





3 1293 01020 8324

THERM

This is to certify that the

thesis entitled

POLYMORPHISM OF THE MITOCHONDRIAL DNA CONTROL  
REGION IN THE PUERTO RICAN POPULATION

presented by

Amin Abdel-Rahman Abujoub

has been accepted towards fulfillment  
of the requirements for

Master of Science degree in Clinical Laboratory  
Science

Robert W. Bull

Major professor  
Robert W. Bull, D.V.M.

Date August 22, 1994

**LIBRARY  
Michigan State  
University**

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.

| DATE DUE | DATE DUE | DATE DUE |
|----------|----------|----------|
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |

**MSU Is An Affirmative Action/Equal Opportunity Institution**

c:\circ\dstdue.pm3-p.1

**POLYMORPHISM OF THE MITOCHONDRIAL DNA CONTROL REGION IN THE  
PUERTO RICAN POPULATION**

**By**

**Amin Abdel-Rahman Abujoub**

**A THESIS**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**MASTER OF SCIENCE**

**Medical Technology Program**

**1994**

## **ABSTRACT**

### **POLYMORPHISM OF THE MITOCHONDRIAL DNA CONTROL REGION IN THE PUERTO RICAN POPULATION**

**BY**

**Amin Abdel-Rahman Abujoub**

Polymerase chain reaction (PCR) was used to amplify and sequence the hypervariable segment 1 of the mitochondrial DNA (mtDNA) control region of 50 Puerto Ricans. Comparison of these sequences with the human reference sequence (Anderson et al., 1981) revealed an eight fold excess of substitutions to length mutation events. The substitution observed obeyed the expected bias toward transition rather than transversion type events. Sequence analysis revealed the existence of 33 mitochondrial lineages (mt-lineages) defined by 20 variable positions. These 33 mt-lineages were found to be clustered in 4 main groups, which defined the ethnic origin of Puerto Ricans. Sixty eight percent of the Puerto Ricans mt-lineages were found to be similar to Amerindian mt-lineages, and 26% of the mt-lineages were found to be similar to Southern African mt-lineages.

To my father and mother  
and my lovely wife Aida  
for their love, support, and patience.

and

To my lovely daughter and son  
Rawan and Shadi

## **ACKNOWLEDGMENT**

I would like to thank my advisors Dr. R. W. Bull, and Dr. J. A. Gerlach for their encouragement and understanding.

Special thanks to my committee members Dr. D. Estry, and Dr. J. M. Kaguni for their valuable time and suggestions. I'm also thankful to Mr. R. A. Southwick for helping me in drawing some of my figures.

Finally, my deepest appreciation goes to my wife Aida for her love and continuous encouragement through all these years.

## TABLE OF CONTENTS

|                                                                                 |      |
|---------------------------------------------------------------------------------|------|
| List of Tables.....                                                             | vii  |
| List of Figures.....                                                            | viii |
| Introduction.....                                                               | 1    |
| The mitochondrion.....                                                          | 2    |
| Organization of human mtDNA and the control region.....                         | 3    |
| The Puerto Ricans and type i (insuline dependent) diabetes mellitus (IDDM)..... | 14   |
| The Puerto Ricans.....                                                          | 16   |
| Objectives.....                                                                 | 19   |
| Materials and methods.....                                                      | 20   |
| DNA isolation.....                                                              | 20   |
| Proteinase digestion.....                                                       | 20   |
| Mitochondrial DNA amplification.....                                            | 21   |
| Fidelity of Amplification.....                                                  | 22   |
| DNA electroelution.....                                                         | 23   |
| Cycling sequencing.....                                                         | 24   |
| Sequencing gel electrophoresis.....                                             | 25   |
| Autoradiography.....                                                            | 26   |
| Sequence analysis.....                                                          | 26   |
| Results.....                                                                    | 29   |
| Distribution of mutated sites.....                                              | 32   |

|                                                   |    |
|---------------------------------------------------|----|
| Substitutions versus length mutations.....        | 52 |
| Transition versus transversions.....              | 52 |
| Sequence diversity.....                           | 54 |
| Phylogenetic analysis.....                        | 54 |
| Discussion.....                                   | 58 |
| Comparison of all sequences.....                  | 58 |
| Sequence diversity and phylogenetic analysis..... | 60 |
| Future prospective.....                           | 63 |
| List of references.....                           | 64 |

