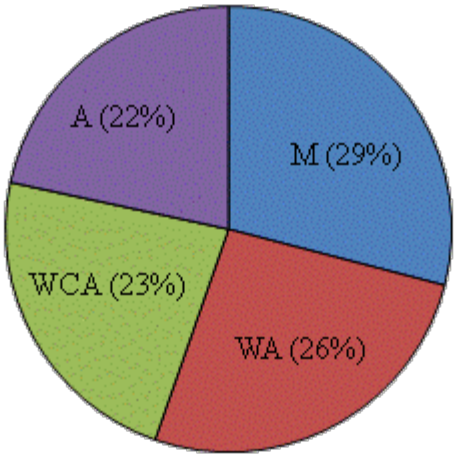
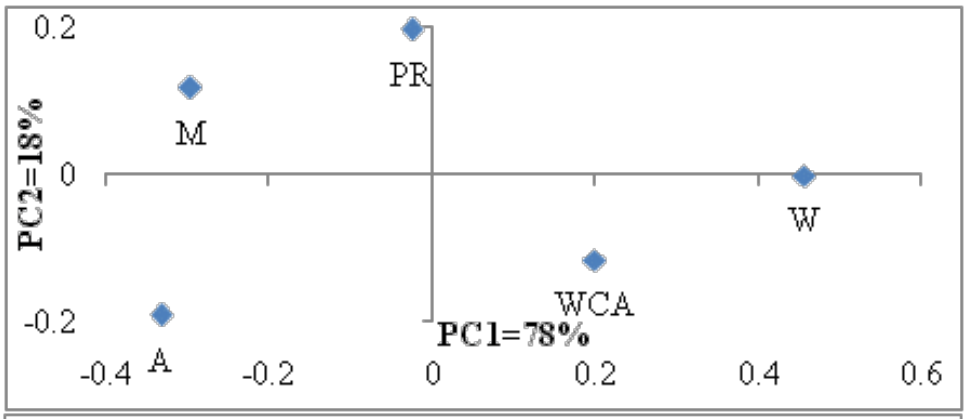
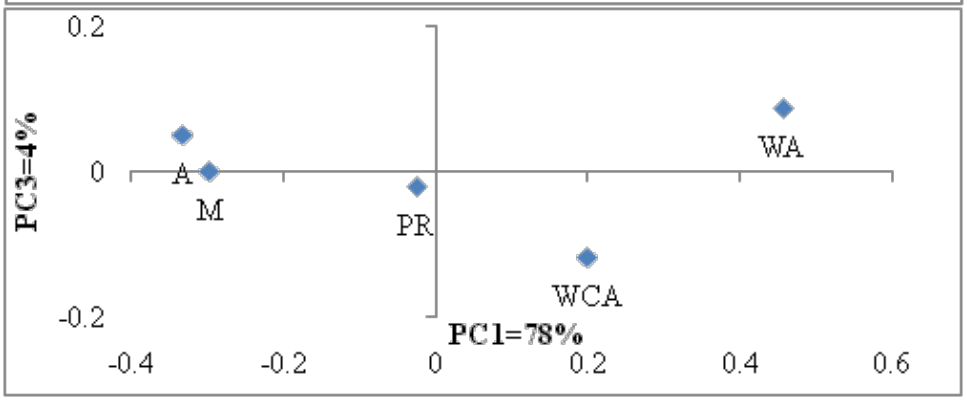


Figure 2. Bayesian approach of L3e haplogroup origin in Puerto Rico. The African continent contribution to the L3e sub-haplogroups in our present Puerto Rican population are as follows: Mozambique (M), West Africa (WA), West Central Africa (WCA) and Angola (A).

