HAPLOGROUP FREQUENCIES OF Y CHROMOSOMES WITH THE 92R7T ALLELE IN PUERTO RICO

By

Katherine Martínez-Vargas

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

MASTER IN SCIENCES

in

BIOLOGY

UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS

December, 2007

Approved by:

Fernando J. Bird-Picó, Ph.D. Member, Graduate Committee	Date
Richard D. Squire, Ph.D. Member, Graduate Committee	Date
Juan C. Martínez-Cruzado, Ph.D. President, Graduate Committee	Date
Rima Brusi, Ph.D Representative of Graduate Studies	Date
Lucy Bunkley-Williams, Ph.D. Chairperson of the Department	Date

ABSTRACT

For over two decades, Y chromosome polymorphisms have successfully been used as lineage markers in evolutionary studies to determine human origins, migrations waves, and admixture. Determining the frequency and geographic distribution of the Y chromosome haplogroups is essential in order to determine paternal ancestry in Puerto Rico. Preliminary studies undertaken in 2002 suggested that most Puerto Rican men had the derived state of the 92R7 allele; 92R7T. This allele defines the P clade which includes European and Native American haplogroups. A substantial number of 92R7T Y chromosomes could not be classified into any specific haplogroup, thus remaining classified simply as 92R7T Y chromosomes. In this study, 99 individuals were sampled to test Puerto Rican Y chromosomes for the 92R7T allele and classified into well-defined haplogroups by identifying five polymorphisms. By using molecular techniques such as polymerase chain reaction (PCR), restriction fragment length polymorphism (RFLP) and DNA sequencing, samples with the 92R7T allele were identified and then classified into haplogroups by the analysis of the following single nucleotide polymorphisms (SNPs): P25, M242, M3, SRY10831, and M207. The results showed the presence of the 92R7T allele in 57 of the samples of which 54 are of European origin, belonging to haplogroups R1a, R1b1, and R(xR1a, R1b1), one to Native American haplogroup Q3, and the other two could not be classified into well defined haplogroups. Thus this study revealed a strong patrilineal contribution of European population to modern Puerto Rican and a very poor Native American contribution.

RESUMEN

Por más de dos décadas los polimorfismos en la región no-recombinante del cromosoma Y han sido utilizados como marcadores de linajes en estudios de evolución para determinar el origen de la humanidad, migraciones y mestizaje entre poblaciones. Determinar la frecuencia y distribución geográfica de los haplogrupos del cromosoma Y es esencial para determinar la ascendencia paternal en Puerto Rico. Estudios preliminares llevados a cabo en el 2002 sugieren que la mayoría de los hombres puertorriqueños poseen el alelo derivado de 92R7; 92R7T. Este alelo define el clade P, el cual incluye haplogrupos europeos y nativos de América. Un número sustancial de cromosomas Y con el alelo 92R7T no pudo ser clasificado en haplogrupos más definidos, permaneciendo así clasificados simplemente como cromosomas Y con el alelo 97R7T. En este estudio, 99 muestras fueron analizadas con el propósito de detectar cromosomas Y puertorriqueños con el alelo 92R7T y poder así clasificarlas en haplogrupos bien definidos por medio del estudio de cinco polimorfismos. Usando técnicas moleculares tales como reacción de polimerasa en cadena, patrones de restricción polimórfica y secuenciación de ADN, las muestras con el alelo 92R7T fueron identificadas y clasificadas en haplogrupos mediante el análisis de los siguientes polimorfismos de un solo nucleótido (SNP): P25, M242, M3, SRY10831, y M207. Los resultados muestran la presencia del alelo 92R7T en 57 de las muestras de las cuales 54 son de origen Europeo, perteneciendo a haplogrupos R1a, R1b1, R(xR1a, R1b1), solo una pertenece al haplogrupo Q3 nativo de América y las otras dos no pudieron ser clasificadas en haplogrupos definidos.

Por lo tanto este estudio revelo una fuerte contribución paterna de origen Europeo a la población moderna puertorriqueña mientras que la contribución indígena fue bien pobre.

To my husband William and my son Victor, thank you for your unconditional support and love through this process… I love you!!!

ACKNOLEDGEMENTS

To my graduate committee, doctors Juan C. Martínez-Cruzado, Richard Squire and Fernando Bird-Picó, thank you for your advise, motivation and support during this process. I would like to especially recognize Richard Squire, thank you for always believing in me and for all your advice and guidance during a very crucial time of my life. I would also like to especially thank Juan C. Martínez-Cruzado for the opportunity and for all of your encouragement during the hard times.

To my parents, Victor M. Martínez-Martínez and Zaida Vargas-Fernández, and my sister, Jacqueline Martínez-Vargas, thank you for all of your love and unconditional support during all these years.

To my colleagues, thank you for all the unforgettable experiences we shared and the lifetime friendships.

To my husband, William Gonzalez-Rodríguez, thank you for always believing in me and for all of your love and unconditional support during all these years. I would have never done it without you.

To my beautiful and lovely son, Victor Johvan Rodríguez-Martínez, you are my inspiration and my everything… mami loves you!!!

TABLE OF CONTENTS

ABSTRACT ……………………………………………………………..…………..ii

RESUMEN ...……………………………………………………...…………iii

ACKNOWLEDGEMENTS ……….………………………………..…………iv

TABLE OF CONTENTS …………….……………………………....……………vi

TABLE LIST ……………………….………………………………….………….vii

FIGURE TABLE ……………………..……………………………….…………...viii

1 INTRODUCTION ……………….……………….……………………………1
2 LITERATURE REVIEW…………..…………….……………………………5

2.1 Y chromosome………………………………………..……………..……………...5
2.2 NRY polymorphisms……………………………….……………….…………….…7
2.3 NRY nomenclature……………………………….……………….……………...10
2.4 Peopling of America ...……………...………………………………...……………14
2.5 Relevant Y chromosome markers and the haplogroups they define……………………..15

3 MATERIALS AND METHODS………………………...………………18

3.1 Study Sample……………………………………………………….…………………18
3.2 DNA extraction…………………………………………………..…………………18
3.3 Identification of samples containing the 92R7T allele and haplogroup determination………………………………………………………..…………...19

3.3.1 PCR………………………………………………………………………….…21
3.3.2 RFLP…………………………………………………………………………...21
3.3.3 DNA sequencing……………………………………………………………….21

4 RESULTS ………………………………………………………….…………25
5 DISCUSSION ……………………………………………………….………..29
6 CONCLUSION ……………………………………………………………33
7 RECOMMENDATIONS ……………………………………………..34
8 REFERENCES …………………………………………………………35
9 APPENDIX ……………………………………………………….………..40

9.1 Consent form……………………………………………………….………………40
9.2 Pictures………………………………………………………………………………44

TABLE LIST

1. PCR conditions to amplify tested sites …………………………………………23

2. RFLP conditions for PCR products ……………………………………………24

FIGURE LIST

1. 2005 P-92R7 Y-Chromosome Phylogenetic Tree…………………..……..….3

2. Y-Chromosome ideogram……………………….. ……………………..…6

3. Ideogram of the Y Chromosome showing the locations of the pseudosomal regions (PAR), the non-recombining region (NRY), and the testis determining gene, SRY……………………………………………………..…..…7

4. The Phylogenetic tree of NRY haplogroups…………………..……...…..13

5. Hierarchical order for identification of Y chromosome haplogroups……….20

6. Haplogroup determination of samples following hierarchical order………...28

INTRODUCTION

Ever since it's colonization five centuries ago, Puerto Rico has been subject to admixture between diverse population groups. The original population groups were Native Americans (Arawak-speaking Taino Indians). During the 16th century, Europeans (mainly Spanish colonizers) and Sub-Saharan Africans started arriving to the island.

Studies have been carried out that characterized Y chromosome haplogroups and their frequencies in the human groups that have most contributed to the Puerto Rican gene pool according to traditional history; Native Americans (Pena et al. 1995; Underhill et al. 1996, 1999, 2000; Lell et al. 2002), Iberians (Flores et al. 2004), and West Africans (Scozzari et al. 1997, 1999; Fernández et al. 2003). Since the native population of the island was decimated, the most frequent haplogroups on the island were thought to be from Europe and West Africa. Nonetheless, a study based on the maternally inherited mitochondrial DNA (mtDNA) revealed results that challenged the conventional wisdom that the indigenous population had disappeared by the end of the sixteenth century. Results showed a significantly predominant 61.3% native American ancestry, a 27.2% of West African contribution and a considerably lower 11.5% West Eurasian component (Martinez-Cruzado et al. 2005). Such results suggest a substantial native American contribution to the Puerto Rican gene pool and demand that the male contribution be assessed as well.

The first attempt to study the Y chromosome was not as definitive as the previous mtDNA study since samples could not be identified into well defined

haplogroups. Preliminary studies performed in 2002 suggested that 53 % of Puerto Rican men had the derived state of the 92R7 allele, 92R7T (Martinez-Cruzado et al., unpublished data). This allele defines the P clade, which includes several European and Native American haplogroups (Figure 1). A substantial number of 92R7T or haplogroup P Y chromosomes could not be classified into any of these haplogroups, thus remaining classified simply as 92R7T Y chromosomes. The classification of Y chromosomes as having the 92R7T Y allele is insufficient to determine their continental origin, as haplogroup P Y chromosomes are found in Europe as well as Asia and the New World.

In 2003, a new single nucleotide polymorphism (SNP), M242, which arose in East Asia and was carried by all men in the first migratory wave into the New World, was discovered (Bortolini et al. 2003; Seielstad et al. 2003). This SNP distinguished Native American from European haplogroups within the P clade. It occurred in Central Asia prior to the emergence of the M3 mutation which defines Native American haplogroup Q3 (Schurr and Sherry 2004). Its derived allele (M242-T) is found in all Native American 92R7T Y chromosomes (Bortolini et al. 2003) but in very few Asian, and thus defines clade Q. The newly discovered SNP and the detailed characterization of all Y chromosome haplotypes by the Y Chromosome Consortium (2002) provided all the necessary information to trace the biological ancestry of all Puerto Rican 92R7T Y chromosomes.