## **List of Tables**

|                                                                                      |           |
|--------------------------------------------------------------------------------------|-----------|
| <b>Table 1. Sequence of primers used for amplification and sequencing.....</b>       | <b>23</b> |
| <b>Table 2. Puerto Ricans mitochondrial DNA sequences.....</b>                       | <b>34</b> |
| <b>Table 3. Base substitution and length mutations.....</b>                          | <b>45</b> |
| <b>Table 4. Nucleotide positions in the reference sequence</b>                       | <b>49</b> |
| <b>Table 5. Analysis of nucleotide substitutions, deletions, and insertions.....</b> | <b>53</b> |
| <b>Table 6. Mitochondrial lineages.....</b>                                          | <b>55</b> |

## **List of Figures**

|                                                                            |           |
|----------------------------------------------------------------------------|-----------|
| <b>Figure 1. The organization of the human mtDNA genome....</b>            | <b>5</b>  |
| <b>Figure 2. The control region and its functional elements</b>            | <b>11</b> |
| <b>Figure 3. The hypervariable segments of the control<br/>region.....</b> | <b>13</b> |
| <b>Figure 4. Sequencing gel autoradiograph.....</b>                        | <b>27</b> |
| <b>Figure 5. PCR amplification of mitochondrial DNA.....</b>               | <b>30</b> |
| <b>Figure 6. Purified PCR product.....</b>                                 | <b>31</b> |
| <b>Figure 7. Distribution of mutations.....</b>                            | <b>51</b> |
| <b>Figure 8. Phylogenetic tree.....</b>                                    | <b>57</b> |

## Introduction

The mitochondrial genome of animals has captured the interest of biologist in a number of disciplines during the past decade. The simple genetic organization provides an ideal system for the study of gene expression and the coordinate regulation of the organelle and nuclear genome. The genome also provides a record of molecular evolution that is generally believed to be free of the effects of biparental inheritance and recombination. In recent years, analysis of mitochondrial deoxyribonucleic acid (DNA) (mtDNA) has proved to be a powerful tool in evolutionary population genetic structure studies. Restriction endonucleases together with agarose gel electrophoresis, and DNA sequencing has revealed extensive nucleotide sequence diversity within and between nonspecific populations (Avice et al., 1988: 1989). The utility of the genome for studies of human evolution has been well demonstrated by many laboratories (Brown et al., 1980; Johnson et al., 1983; Horai and Matsunga, 1986; Cann et al., 1987; Wilson et al., 1987; Stoneking et al., 1990). A dramatic and controversial outcome of these studies was the proposal that the single common ancestor of all human mtDNA lived in Africa (Cann et al., 1987).

### **The mitochondrion**

It is believed that mitochondria, evolved from procaryotes that were engulfed by primitive eukaryotic cells and developed a symbiotic relationship with them about 1.5 billion years ago (Ernster et al., 1981; Clark, 1990). This would explain why the mitochondria contains it's own genome, which codes for some of their proteins. Since then, however this organelle has lost much of it's genome and has become heavily dependent on proteins that are encoded by genes in the nucleus (Ernster et al., 1981; Greenberg et al., 1983; Cote et al., 1990). Conversely the host cells have become dependent on their mitochondria for the generation of most of the adenosine triphosphate (ATP) they need to carry out biosynthesis.

Mitochondria are small (0.5-1.0 by 5-10 micrometer (um)) oval cytoplasmic organelles found only in eukaryotes (Spuhler, 1988). Hundreds of these self replicating organelles may be found in a single mammalian cell (Bogenhagen and Clayton, 1974). MtDNA is a covalently closed circular double-stranded DNA (dsDNA) molecule (Brown et al., 1978; Aquadro and Greenberg, 1983; Clark, 1990). Each mitochondrion contains several mtDNA molecules, and hence the presence in each mammalian cell of hundreds to thousands (Giles et al., 1980; Palca, 1990) of mtDNA increases the ease of isolation and examination of these molecules.