A



B



C

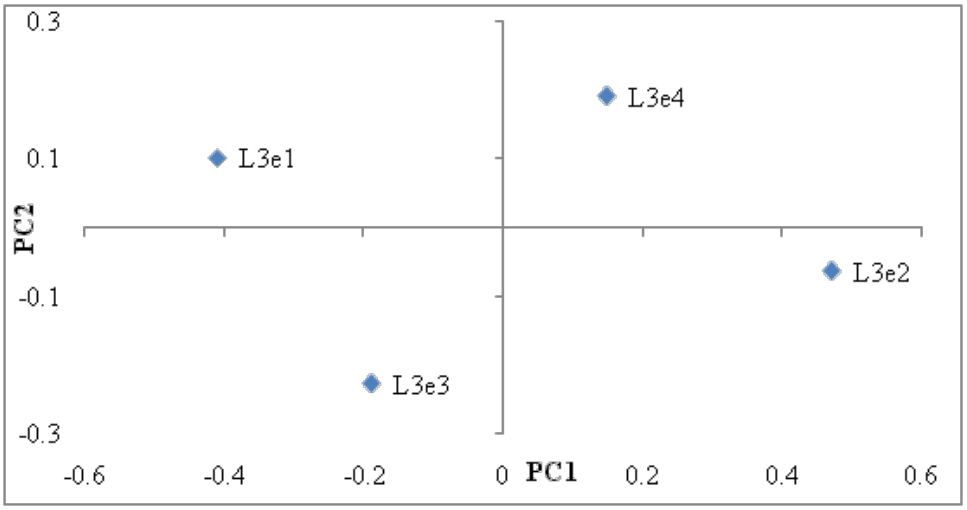


Figure 3. Analysis of Principal Components among Puerto Rican L3e sub-haplogroups and their relative comparison with their distribution in continental Africa. **A.PC1 vs. PC2** covers 96% of the variance. **B.PC1 vs. PC2** covers 80% of the variance. **C**. Shows the Allele frequencies of the L3e analysis for the corresponding PCA. Geographic areas: West Central (WC), Mozambique (M), Angola (A), West (W) within Africa and Puerto Rico (PR).

A

B

C

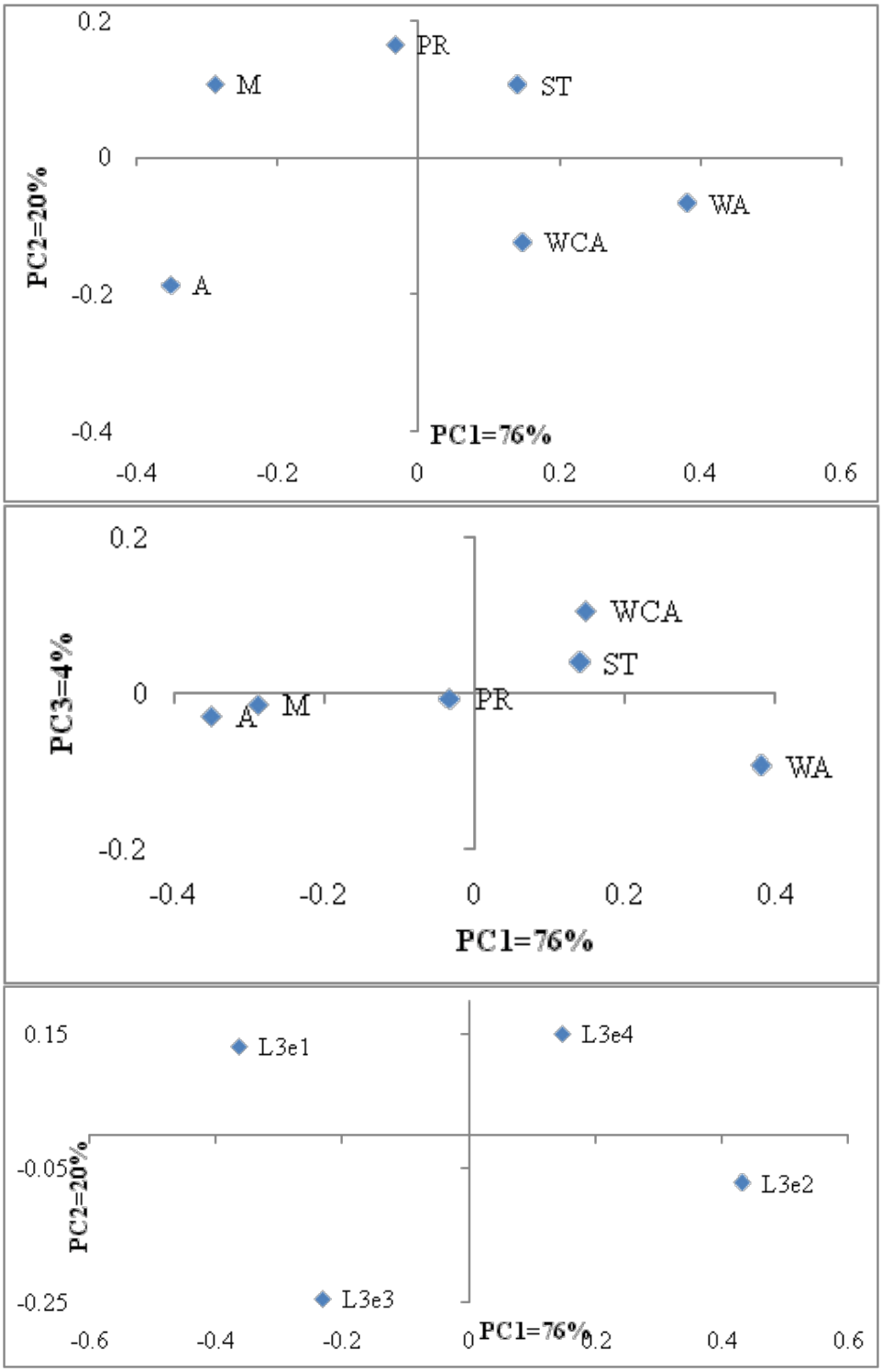
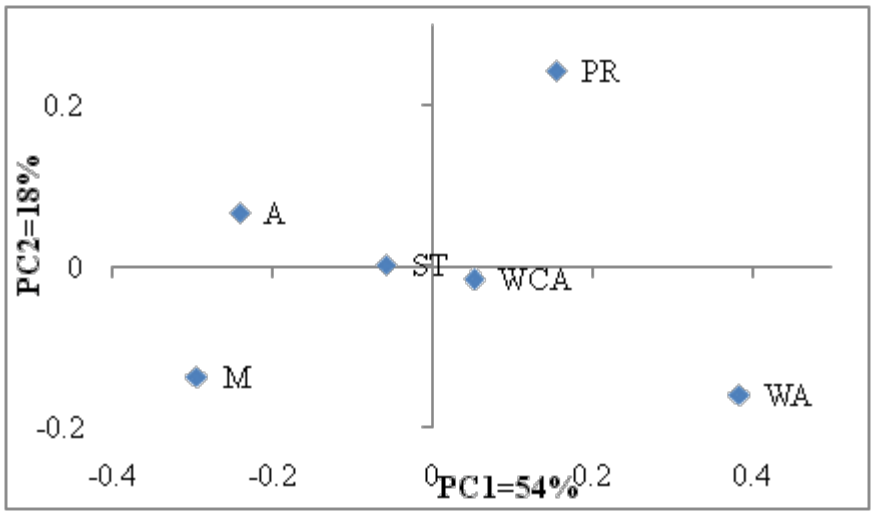
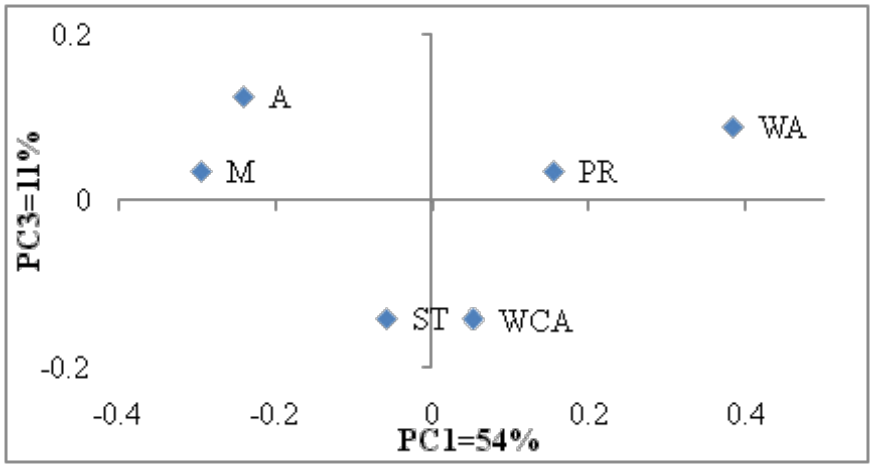