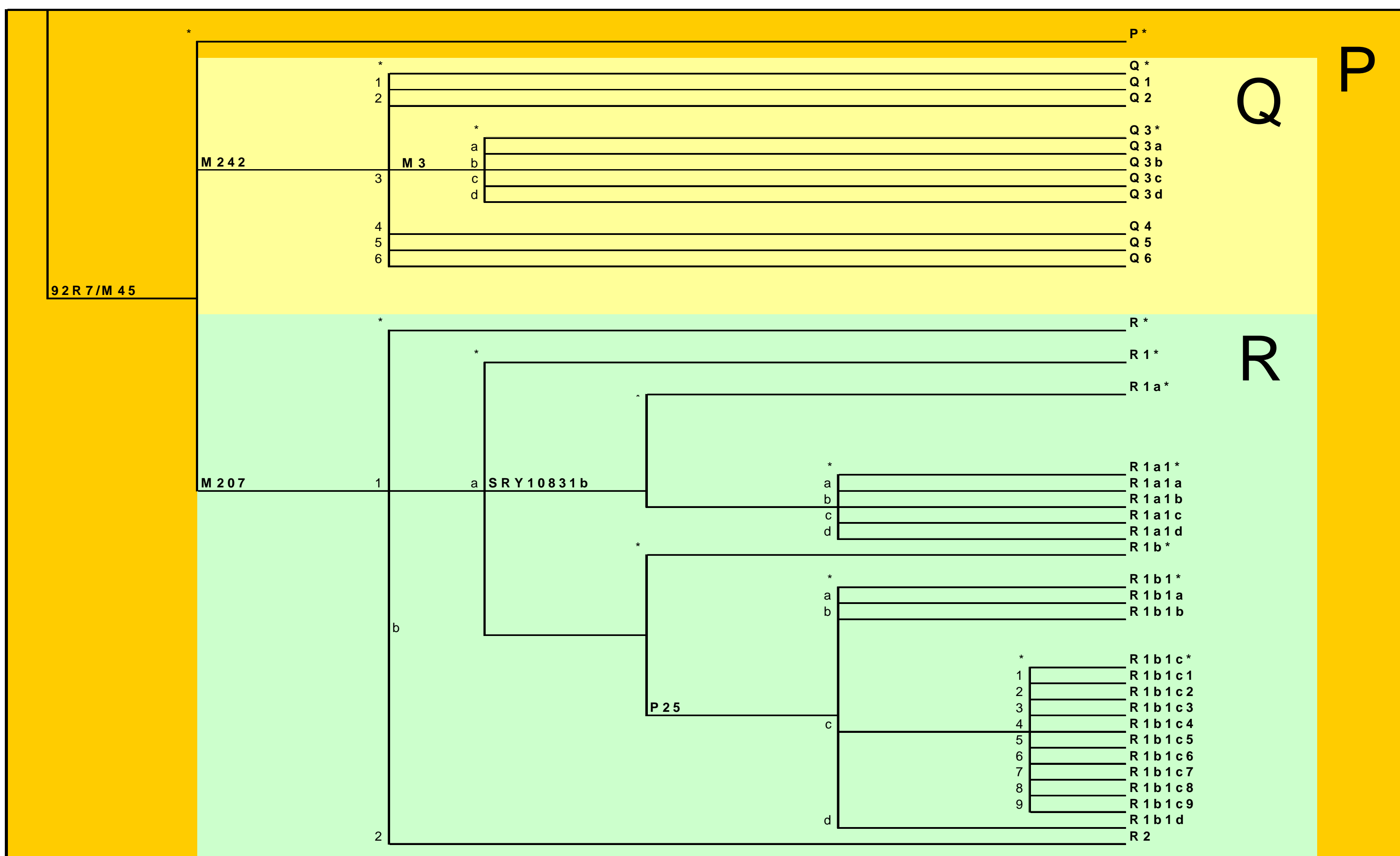


Figure 1. 2005 P-92R7 Y-Chromosome Phylogenetic Tree showing haplogroups P, Q and R. Recreated from phylogenetic tree at www.FamilyTreeDNA.com.

This study is the first attempt in Puerto Rico to classify 92R7T Y chromosomes into well-defined haplogroups that leave no doubt of their biological ancestry. By obtaining this preliminary data, we aim to gain and enhance the understanding of which 92R7T Y chromosome haplogroups are present in Puerto Rico and in what frequencies in order to develop an optimized methodology for a representative and randomized study in the Puerto Rican population. These results, in combination with the mtDNA analysis, will assist us to obtain a more complete picture of the human ancestry in Puerto Rico.

LITERATURE REVIEW

2.1 Y Chromosome

The Y chromosome is one of the sex-determining chromosomes in humans. It causes testis differentiation, thus determining maleness in an epistatic way through the action of a single gene, SRY (Sex determining region of Y chromosome) (Sinclair et al. 1990**;** Hurles et al. 2001). The Y chromosome is approximately 58 Million bp (Mb) long and contains 307 genes (Figure 2). The Y chromosome is hemizygous and lacks recombination for most (95%) of its length (Skaletsky et al. 2003). This region where there is no X-Y crossing-over in male meiosis is called the non-recombining region of Y (NRY), non-recombining portion Y (NRYP) or the male-specific region (MSY). Only two small segments called the pseudoautosomal regions (PAR) located at the end of each the long and short arms (Figure 3) recombine with the X chromosome (Hurles et al. 2001). Due to the lack of recombination in NRY, the Y chromosome sequence can only be altered by random mutations which are passed almost unaltered from generation to generation preserving a record of their history. Given that the Y chromosome is male-specific, passing from father to son, these mutations or polymorphisms on the NRY provide a unique system for the study of human origins, male migration, and admixture (Underhill et al. 1996).

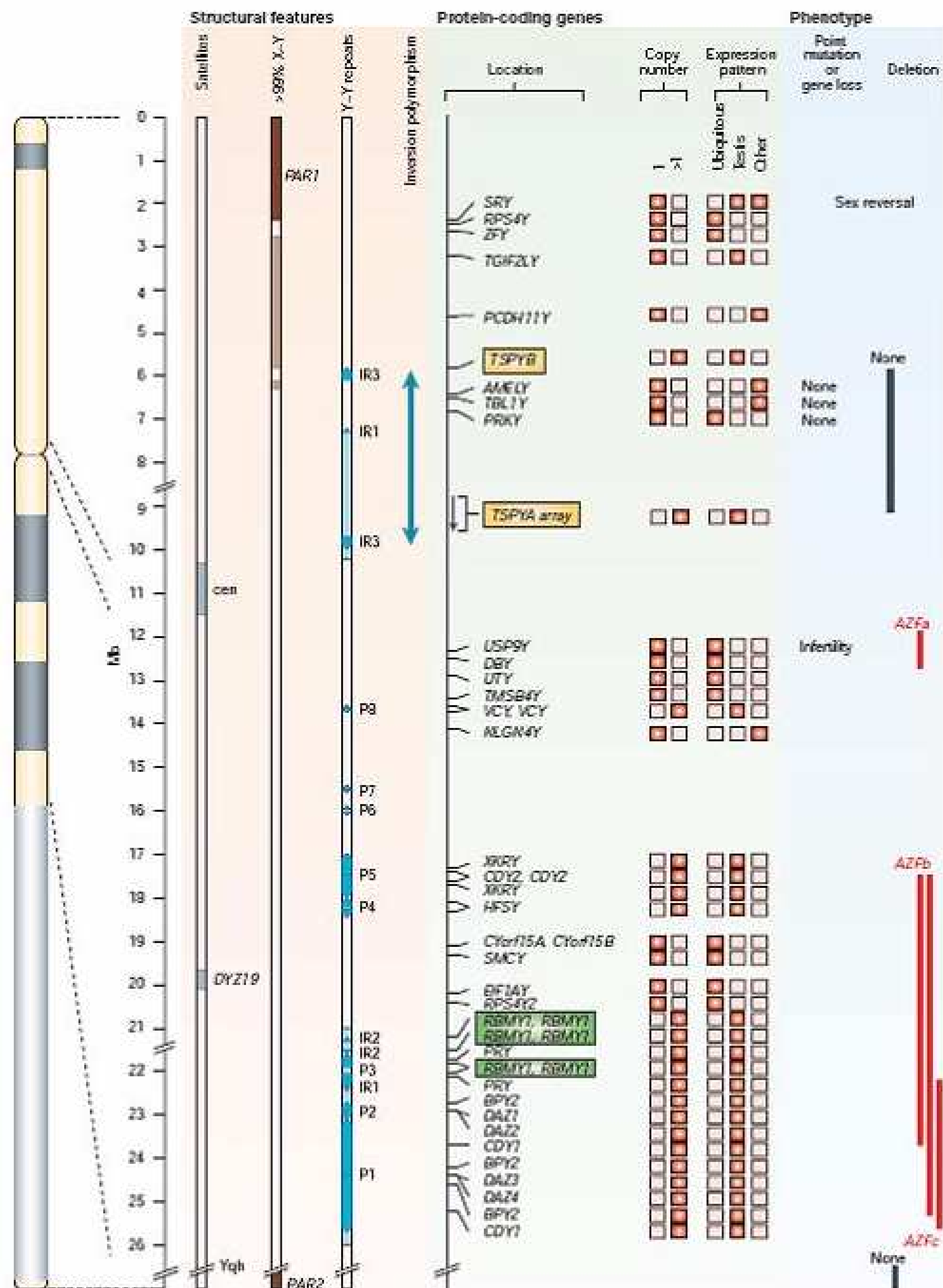


Figure 2. Y-Chromosome ideogram. From left to right: citogenetic features of the chromosome, structural features, location for Y-specific protein-coding genes and the phenotypes associated with gene inactivation or loss. Taken from Jobling and Tyler-Smith (2003)

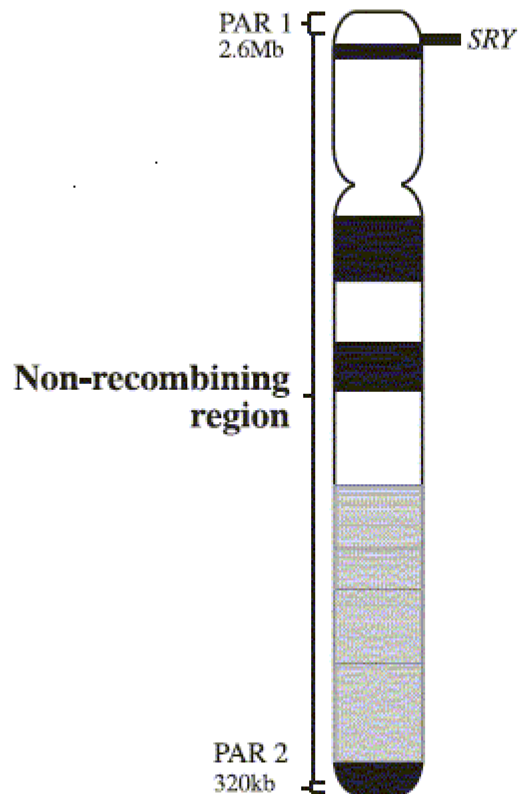


Figure 3. Ideogram of the Y Chromosome showing the locations of the pseudosomal regions (PAR), the non-recombinaing region (NRY), and the testis determining gene, SRY. Obtained from Hurles and Jobling (2001).

2.2 NRY Polymorphisms

The Y chromosome is thought to have a mutation rate slightly higher than that of autosomal loci (Hurles et al. 2003) and has been found to contain the same types of polymorphic loci as found on the other chromosomes (Jobling and Tyler-Smith 1995; de Knijff et al. 1997; Underhill et al. 1997). Y chromosome markers are best classified by their mutational rate which helps distinguish between so-

called ‘unique’ mutations events that can be considered to have occurred once in human history and those that are likely to be recurrent (Jobling and Tyler-Smith 1995).