### Organization of human mtDNA and the control region

The circular human mtDNA genome has been extensively characterized. The complete nucleotide sequence of it's 16,596 base pairs (bp) has been determined for a single human (Anderson et al., 1981, 1982). The molecular biology of mitochondrial functions has been comprehensively reviewed by Anderson et al., (1981), Clayton (1982, 1984) and Attardi (1985). This genome was shown to contain the genes for two RNA homologous to the 16S and 23S ribosomal ribonucleic acid (rRNA) of *Escherichia Coli*, and for 22 transfer RNA (tRNA), and 13 open reading frames (ORF). Six of these ORF were identified as the genes for enzymes or components of enzymes involved in oxidative-phosphorylation: Cytochrome b (Cyto b), subunits I-III of Cytochrome C (COI-III), and subunits 6 and 8 of the F<sub>0</sub> ATPase complex (ATPase 6 and ATPase 8). The remaining 7 ORF, designated URF 1-6 and 4L, were later shown to encode subunits of the respiratory chain nicotinamide dinucleotide dehydrogenase (NADH) complex and have since been referred to as N 1-6 and N4L (Chomyn et al., 1985, 1986). But as mentioned earlier most of the proteins required for mitochondrial function are nuclear encoded and imported from the cytoplasm. Figure 1 is a schematic diagram of the mt-genome, where all the genes, and the noncoding DNA are shown. This figure was adapted from Mitochondrial Genomes (Wolstenholme and Jeon, 1992).

The gross genetic arrangement of the genome is

**Figure 1. The organization of the human mtDNA genome.** Abbreviations for the genes are as follows: two ribosomal RNA genes (12S and 16S); seven genes for NADH dehydrogenase subunits (N1-N6, N4L); three genes for cytochrome oxidase subunits (COI-COIII); two genes for ATPase subunits (6 and 8); the gene for cytochrome b (Cyt b); and a single letter codes for the 22 transfer RNA genes. The origins of replication of the heavy ( $O_H$ ) and light ( $O_L$ ) strands are designated within the solid bars representing noncoding regions. Right (R) and left (L) represent the direction of replication of the H and L-strands respectively. The inside arrow ( $\rightarrow$ ) show the direction of transcription of the mt-genome. The large noncoding region known as the control region.