Figure 4. Analysis of Principal Components among Puerto Rican L3e sup-haplogroups and their relative comparison with their distribution in Africa. **A. PC1vs PC2**; **B.PC1 vs PC3**; **C.** All the corresponding L3e sub-haplogroups. Abbreviations as in Figure 2. ST = Sao Tomé.

A



B



C

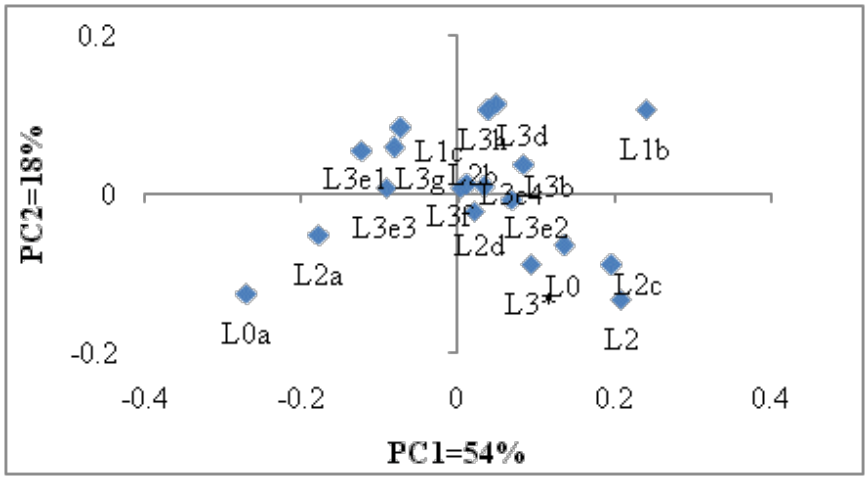


Figure 5. PCA among all Puerto Rican African Sub-Saharan phylogeographic regions and their relative comparison with the phylogenetic haplogroups. **A.** Shows geographic areas in this study within **PC1vs PC2**; **B. PC1 vs PC3**; and **C.** All Sub-Saharan haplogroups including the L3e sub-haplogroups.