Bi-allelic markers include SNPs and certain insertion-deletion (indels) events considered to be rare and therefore unique. The mutation rate for SNPs markers is considered to be an average on the order of 2×10^{-8} per base per generation (Nachman and Crowell 2000). Indels occur at a rate ten times slower and include LINE and SINE insertions, the presence of which always correspond to the derived state. Minisatellites and microsatellites have much higher mutation rates and thus are more frequent among multiallelic markers (Hurles et al. 2001).

SNPs are changes in DNA of a single nucleotide for a different one, resulting in the formation of a different variant or allele. These polymorphisms can sometimes create or destroy restriction sites or short DNA sequences known as Restriction Fragment Length Polymorphisms (RFLP) (Jones 2004). RFLPs are often referred to as biallelic markers because they distinguish between two alleles, one before the mutation (ancestral) and other after the mutation (derived). The group of haplotypes sharing and defined by evolutionary stable binary markers is known as a haplogroup.

Indels, mostly deletions in regions of Y-specific genes, have been found to be related to many diseases leading to male infertility among others. However, not all deletions affect male fertility; sometimes they persist over generations and are sufficiently common to be considered as polymorphisms that define haplogroups. An example is the first described Y-chromosome polymorphism, a

2kb deletion known as the 12f2 marker (Casanova et al. 1985), which defines haplogroup J.

Multiallelic markers include microsatellites and minisatellites. Microsatellites or short tandem repeats (STRs) are very small DNA sequences of 2-5 nucleotides in length that are repeated several times while minisatellites are repeats of DNA sequences of 8-100 base pairs. In some, but not all, the number of repeats is variable and several versions of different lengths can be found. Thus, multiallelic loci can generate various haplotypes within each haplogroup (Jones 2004).

Many specific polymorphisms have been used to construct informative haplogroups that are specific to geographical regions and to propose possible historic population movements (Jones 2004). Many of these have been utilized in evolutionary studies (Jobling and Tyler-Smith 1995; Underhill et al. 2000, 2003; Hammer and Zegura 2002), forensics (Jobling et al. 1997), medical genetics (Jobling and Tyler-Smith 2000) and genealogical reconstruction (Jobbling 2001). Biallelic markers which are slow evolving and considered unique events are very useful in evolutionary studies. Microsatellites are the markers of choice for paternal casework and criminal investigations (Santos et al. 1999). Minisatellites can be used in paternity and forensic casework as well as evolutionary studies; however their use in routine laboratory work is more complicated. These are reasons why in Y-chromosome studies STRs are widely used while minisatellites have been used only in some investigations (e.g. Jobling et al. 1998; Bao et al. 2000; Jin et al. 2003). The combined use of biallelic and microsatellites can help

characterize the variability of certain haplogroups. For example, Scozzari et al. 1999 combined the use of biallelic markers and STRs in order to infer affinity among African populations.

2.3 NRY Nomenclature

Researchers have employed a number of SNP and STR loci to define paternal lineages. Many of these researchers used at least 7 different nomenclature systems: (α) Jobling and Tyler-Smith (2000) and Kaladjieva et al. (2001); (β) Underhill et al. (2000); (γ) Hammer et al. (2001); (δ) Karafet et al. (2001); (ε) Semino et al. (2000); (ζ) Su et al. (1999); and (η) Capelli et al. (2001), making it very difficult to compare results. Hence, the Y Chromosome Consortium (2002) developed a hierarchical nomenclature system that unified all previous nomenclatures and allowed the inclusion of additional mutations and haplogroups yet to be discovered. Using this new nomenclature they constructed a comprehensive NRY phylogenetic tree (Figure 4); a diagram that represent the evolutionary relationships between lineages (Jobling and Tyler-Smith, 2003). According to the new nomenclature as it is described by the YCC (2002):

- Major clades are identified by capital letters (A-R) which constitute the front symbols of all subsequent subclades.
- Letter Y was assigned to the most inclusive haplogroup comprising all haplogroups from A to R.
- Paragroups are lineages that are not defined by the presence of a derived marker and are indicated by an asterisk. For example P*

represents chromosomes belonging to clade P but not to its known sub-clades.

- The nomenclature system allows the union of two letters for all clades that share a derived state. For example, clade DE includes all chromosomes within haplogroups D and E which share the derived state of YAP.
- Subclades nested within each major haplogroup defined by a capital letter are named using an alternating alphanumeric system. For example, within haplogrup Q, there are six basal haplogroups named Q1, Q2, Q3, Q4, Q5, and Q6, and the underived paragroup becomes Q*.
- Nested clades within each of these haplogroups are named in similar way, except that lower-case letters are used instead of numerals. For example, within haplogroup Q3, there are basal haplogroups named Q3a, Q3b, Q3c, and Q3d, and the underived paragroup becomes Q3*.
- The naming system continues to alternate between numerals and lower-case letters until the very last branches are labeled, thus the names of each haplogroup contain information of its location on the tree.
- Haplogroups can also be named by mutation by using clade letter followed by a "-" symbol and then the name of the SNP that defines it. For example haplogroup R1b1 (name by lineage) can also be named R-P25 (see Figure 1).

- When not all markers within a clade are typed, an “x” is used for excluding followed by the lineages that have been shown to be absent. This system could be applied for both lineage-based and mutation-based nomenclatures. For example, any 92R7 derived chromosome ancestral for P25 would be named P(xR1b1) or P-92R7(xP25).

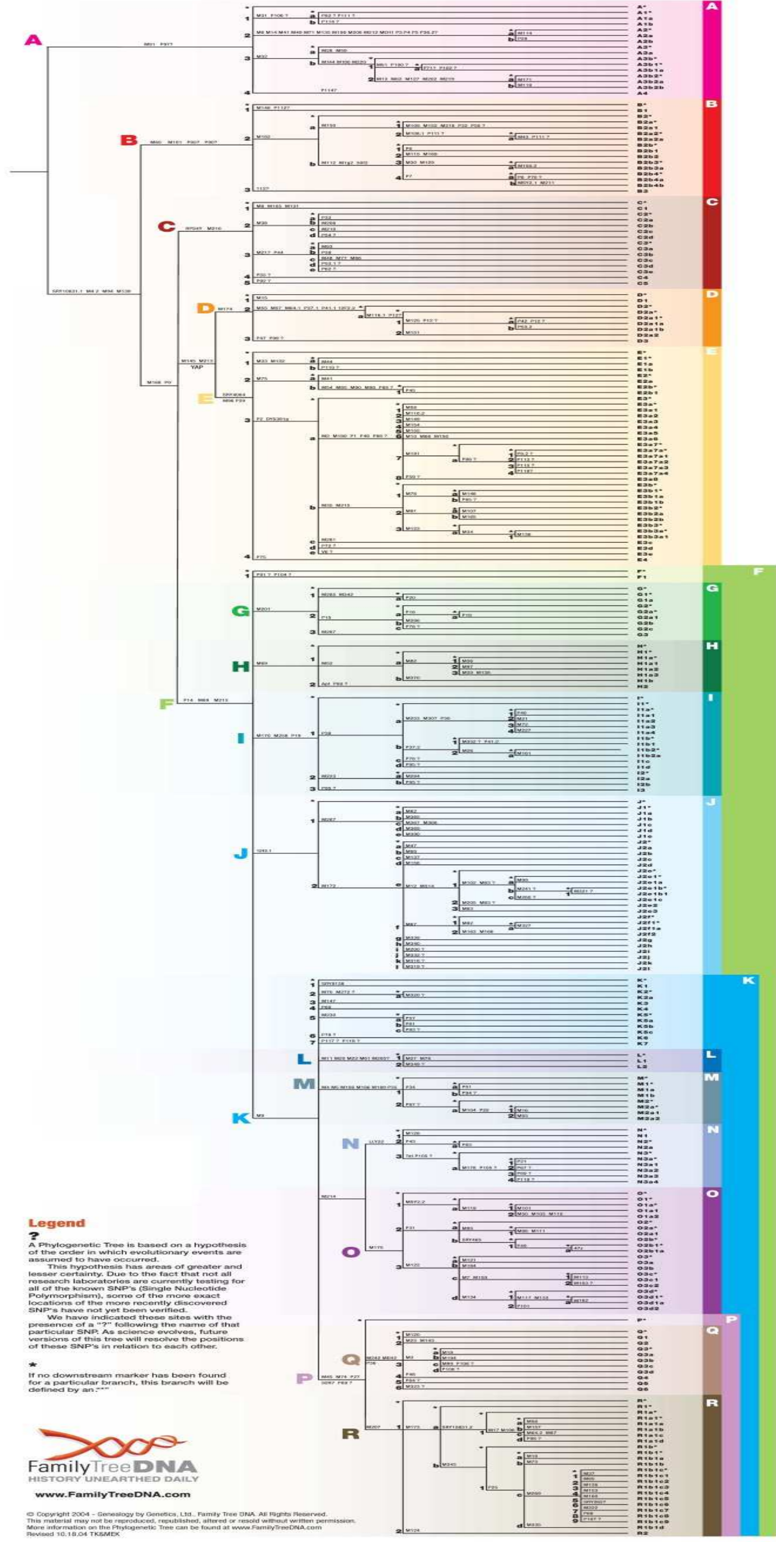


Figure 4. The Phylogenetic tree of binary NRY haplogroups. Obtained from the Family Tree DNA at: http://www.familytreedna.com/haplotree.html#top.

2.4 Peopling of the Americas

For many years researchers have used Y chromosome polymorphisms to answer questions about the peopling of the Americas. Where, when, and how have been the main focus and motivation for these investigations. Initial Y chromosome analysis found one haplogroup (Q3-M3) at high frequencies and was thought to be the single founder Native American lineage (Underhill et al. 1996). The descendent of haplogroup P was interpreted as being indicative of a unique migratory wave (Underhill et al., 1996; Bianchi et al., 1998; Santos et al., 1999) to the continent which was further supported by the finding of Ruiz-Linares et al. (1999), which indicated that the ancestral M3 allele was Native American in origin. However, the same year Bergen et al. identified a mutation, $RPS4Y_{711}$, which defines haplogroup C, is restricted to eastern Asia and America, and marked a Native American founder lineage outside P-M45. Consequently, some challenged the proposal of a single migratory wave to the Americas and suggested that there were two major migrations from East Asia into the New World that gave rise to ancestral Amerindians (Karafet et al., 1999; Lell et al., 2002; Bortolini et al., 2003; Schurr and Sherry, 2004). The first migration occurred 20,000 to 15,000 calendar years before present (cal BP) from Southern Central Siberia, extending towards South America, and introducing haplogroups P ancestral to the major Native American founding lineage, haplogroup Q3. The second migration initiated from the Lower Amur/Sea of Okhotsk region, brought a differentiated haplogroup P with its associated variant R1-M173 and haplogroup C-M130 to only North and Central America (Lell et al. 2002). The justification

for this second migration was explained on the basis of P-M45 lineage being differentiated into two major subdivisions: M45a, which is ancestral for Q3-M3 and is found throughout the Americas, and M45b, which incorporates the R1-M173 variant and is concentrated in North and Central America. The problem is that P chromosomes are found in both Native Americans and Europeans.