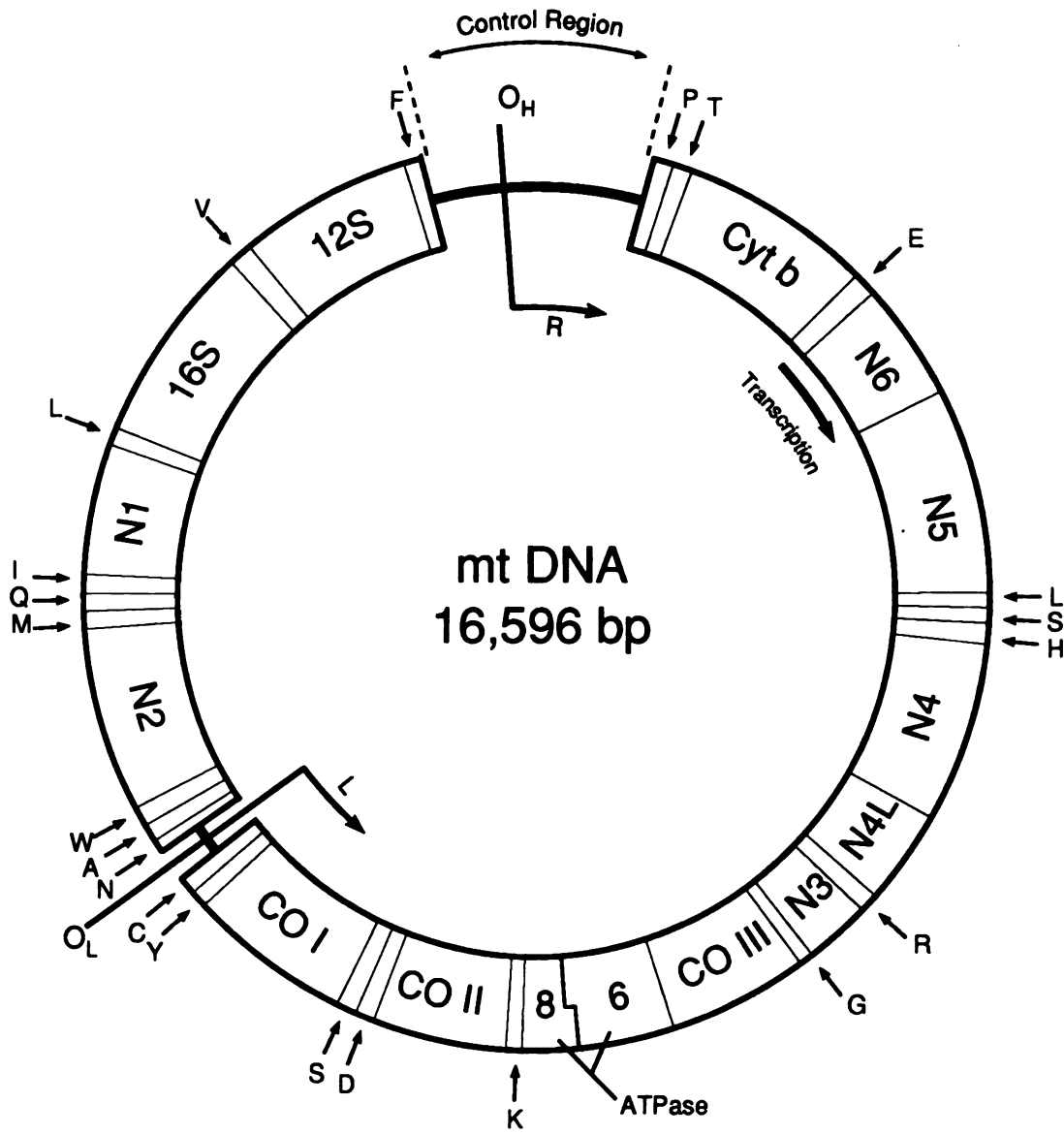


Figure 1

remarkably conserved, it can be divided into two general domains. A coding region which has no introns and usually has just a few base pairs of non-coding DNA. The close packing of genes observed is likely related to the mechanism by which most mammalian mtDNA appear to be transcribed: that is the generation of primary transcripts of entire strands are produced by precise cleavage, possibly, in some cases, as a function of tRNA secondary structure. The other domain is the non-coding region (control region) which lies between tRNA proline (tRNA<sup>pro</sup>) and tRNA phenylalanine (tRNA<sup>phe</sup>) and contains almost all the non-coding DNA of the entire molecule. This region is 1122 bp (Anderson et al., 1981). The middle segment of the control region is known as the displacement (D)-loop. The D-loop is a triple stranded region generated by the synthesis of a short piece of heavy-strand DNA, the 7SDNA (Clayton, 1982).

The complementary strands of mammalian mtDNA molecules differ sufficiently in Guanine (G) and thymine (T) content that they can be separated in alkaline cesium chloride gradients. The complementary strands of these mtDNA molecules thus acquired the designation heavy (H) and light (L) strands, that have been used as strand definitions in replication and transcription studies of mammalian mt-genome (Clayton, 1982, 1984, 1991).