In 2003, a new SNP, M242, which can distinguish Native American from European haplogroups within the P clade, was discovered (Bortolini et al. 2003; Seielstad et al. 2003). This SNP arose in Central Asia prior to the emergence of the M3 mutation (Schurr and Sherry, 2004). Its derived allele (M242-T) is found in all Native American P chromosomes (Bortolini et al. 2003) but in very few Asians, thus defining Q, an almost exclusively Native American clade.

2.5 Relevant Y chromosome markers and the haplogroups they define.

For the purpose of this study the following SNPs were used to determine the haplogroup frequencies of Y chromosomes in Puerto Rico.

92R7

In 1994, Mathias and his colleagues reported a G→A transition that defines haplogroup P, which is thought to have originated in Central Asia 35,000 to 40,000 years ago and is ancestral to haplogroups Q and R. Haplogroup P can also be described by SNPs M45, M74, and P27.

M207

In 2001, Underhill and his colleagues reported an A→G transition residing in Intron 3a of the Ubiquitously Transcribed Tetratricopeptide repeat gene (UTY1 ex03). This mutation defines haplogroup R which is thought to have originated

30,000 years ago and is mainly represented by two lineages: R1a and R1b. Haplogroup R can also be described by SNP M306.

P25

In 2000, Hammer and his colleagues reported a C→A transversion at the DYS194 locus that originated approximately 10,000 ± 5,100 years ago. This mutation defines haplogroup R1b1 which is a subgroup of R1b, the most frequent haplogroup in western European populations (Adams et al. 2006). At least this was the case until it was found that P25 is a paralogous sequence variant rather than a SNP (Adams et al. 2006). Three copies of the P25 sequence lie within the giant palindromic repeats on Yq, and one copy has undergone a C to A transversion that defines haplogroup R1b (designated C/C/A). However, reverse conversion has been shown to occur where the derived P25 A-allele is replaced by the ancestral C-allele (yielding C/C/C). Because of its inherent instability, it is suggested that P25 be used with caution and perhaps be replaced with the more reliable binary maker M269.

SRY10831b

In 1998, Hammer and his colleagues reported the occurrence of two mutational events residing at the SRY gene position 10831; an A→G transition (SRY10831a) and a G→A reversion (SRY10831b). SRY10831b defines European haplogroup R1a which is believed to have originated in Eurasia approximately 10,000 to 15,000 years ago.

M242

In 2003, Seielstad et al. reported a C→T transition residing in intron 1(IVS-866) of the DEAD-box RNA helicase Y (DBY) gene. The M242 mutation arrived after M45/M74/92R7 but before M3 and is believed to have occurred 15,000 to 18,000 years ago in Central Asia and entered the Americas soon after (Schurr et al. 2004). The M242 marker defines haplogroup Q which is ancestor to many Siberians and almost all of the indigenous peoples of the Americas (through its subgroup Q3). This haplogroup is surprisingly diverse. At least six primary subclades have been sampled and identified in modern populations.

M3

In 1996, Underhill and his colleagues reported a C→T transition residing at the DYS199 locus which was only found in Native American populations. This mutation is believed to have occurred in North America approximately 10,000 to 15,000 years ago. The M3 marker defines haplogroup Q3 which is strictly associated with Native American populations.

MATERIALS AND METHODS

3.1 POPULATION SAMPLE

A total of 99 samples were obtained from volunteers at the University of Puerto Rico at Mayagüez. Mouthwash samples were collected from all subjects following written informed consent (see appendix 1 for consent form). All volunteers rinsed their mouths vigorously with 10 mL of mouthwash for 45 seconds and then spat into a sterile cup. The solution was then transferred into a 15 mL conical tube for DNA extraction.

3.2 DNA EXTRACTION

DNA samples were prepared according to a protocol kindly provided by Bert Ely (pers. comm.) at the University of South California. Each sample was centrifuged for 10 minutes at 5000 rpm in a table-top centrifuge. The supernatant was then discarded and the pellet resuspended in 200 µL of DNAzol (Invitrogen) and 10 µL of Proteinase K (20 mg/mL) (QIAGEN) and incubated at room temperature overnight. Then, the samples were transferred into 1.5mL microfuge tubes and centrifuged at 14,000 rpm for 10 minutes. Supernatants were transferred into new 1.5 mL microfuge tubes, to which 200 µL of ice cold 100% EtOH was added. In order to help the DNA precipitate, tubes were inverted from 5 to 8 times and placed on ice for 2 minutes. The tubes were centrifuged at 14,000 rpm for 5 minutes to allow the DNA to precipitate. The supernatants were discarded and to assure clean pellets, the samples were washed twice with ice cold 75% EtOH. To remove the excess EtOH the tubes were inverted and left on a paper towel with the caps open. After all of the EtOH evaporated, the pellets

were resuspended in 200 µL of TE Buffer 1X (10mM Tris-HCl pH 8.0 and 10mM Na_2EDTA pH 8.0) and stored at -20°C.

3.3 IDENTIFICATION OF SAMPLES CONTAINING THE 92R7T ALLELE AND HAPLOGROUP DETERMINATION

By using molecular techniques such as polymerase chain reaction (PCR), restriction fragment length polymorphism (RFLP) and DNA sequencing, samples with the 92R7T allele were identified and then classified into haplogroups by the analysis of the following SNPs: P25, M242, M3, SRY10831, and M207. These were tested hierarchically (Figure 5). All samples were tested at 92R7 first. Those containing the derived (92R7T) allele were then tested at P25. Those not shown to have the derived state of P25 and thus not belonging to haplogroup R1b1 were tested for M242 (dbSNP accession number ss9805824). Samples having the derived state at this locus were tested at M3 to determine their belonging to haplogroup Q or Q3. Those not shown to have the derived state of M242 were tested for SRY10831. Samples that did not have the derived state at SRY10831 and thus not belong to haplogroup R1a were tested at M207. Derived samples were classified as R(xR1a, R1b1) and those ancestral as paragroup P*.

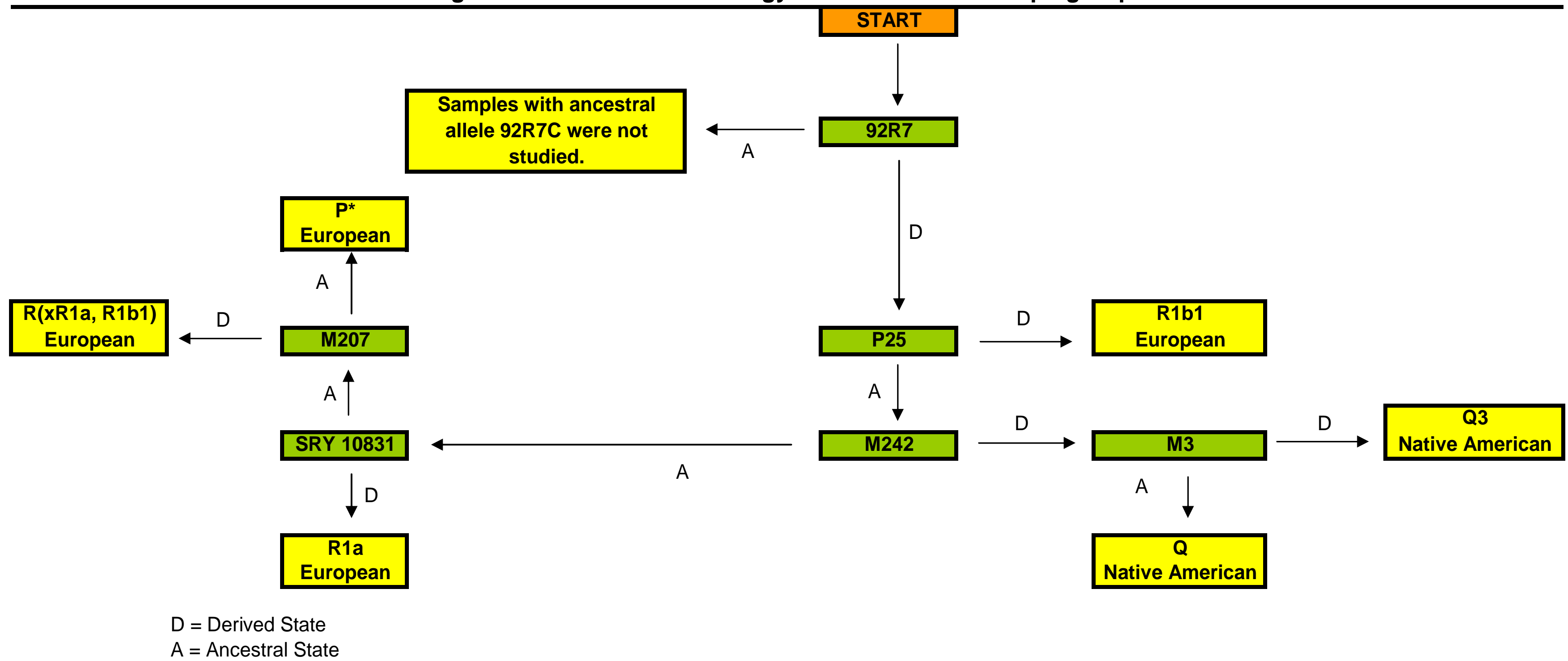


Figure 5. Identification strategy of Y chromosome haplogroups showing hierarchical order for sample testing.

3.3.1 PCR

All PCR reactions were performed at 1X PCR Buffer (10mM Tris-HCl pH 8.3, 50mM KCl, 1.5mM $MgCl_2$, 0.001% gelatin), 2.5mM $MgCl_2$, 400µM dNTP, 1µM each primer, 7µl of DNA sample, 0.2mg/ml BSA (Bovine Serum Albumin) and 2.5 units of Taq DNA polymerase in a total volume of 25µl. Primers and amplification cycles used for all sites tested and their respective products are shown in Table 1. An agarose gel electrophoresis was performed to verify all PCR products.

3.3.2 RFLP

The restriction digestions were performed using 15 µl of the amplification reaction, 2 µl of the 10x buffer recommended by the manufacturer (New England Bio Labs) and 1 µl of the restriction enzyme (10 to 20 U/ul) in a total volume of 20µl. In addition, 0.1 µl BSA (10mg/ml) per every 10 µl of reaction was added. Only 92R7, P25, M3, and SRY 10831 were subjected to restriction digestion. All digestions were placed at 37ºC overnight to allow plenty of time for the reaction to occur. Then agarose gel electrophoresis was performed. (See Table 2 for enzymes, digestion product and % of agarose gels used.)

3.3.3 DNA Sequencing

To determine the sequences of M242 and M207, PCR fragments were purified using the High Pure PCR Product Purification Kit (Roche Molecular Biochemicals) as instructed by the manufacturer. Fifty µl of each purified PCR product were sent together with 10 ul of the respective sequencing primers at 1µM concentration to the New Jersey Medical School Molecular Resource Facility for

automated sequencing.   Then sequences were analyzed with Chromas (Technelysium) software to identify polymorphisms (M242C, M242T, M207A and M207G).

Table 1. PCR conditions to Amplify Tested Sites

Sites Tested	Primers: Sequence (5' - 3')	PCR: Cycle	PCR: Product
92R7	F GAC CCG CTG TAG ACC TGA CT R GCC TAT CTA CTT CAG TGA TTT CT	Start at 94ºC 2:30 min. followed by 35 cycles at 94ºC 1 min.; 59ºC 1 min.; 72ºC 1 min. Finish with 10 min at 72ºC. Hold at 4ºC	3 fragments of 709 bp
P25	F CTC AAA TAC ACA AAA CCA GG R TCA AGA CAA AGG CTA AAG C	Start at 94ºC 2:30 min.; followed by 35 cycles of 94ºC 45 sec.; 49ºC 1 min.; 72ºC 1 min. Finish with 10 min at 72ºC. Hold at 4ºC	3 fragments of 490 bp
M242	F CAC TGA CGA CGT ATT AAC G R CCT AGA ACA ACT CTG AAG C	Start at 94ºC for 3:00 min.; followed by 35 cycles at 94 for 30 sec.; 55ºC for 45 sec; 72ºC for 1 min. Finish with 10 min at 72ºC. Hold at 4ºC	398 bp
M3	F TAA TCA GTC TCC TCC CAG CA R AGG TAC CAG CTC TTC CCA ATT	Start at 95ºC for 2:30 min. followed by 35 cycles at 94ºC for 40 sec.; 59ºC for 30 sec.; 72ºC for 40 sec. Finish with 10 min at 72ºC. Hold at 4ºC	202 bp
SRY 10831b	F TCT GAC TCT TTG GTT CAC CA R AAG TGT TGG TTC TCC TGT A	Start at 94ºC for 2:30 min. followed by 35 cycles at 94ºC for 45 sec.; 49ºC for 1 min.; 72ºC for 1 min. Finish with 10 min at 72ºC. Hold at 4ºC	310 bp 190 bp
M207	F AGG AAA AAT CAG AAG TAT CCC TG R CAA AAT TCA CCA AGA ATC CTT G	Start at 94ºC for 3:00 min. followed by 35 cycles at 94ºC for 30 sec.; 59ºC for 30 sec.; 72ºC for 1 min. Finish with 10 min at 72ºC. Hold at 4ºC	422 bp

Table 2. RFLP Conditions for PCR Products

Locus	PCR Product	RFLP			
			Product		
		Enzyme	Ancestral State (bp)	Derived State (bp)	Gel
92R7	3 fragments of 709 bp	*Hind*III	709 512 197	709	1% agarose in TBE 1X
P25	3 fragments of 490 bp	*HpyCH4* V	270 118 102	372 270 118 102	2.5% agarose in TBE 1X
M3	202 bp	*Mfe* I	118 21	202	3% agarose in TBE 1X
SRY10831b	310 bp 190 bp	*Dra* III	153 83	310 190	3% agarose in TBE 1X

RESULTS

Following the established hierarchical order (Figure 1), all 99 samples were first analyzed to determine if they possessed the derived state, 92R7T. Results of the first test showed that 28 samples had the derived state 92R7T, 38 had the ancestral state 92R7C and the remaining 33 could not be determined at this time. For the purpose of this discussion, the results samples will be labeled as A, B and C according to the results of the 92R7 tests (Figure 6). In group A are all 92R7T samples, in group B are all 92R7C samples and in group C are all of those samples for which 92R7 allele could not be determined. Since only those samples with the 92R7T allele are of interest, all samples in group B were eliminated from the study. Only groups A and C were further tested for all other SNPs starting with P25.

In group A, out of the 28 92R7T samples, 23 had the derived state for P25 and thus belong to haplogroup R1b1, 4 samples (B2, D5, G3 and I2) had the ancestral state, and only one sample (B5) could not be determined. After many unsuccessful attempts to test for all of the other SNPs, sample B5 was simply classified as P. Samples B2, D5, G3 and I2 were then tested for M242. Out of all four samples only D5 had the derived state at M242, and thus was further tested at M3 and determined to be M3 derived, for which it was classified as belonging to Native American haplogroup Q3. Samples G3 and I2 had the ancestral state for M242 and were therefore tested for SRY10831b. Only sample G3 had the derived state for SRY10831b and was classified as belonging to

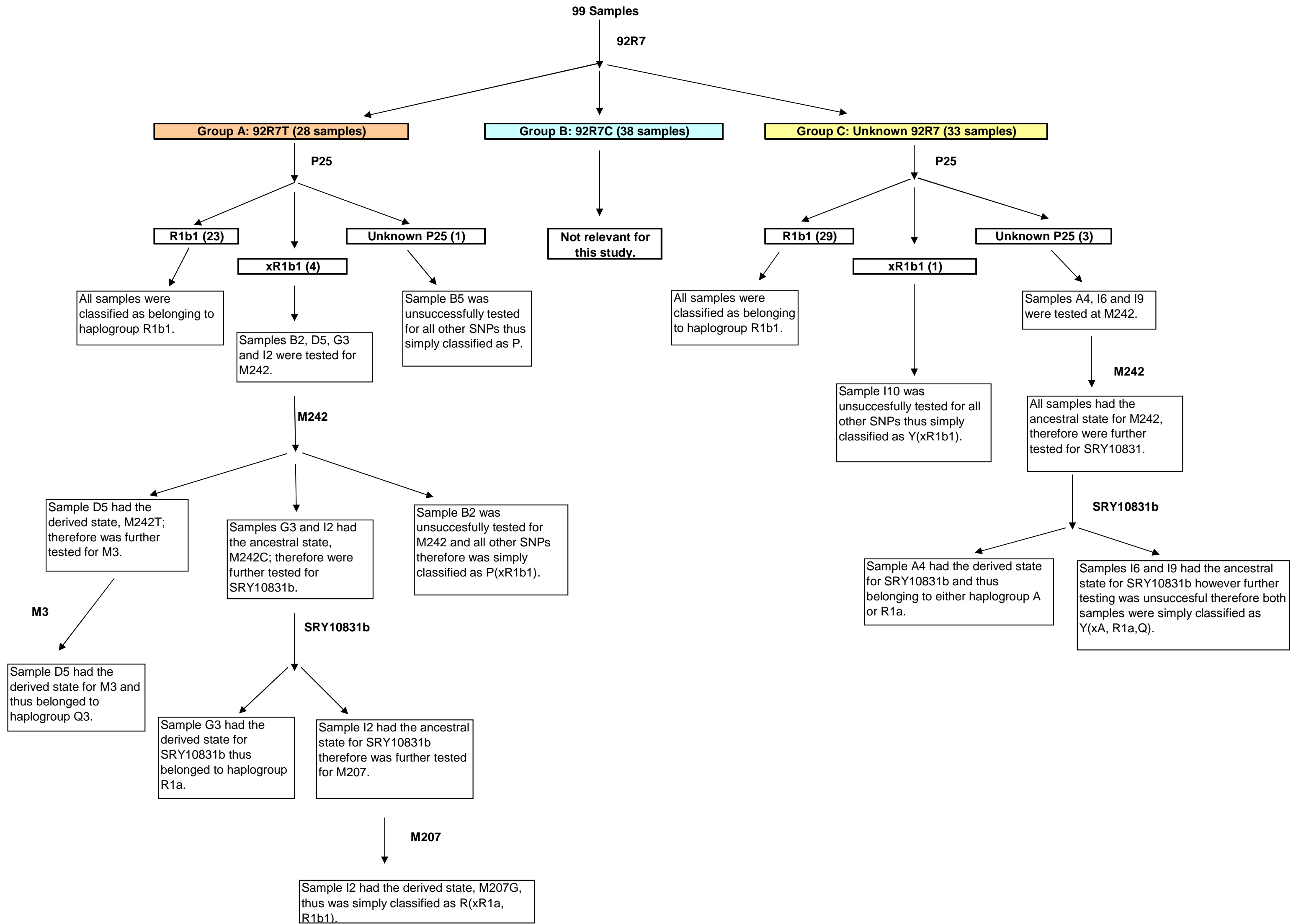
haplogroup R1a. Since sample I2 was ancestral for SRY10831b, it was then tested for M207 and found to have the derived state. Thus, it was classified as R(xR1a, R1b1) since I2 was derived for haplogroup for haplogroup R but ancestral to R1a and R1b1. Unfortunately sample B2 was not successfully tested for M242 or any other SNPs. However, using the information already obtained B2 was classified as P(xR1b1) since it was derived for haplogroup P but ancestral to R1b1.

After testing the 33 unknown samples in group C for P25 it was determined that 29 of the samples had the derived state and therefore belonged to haplogroup R1b1 and must have the 92R7T allele. Of the remaining 4 samples (A4, I6, I9 and I10), only I10 was ancestral for P25. The other 3 could not be determined. I10 was tested multiple times unsuccessfully for all other SNPs and as a result classified as Y(xR1b1). Samples A4, I6 and I9 were tested for M242 and determined that all of them had the ancestral state therefore were then tested for SRY10831. Only sample A4 had the derived state for SRY10831 and given that it was unknown for 92R7 it was classified as belonging to either haplogroup A or R1a. 92R7 is needed to differentiate between the two haplogroups since SRY10831 is used to determine both. Samples I6 and I9, although they had the ancestral state for SRY10831b, further testing was not successful. Hence they were classified as Y(xA,Q,R1a).

In summary, 38/99 samples were eliminated since they possessed the 92R7C allele and 61 were further analyzed to determined their patrilineal

biological ancestry. Results showed that a total of 57 of the samples had the 92R7T allele, from which 52 belong to the haplogroup R1b1, 1 to R1a, 1 to Q3, 1 to R(xR1a, R1b1), 1 was simply classified as P(xR1b1), 1 (1.7%) as 92R7T or P and lastly, 1 was classified as belonging to either haplogroups A or R1a. Only 3 samples could not be identified. Overall, the vast majority of the 92R7T Y chromosomes were classified into well defined haplogroups that leave no doubt of their biological ancestry. Approximately 88.5% of the 92R7T Y chromosomes in Puerto Rico are of European origin (R1a, R1b1, R*/R1*) while only about 1.7% are of Native American origin (Q3).

Figure 6. Haplogroup determination of samples following hierarchical order



DISCUSSION

In this study, the frequency of Y chromosome haplogroups within the P clade was determined in Puerto Rico in an effort to better understand the assortment of these Y chromosomes between those of Native American origin and those European. Previous study show that most Puerto Rican men have the derive state 92R7T which defines the P clade however, a substantial number of these men could not be classified into any of the Europeans nor the Native American haplogroup that are included within this clade, thus remaining classified simply as 92R7T Y chromosomes. Of those who were successfully identified it was observed that an overwhelmingly majority (73.7%) of the Puerto Rican men are West Eurasian in origin, while 25.1% are Sub-Saharan African and 1.2% Native American

In this study, the state of the 92R7 marker was determined for 95 of the 99 samples, either directly or by the identification of the samples as belonging to a clade within haplogroup P. Of these, 57 (57.6%) showed the presence of the 92R7T allele. This is a frequency higher than that found for Puerto Rico in the previous study, but the difference is not significant. The ancestry of 55 of these 57 92R7T samples was determined, 54 of which were shown to be of European origin and only one Native American. The haplogroups of the 54 92R7T samples of European origin were 52 R1b1, one R1a and one R*/R1*. The Native American sample belonged to haplogroup Q3.

There were six samples not shown to have the 92R7C allele for which the ancestry is unknown. However, three of these were shown not to belong to a Native American haplogroup. These were samples I6 and I9, classified as Y(xA,R1a,Q) and A4, classified as either A (Sub-Saharan African) or R1a (European). Two of the remaining

three samples (B2 and B5) were shown to have the 92R7T allele. These were classified simply as P and as P(xR1b1) respectively. The remaining sample (I10) was classified as Y(xR1b1).

It is important to point out that this study was not designed to compare any other haplogroups outside the P clade like the previous study where African and other European haplogroups were included. However, the results of this study are comparable to the previous one since both show that Europeans have contributed the most to the patrilineal biological ancestry in Puerto Rico while Native American have contributed the least. In order to determined the frequency of African and other European haplogroups within the Puerto Rican male population further analysis of the samples with the 92R7C allele is necessary.

According to history the European paternal ancestry in Puerto Rico was expected to be high since is likely coming from Spanish colonizers that came to the island over 500 years ago. One can conclude this to be accurate since from the European haplogroups found in our sample; R(xR1a, R1b1), R1a and R1b1, the latter which is a subgroup of the R1b, the most frequent haplogroup in western European populations (Adams et al. 2006) was the most frequent haplogroup with a frequency of 52.5%.

Also as expected and according to history, the Native American patrilineal biological ancestry has contributed the least. Although results are quite the opposite from those obtained from mtDNA analysis were the Native Americans maternal contribution was 61.3%, this study is evidence that there is indeed paternal Native Amerindian ancestry in Puerto Rico as well. It should be stressed that Native American Y-chromosomes belonging to haplogroup C-M130 have never been found south of

Central American (Lell et al. 2002), and thus Puerto Rican samples shown not to belong to the only other Native American haplogroup Q are regarded as non-Native American in origin.

From the 38.4% of the samples with the 92R7C allele, one can expect this cluster to have a combination of other European haplogroups but mainly Sub-Saharan African haplogroups likely from those who were brought to the island as slaves during the 16th century. As it was shown in the previous study where Sub-Saharan African Y chromosome haplogroups represented 25.1% of the Puerto Rican Y chromosomes; second only to the Europeans haplogroups.

Another important aspect that should be pointed out is the effectiveness of the hierarchical strategy used in this study. After testing for only two SNPs (92R7 and P25) over 90% of the samples were identified. However, caution should be exercised when testing for 92R7 since PCR amplification was not achieved for almost a third of the samples. Perhaps a better approach will be to invert the order by testing all samples for P25 first and then for 92R7 all of those proven to be ancestral for P25. Although there were some problems with the identification of the remaining samples it was mainly due to low concentration of DNA or perhaps the presence of PCR inhibitors in the sample.

While this study is merely preliminary and is neither randomized nor representative of the Puerto Rican population, an overwhelming majority of European haplogroups and a very low presence of Native American haplogroups are to be expected even in a randomized and representative population sample. Only 92R7T Y chromosomes stand a chance of being of Native American origin in Puerto Rico, but

according to this study, if a Puerto Rican is said to have a 92R7T Y chromosome, chances are this person belongs to a European haplogroup rather than Native American.

CONCLUSION

The results obtained in this study reveal a strong patrilineal contribution of European population to modern Puerto Rican and a very poor Native American contribution. Although this study is neither a randomized nor representative of the Puerto Rican population it is expected that the overwhelming majority of Puerto Rican men belong to European haplogroups rather than Native American.

Lastly, this study also proved that the hierarchical strategy employed for Y chromosome haplogroup identification was effective in classifying 92R7T Y chromosomes into well defined haplogroups leaving no doubt of their biological ancestry. Determining P25 first for all of those 97R7T Y chromosomes significantly decrease the number of samples to be tested for all other SNPs minimizing time and cost.

RECOMMENDATIONS

1. Upcoming Y chromosome research must include coverage of other haplogroups outside of the P clade not covered in this study, such as African haplogroups, which historically have significantly contributed to the Puerto Rican population. Samples with 92R7C allele need to be studied and classified into well defined haplogroups in order to obtain a complete data set that can be used in combination with mtDNA analysis to obtain a more comprehensive picture of the human ancestry in Puerto Rico.
2. Due to the unreliability of the now known to be a paralogous sequence variant, P25 should be replaced by a more reliable SNP like M343 which defines haplogroup R1b, leaving no doubt of patrilineal biological ancestry.
3. Lastly, due to the inconsistency with the 92R7 PCR test, perhaps a better approach will be to replace it with M45 and invert the order by testing all samples for P25 or M343 first and then for M45.

CITED LITERATURE

Adams, S.M., King, T.E., Bosch, E., and Jobling, M.A. 2006. The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion. *Forensic Sci Int* 25:159(1):14-20.

Bao, W., Zhu, S., Pandya, A., Serjal, T., Xu, J., Shu, Q., Du, R., et al. 2000. MSY2: a slowly evolving minisatellite on the human Y chromosome which provides a useful polymorphic marker in Chinese populations. *Gene* 244:29-33.

Bianchi, N.O., Catanesi, C.I., Bailliet, G., Martinez-Marignac, V.L., Bravi, C.M., Vidal-Rioja, L.B., Herrera, R.J., and Lopez-Camelo, J.S. 1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet 61:1862-1871.

Bortolini, M.C., Salzano, F.M., Thomas, M.G., Thomas, M.G., Stuart, S., Nasanen, S.P.K., Bau, C.H.D., Hutz, M.H., Layrisse, Z., Petzl-Erler, M.L., Tsuneto, L.T., Hill, K., Hurtado, A.M., Castro-de-Guerra, D., Torres, M.M., Froot, H., Michalski, R., Nymadawa, P., Bedoya, G., Bradman, N., Labuda, D., and Ruiz-Linares, A. 2003. Y-Chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73:524-539.

Capelli, C., Wilson, J.F., Richards, M., Snumpf, M.P., Gratrix, F., Oppenheimer, S., Underhill, P.A., et al. 2001. A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am J Hum Genet* 68:432-443.

Cassanova, M., Leroy, P., Boucekkine, C., Weissenbach, J., Bishop, C., Fellous, M., Purrello, M., Friori, G., and Siniscalco,M. 1985. A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distances. *Science* 230:1403-1406.

de Knijff, P., Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., et al. 1997. Chromosome Y microsatellites: population genetic and evolutionary aspects. Int *J Legal Med* 110:134–140.

Fernández, S., Paracchini, S., Meyer, L. H., Floridia, G., Tyler-Smith, C. and Vogt, P. H. 2004. A Large AZFc Deletion Removes *DAZ3/DAZ4* and Nearby Genes from Men in Y Haplogroup N. Am J Hum Genet 74:180-187.

Flores, G., Maca-Meyer, N., González, A.M., Oefner, P.J., Shen, P., Pérez, J.A., Rojas, A., Larruga, J.M., and Underhill, P.A. 2004. Reduced Genetic Structure of the Iberian Peninsula Revealed by Y-chromosome Analysis: Implications for population demography. *European J Hum Genet* 12:855-863.

Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R., and Zegura, S.L. 1998. Out of Africa and Back Again: Nested Cladistic Analysis of Human Y Chromosome Variation. Mol Biol Evol 15(4):427-441.

Hammer, M.F., Karafet, T.M., Redd, A.L., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H., and Zegura, S.L. 2001. Hierarchical Patterns of Global Human Y-Chromosome Diversity. *Mol Biol Evol* 18(7):1189-1203.

Hammer, M.F., Redd, A.J., Wood, E.T., Bonner, M.R., Jarjanazi, H., Karafet, T., Dantachiara-Benerecetti, S., Oppenheim, A., Jobling, M.A., Jenkins, T., and Bonne-Tamir, B. 2000. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. PNAS 97(12):6769-6774.

Hammer, M.E., and Zegura, S.L. 2002. The human Y chromosome haplogroup tree: Nomenclature and phylogeography of its major divisions. *Annu Rev Anthropol* 31:303-321.

Hurles, M.E., and Jobling, M.A. 2003. A singular chromosome. *Nat Genet* 34:246-247.

Hurles, M.E., and Jobling, M.A. 2001. Haploid chromosomes in molecular ecology: lessons from the human Y. *Molecular Ecology* 10:1599-1613.

Jin, Z.B., Huang, X.L., Nakajima, Y., Yukawa, N., Osawas, M., and Takeichi, S. 2003. Haploid allele mapping of Y-chromosome minisatellites, MSY1 (DYF155S1), to a Japanese population. *Leg Med (Tokyo)* 5:87-92.

Jobling, M.A. 2001. In the name of the father: surnames and genetics. *Trends Genet* **17**, 353-357.

Jobling, M.A., Bouzekri, N., and Taylor, P.G. 1998. Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellites, MSY1 (DYF155S1). *Hum Mol Genet* 7:643-653.

Jobling, M.A., Pandya, A., and Tyler-Smith, C. 1997. The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* 110(3):118-124.

Jobling, M.A., and Tyler-Smith, C. 1995. Fathers and sons: the Y chromosome and human evolution. *Trends Genet* 11(11):449-456.

Jobling, M.A., and Tyler-Smith, C. 2000. New uses for new haplotypes: the human Y chromosome, disease and selection. *Trends Genet* 16(8):356-362.

Jobling, M.A., and Tyler-Smith, C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4(8):598-612.

Jones, P.N. 2004. American Indian mtDNA and Y Chromosome Genetic Data: A Comprehensive Report of their Use in Migration and Other Anthropological Sudies.

Kalaydjieva, L., Calfell, F., Jobling, M.A., Angelicheva, D., de Knijff, P., Rosser, Z.H., Hurles, M.E., et al. 2001. Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur J Hum Genet* 9:97-104.

Karafet, T., Xu, L., Du, R., Wang, W., Feng, S., Wells, R.S., Redd, A.J., et al. 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 69:615-628.

Karafet, T.M., Zegura, S.L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., Wiebe, V., Griffiths, R.C., Templeton, A.R., and Hammer, M.F. 1999. Ancestral Asian Source(s) of New World Y-Chromosome Founder Haplotypes. Am J Hum Genet 64:817-831.

Lell, J.T., Sukernik, R.I., Starikovskaya, Y.B., Su.B., Jin, L., Schurr, T.G., Underhill, P.A., and Wallace, D.C. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 70:192-206.

Martínez-Cruzado, J.C., Toro-Labrador, G., Viera-Vera, J., Rivera-Vega, Michelle., Startek, J., Latorre-Esteves, M., Román-Colón A., Rivera-Torres, R., Navarro-Millan, I.Y., Gomez-Sanchez, E., Caro-Gonzalez, H.Y., and Valencia-Rivera, P. 2005. Reconstructing the Population History of Puerto Rico by means of mtDNA Phylogeographic Análisis. American Journal of Physical Anthropology 12:131-155.

Mathias, N., Bayes, M., and Tyler-Smith, C. 1994. Highly informative compound haplotypes for the human Y chromosome. *Hum Mol Genet* 3:115-124.

Nachman, M.W., and Crowell, S.L. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297-304.

Pena, S.D., Santos, F.R., Bianchi, N.O., Bravi, C.M., Carnese, F.R., and Rothhammer, F. 1995. A major founder Y-chromosome haplotype in Amerindians. *Nat Genet* 11:15-16.

Ruiz-Linares, A., Ortiz-Barrientos, D., Figueroa, M., Mesa, N., Munera, J.G., Bedoya, G., Vélez, I.V., Garcia, L.F., Perez-Lezaun, A., Bertranpetit, J., Feldman, M.W., and Goldstein, D.B. 1999. Microsatellites provide evidence for Y chromosome diversity among the founders of the New World.

Santos, F.R., Pandya, A., Tyler-Smith, C., Pena, S.D.J., Schanfield, M., Leonard, W., Osipova, L., Crawford, M.H., and Mitchell, J. 1999. The Central Siberian Origin for Native American Y Chromosomes. *Am J Hum Genet* 64:619-628.

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.

Scozzari, R., Cruciani, F., Malaspina, P., Santolamazza, P., Ciminelli, B.M., Torroni, A., Mediano, D., Wallace, D.C., Kidd, K.K., Olckers, A., Moral, P., Terrenato, L., Akar, N., Qamar, R., Mansoor, A., Mehdi, S.Q., Meloni, G., Vona, G., Cole, D.E.C., Cai, W., and Novelletto, A. 1997. Differential structuring of human populations for homologous X and Y microsatellites loci. *Am J Hum Genet* 61:719-733.

Scozzari, R., Cruciani, F., Santolamazza, P., Malaspina, P., Torroni, A., Sellitto, D., Arredi, B., Destro-Bisol, G., De Stefano, G., Rickards, O., Martinez-Labarga, C., Mediano, D., Biondi, G., Moral, P., Olckers, A., Wallace, D.C., and Novelletto, A. 1999. Combined use of biallelic and microsatellite Y-Chromosome polymorphisms to infer affinities among African populations. Am J Hum Genet 65:829-846.

Seielstad, M., Yuldsheva, N., Singh, N., Underhill, P.A., Oefner, P., Shen, P., and Wells, R.S. 2003. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. Am J Hum Genet 73:700-705.

Semino, O., Passarino, G., Oefner, P.J., Lin A.A., Abruzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiæ, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A.S., Cavalli-Sforza, L.L., and Underhill, P. 2000. The Genetic Legacy of Paleolithic Homo Sapiens in Extant Europeans. *Science* 290:1155-1159.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., and Goodfellow, P.N. 1990. A gene from the human sex-determining region encodes a protein with homology to conserved DNA-binding motif. *Nature* 346:240-244.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, K.m Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.F, Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.P, Waterston, R.H., Wilson, R.K., Rozen, S., and Page, D.C. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequences classes. *Nature* 423:825-837.

Su, B., Xiao, J., Underhill, P.A., Deka,R., Zhang, W., Akey, J., Huang, W., Shen, D., Lu, D., Lou, J., Chu, J., Tan, J., Shen, P., Davis, R., Cavalli-Sforza, L., Chakraborty, R., Xiong, M., Du, R., Oefner, P., Chen, Z., and Jin, L. 1999. Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last ice age. *Am J hum Genet* 65:1718-1724.

Underhill, P.A. 2003. Inferring Human History: Clues from Y-Chromosome Haplotypes. Cold Spring Harbor Symposia on Quantitative Biology. Vol LXVIII. Cold Spring Harbor Laboratory Press: 487-493.

Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., and Cavalli-Sforza, L.L. 1997. Detection of numerous Y chromosomes biallelic polymorphisms by denaturing high-performance liquid chromatography (DHPLC). *Genome Res* 7:996-1005.

Underhill, P.A., Jin, L., Zemans, R., Oefner, P.J., and Cavalli-Sforza, L.L. 1996. A pre-Colombian Y chromosome-specific transition and its implications for human evolutionary history. Proc Nalt Acad Sci USA 93:196-200.

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazón-Lahr, M., Foley, R.A., Oefner, P.J., Cavalli-Sforza, L.L. 2001. The Phylogeography of Y Chromosome Binary Haplotypes and the Origins of Modern Human Populations. *Annals of Human Genetics* 65:43-62.

Underhill, P.A., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonne-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J.R., Mehdi, S.Q., Seielstad, M.T., Wells, R.S., Piazza, A., Davis, MR.W., Feldman, M.W., Cavalli-Sforza, L.L., and Oefner, P.J. 2000. Y Chromosome sequence variation and the history of human populations. *Nat Genet* 26:358-361.

Y Chromosome Consortium. 2002. A nomenclatura System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Res* 12:339-348.

APPENDIX

CONSENTIMIENTO INFORMADO

PROYECTO DE CONTRIBUCIONES ETNICAS POR VIA PATENAL AL ACERVO GENETICO DE PUERTO RICO

1. Introducción y Objetivo

   Hola. Yo soy Katherine Martínez-Vargas, estudiante del Recinto Universitario de Mayagüez, y estoy llevando a cabo este proyecto cuyo objetivo es identificar la contribución paternal de los distintos grupos étnicos a la genética de la población humana en varias regiones de Puerto Rico. El conocimiento adquirido con esta investigación ayudara a explicar el desarrollo histórico, social y cultural de los puertorriqueños.

2. Beneficios

   Este proyecto ayudará a expandir nuestros conocimientos sobre la historia de Puerto Rico incluyendo su desarrollo genético, social y cultural en diversas regiones geográficas. El proyecto no me rendirá beneficios económicos a mí ni a mi institución. Tampoco rendirá beneficios económicos a ningún participante.

3. Riegos para el voluntario

   El voluntario se enjuagara la boca vigorosamente por 45 segundos con una solución comercial para enjuague bucal y descargara la solución resultante en un envase limpio. Los riesgos son los mismos que cuando se hace un enjuague bucal; el peor de ellos seria un ahogamiento temporal. Además, la información que se obtendrá de las células bucales es de naturaleza genética, personal y familiar, que pertenece única y exclusivamente al voluntario y que, en caso de que fuera obtenida por personal no autorizado, podría ser utilizada ilegalmente para discriminar en contra del voluntario o su familia.

4. Procedimiento

   El voluntario se enjuagara la boca vigorosamente por 45 segundos con una solución comercial de enjuague bucal y vertirá la solución resultante en un envase limpio. La solución será transferida aun envase con tapa y llevada al laboratorio para análisis. Dependiendo de los resultados de la muestra y de la cantidad de material genético que se pueda extraer de la misma, existe la posibilidad de que haya que regresar al voluntario para obtener mas células bucales.

5. Entrevista

   Como parte del procedimiento, se llena una entrevista. Antes de la misma se le informa al voluntario su derecho a negarse a contestar cualquiera de las preguntas. En la entrevista se le pide al voluntario su nombre, dirección, y número de teléfono. Esta información puede ser necesitada si hubiera que retornar al voluntario para obtener mas células escuamosas o para hacerle llegar información relacionada a los resultados de su muestra, del voluntario para obtener mas células escuamosas o para hacerle llegar información relacionada a los resultados de su muestra, del voluntario así solicitarlo.

   La entrevista también pide información sobre la ascendencia por la línea paterna del voluntario. Debido a que el cromosoma Y se hereda del padre a sus hijos varones, esta información le proveerá al proyecto una mejor idea de los orígenes geográficos de los cromosomas Y. Además, la entrevista solicita información en cuanto al nivel de instrucción e ingresos del voluntario. El propósito es comparar la incidencia de grupos étnicos particulares en la ascendencia paterna con el nivel socioeconómico de los voluntarios. Esto permitirá identificar relaciones existenciales entre la etnicidad de la ascendencia paternal y el nivel socioeconómico del puertorriqueño.

Finalmente, la entrevista le asigna un número a la muestra como parte del procedimiento necesario para garantizar la confidencialidad de la información obtenida. Manteniendo la entrevista en un sitio seguro, la información necesaria para relacionar cada persona a cada muestra se mantendrá accesible únicamente al Dr. Juan C. Martínez-Cruzado, quien es el director del proyecto.

6. Acuerdo de Confidencialidad

Para proteger al voluntario de cualquier discriminación que pudiera ocurrir por la divulgación indebida de su información genética, el director del proyecto y su personal se comprometen mediante este documento a mantener dicha información en estricta confidencialidad de acuerdo a las estipulaciones de la reglamentación vigente. Por su parte, el voluntario autoriza al personal del proyecto a analizar única y exclusivamente su DNA extraído de las células escuamosas.

Finalmente, en el caso improbable de que el voluntario sufriese alguna lesión en el proceso de proveer la muestra, el Recinto Universitario de Mayagüez se compromete a conseguirle tratamiento libre de costo.

7. Derechos del voluntario

El voluntario tiene derecho a que se le aclare cualquier duda y a pedir la información derivada de su muestra. Como voluntario, su DNA puede ser estudiado únicamente bajo su continua autorización y consentimiento informado. Si luego de tomada la muestra, el voluntario necesitase comunicarse con el director del proyecto para aclarar alguna duda, solicitar que se le informe sobre los resultados de su muestra, o retirarse del proyecto, puede comunicarse por los siguientes medios:

Dirección: Juan C. Martínez Cruzado, Ph.D.
Departamento de Biología
Recinto Universitario de Mayagüez
Universidad de Puerto Rico
P. O. Box 9012
Mayagüez, PR 00681-9012

Teléfono: (787) 265-3837

Correo electrónico: jmartinez@stahl.uprm.edu
ju_martinez@rumac.uprm.edu

Esta hoja será explicada y entregada antes de que el voluntario acceda a participar en el proyecto.

Dada en ____________, Puerto Rico a los ____ días del mes de _____________de _______.

Firma: ____________________________________ Fecha: _____________________
Voluntario

Firma: ____________________________________ Fecha: _____________________
Investigador

Firma: ____________________________________ Fecha: _____________________
Testigo

Al seleccionado:
Gracias por aceptar colaborar con nuestro proyecto accediendo a proveernos una muestra bucal. Ahora me gustaría hacerle unas pocas preguntas sobre ustedes y su familia. Estas preguntas nos ayudaran a entender mejor la situación de vida de los diferentes grupos de personas que componen la población puertorriqueña y como esta se ha desarrollado a través del tiempo.

Nota para el entrevistador: **Las preguntas se refieren al seleccionado. No necesariamente al que proveyó la muestra.**

1. ¿Cual es su edad? _____

2. ¿Donde vivía su familia cunado usted nació? __________________________

   Bo. ______________________

3. ¿Cual fue el último grado de escuela o año de universidad que completó? ______

4. ¿Cual es el diploma, certificado o titulo educativo o profesional más alto que obtuvo?

   ____Diploma de escuela elemental
   ____Diploma de escuela intermedia
   ____Diploma de escuela superior
   ____Grado Asociado
   ____Bachillerato
   ____Maestría
   ____Doctorado

5. ¿Cual es su ocupación? ________________________

6. ¿Cuanto es su ingreso?_________________________

7. ¿Es usted el principal proveedor del hogar?_________

Ahora le haré algunas preguntas sobre su familia.

8. ¿Cual es o era la ocupación de su padre? ________________________________

9. ¿Cual es o era la ocupación de su madre? ________________________________

Ahora le haré algunas preguntas sobre su padre.

10. ¿Cual es la fecha de nacimiento de su padre? ___________________________

11. ¿En donde vivía la familia de su padre cuando el nació?

______________________________ Bo. ____________________________

Ahora hablemos sobre su abuelo paterno.

12. ¿Cual es su fecha de nacimiento?___________________________________

13. ¿En donde vivía la familia de su padre cuando el nació? _________________

Bo. ______________________________

14. ¿Cual era su ocupación? ____

Y sobre su bisabuelo paterno.

15. ¿Cual es su fecha de nacimiento?___________________________

16. ¿En donde vivía la familia de su padre cuando el nació? _________________

Bo. ____________________

17. ¿Cual era su ocupación? ____

18. Diga el municipio y el barrio donde vivía su familia en el año:

2000 _____________________________ Bo. __________________________

1975 _____________________________ Bo. __________________________

1950 _____________________________ Bo. __________________________

1925 _____________________________ Bo. __________________________

1900 _____________________________ Bo. _________________________

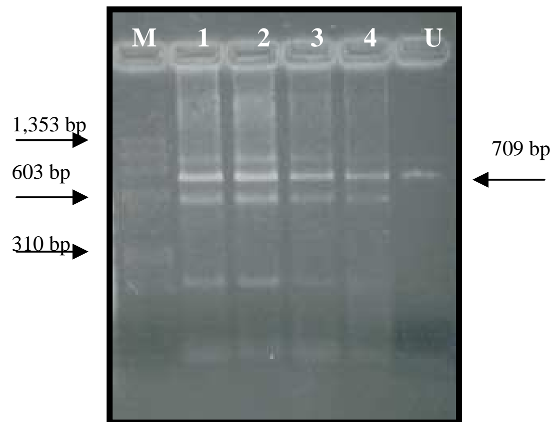
19. ¿Desea conocer los resultados de su muestra? Si____ No____

Num. de muestra __________

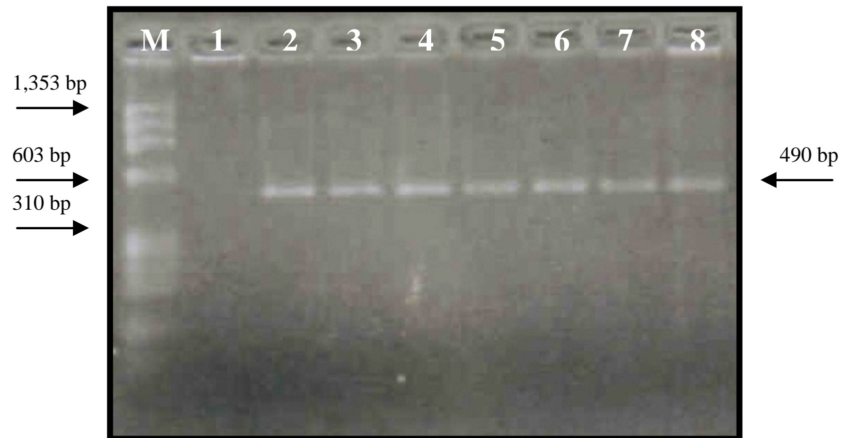
¿Alguna información adicional?

PICTURES

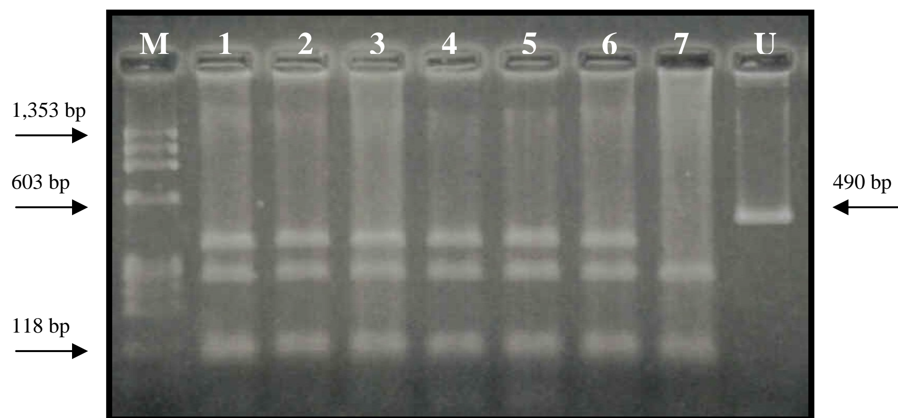
Picture 1. **ΦX174 DNA-*Hae*III Digest Marker and 92R7 PCR product of 709pb.**



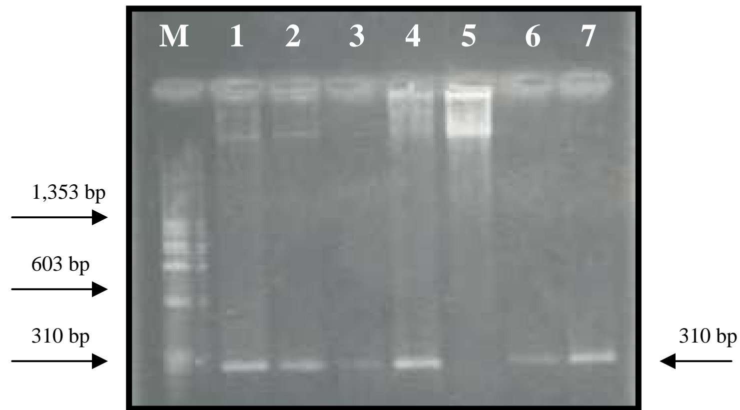
Picture 2. **92R7 DNA-*Hind*III Digest.** ΦX174 DNA-*Hae*III Digest Marker; 92R7 DNA-*Hind*III Digest (1-4); and 92R7 DNA uncut (5). In this case all samples (1-4) are ancestral for 92R7.



Picture 3. **ΦX174 DNA-*Hae*III Digest Marker and P25 PCR product of 490pb.**

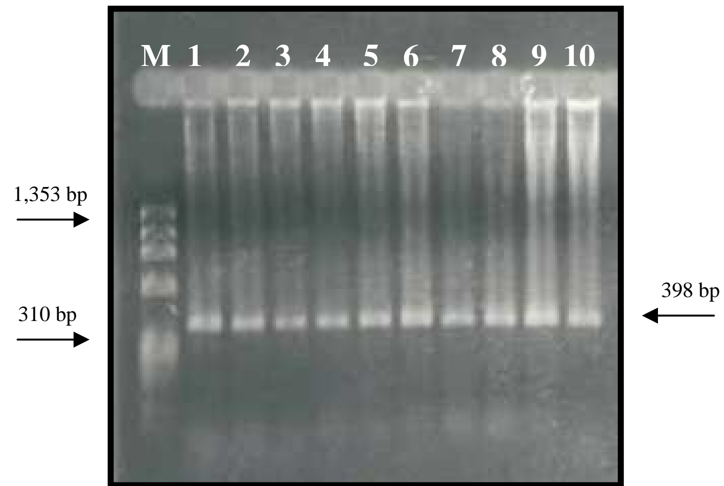


Picture 4. **P25 DNA-HpyCH4V.** ΦX174 DNA-*Hae*III Digest Marker; P25 DNA-*HpyCH4*V Digest (1-7); and P25 DNA uncut. Samples 1-6 are derived for P25 therefore belong to haplogroup R1b1. However, sample 7is ancestral for P25.



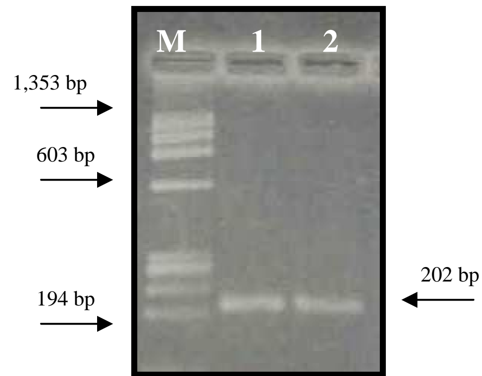
Picture 5. **ΦX174 DNA-*Hae*III Digest Marker and SRY10831b PCR product (only 310bp fragment visible).**

Picture 6. **SRY10831b DNA-*Dra*III.** ΦX174 DNA-*Hae*III Digest Marker; SRY10831b DNA uncut; and SRY10831b DNA-*DRA*III Digest (1-7). In this case all samples (with the exception of 5) all ancestral for SRY10831b. Sample 5 could not be determined.

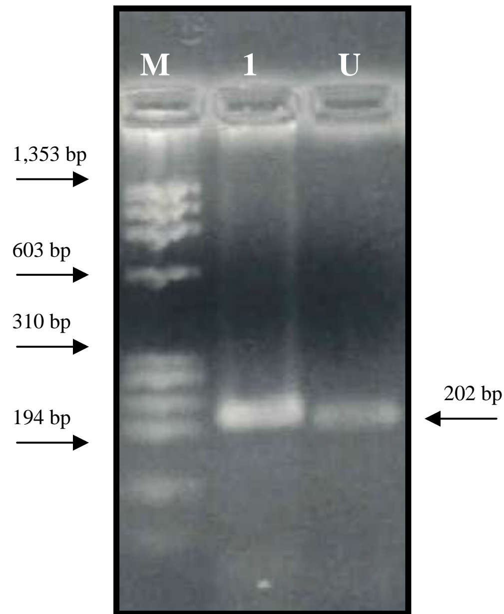


Picture 7. **ΦX174 DNA-*Hae*III Digest Marker and M242 PCR product of 398bp.**

Picture 8. **ΦX174 DNA-*Hae*III Digest Marker and M207 PCR product of 422bp**.



Picture 9. **ΦX174 DNA-*Hae*III Digest Marker and M3 PCR product of 202bp.**



Picture 10. **M3 DNA-*Mfe*I Digest**. ΦX174 DNA-HaeIII Digest Marker, M3 DNA-MfeI Digest and M3 DNA uncut. Sample does not have restriction site for *Mfe*I therefore is derived for M3 and thus belong to haplogroup Q3.