Although the control region contains no structural genes, it is not without functional significance. It has the origin of H-strand replication (Montoy et al., 1982,

1983; Clayton, 1984), where as the origin of the L-strand replication is located in another noncoding region, between the cystine (C), and the asparagine (N) tRNA genes. The control region also has the transcription starting sites for both H- and L-strands, and promoters for both H- and L-strands transcription, light strand promoter (LSP), and heavy strand promoter (HSP), respectively (Chang and Clayton, 1984; Hixon and Clayton, 1985; Cann et al., 1987; Horai and Hayasaka, 1990; Kocher and Wilson, 1991; Stoneking et al., 1991; Vigilant et al., 1991). Both LSP and HSP are bidirectional (Change et al., 1986) and associated with two upstream transcription factors (TF) binding sites designated mtTF-L and mtTF-H respectively (Fisher et al., 1987). Three conserved sequence blocks (CSB) 1,2, and 3 are associated with the origin of replication (Walberg and Clayton, 1981; Chang and Clayton, 1987a and b), and these sequences are similar between species, and conserved during evolution (Low et al., 1988). Another 5 CSB (B-F) in the control region was found to be conserved in most mammals (Southern et al., 1988; Kocher and Wilson, 1991). Another conserved structure in the control region is the D-loop termination associated sequence (TAS) (Walberg and Clayton, 1981; Foran et. al, 1988). These functional elements of the control region have been found to exhibit some sequence conservation in inter-species comparison (Walberg and Clayton, 1981; Hixon and Clayton, 1985; Brown et al., 1986; King and Low, 1987). The control region and its functional elements are schematically

illustrated in Figure 2.

Several features of mtDNA have made it a popular molecule for evolutionary studies of human populations. These characteristics include its high copy number, maternal inheritance with no recombination (Giles et al., 1980; Case and Wallace, 1981), and rapid evolution (Wilson et al., 1985; Stoneking and Cann, 1989). In 1979 Brown et al. published a paper documenting that the rate of mtDNA evolution was 5 to 10 times higher than that of the single copy nuclear DNA (scnDNA). Further more the amount of sequence divergence in the control region exceeds that in the sequence coding for proteins, tRNAs or rRNAs (Aquadro and Greenberg, 1983; Greenberg et al., 1983; Cann et al., 1984; Vigilant et al., 1988, 1989, 1991; Kocher and Wilson, 1991). A comparison of the control region sequence by Greenberg et al (1983) from a total of seven individuals found that divergence in this region was ten-fold greater than that of the mtDNA molecule as a whole. The control region is highly polymorphic, with most of the variation distributed not at random, but rather concentrated in two hypervariable segments (Kocher and Wilson, 1991; Stoneking et al., 1991; Vigilant et al., 1991). The first hypervariable segment lies between nucleotide 1 and 400 of the control region (positions 16024 to 16423 in the reference sequence (Anderson et al., 1981)). The second hypervariable segment lies between nucleotide 600 to 900 of the control region (positions 28 to 328 in the reference

sequence (Anderson et al., 1981)) (Greenberg et al., 1983; Kocher and Wilson, 1991). The central portion of the control region between the two hypervariable segment was found to be conserved, and the reason for this conservation remains obscure (Walberg and Clayton, 1981; Greenberg et al., 1983; Brown et al., 1986). This low level of variation is coincident with the five CSBs (B-F) mentioned earlier (Kocher and Wilson, 1991). Hypervariable region one is found to be approximately twice as variable as hypervariable region two (Vigilant et al., 1991). This reduction in polymorphism is not random, but it is concurrent with the presence of seven of the eight functional elements in that region. Figure 3 is a schematic diagram of the control region showing the location of the two hypervariable segments, as well as the binding sites for the oligonucleotide primers used in this study.

The majority of the polymorphism observed in the control region consists of a single base substitutions rather than length polymorphism (insertions and deletions) of bases (Greenberg et al., 1983; Stoneking et al., 1986a and b; Vigilant et al., 1989; Wrischnik et al., 1987). Comparison between closely related mtDNA sequences revealed that certain base substitutions occur more often than others. Transitions greatly out number transversions in closely related species. Vigilant et al. (1991), and Aquadro and Greenberg (1983) reported a 30:1, and 24:1 transitions to transversions ratio respectively. This

**Figure 2. The control region and its functional elements.**  
A schematic diagram of the control region which is flanked by proline (pro) and phenylalanine (phe) tRNA genes. All the functional and the conserved elements of the control region are shown. The displacement (D) loop DNA associated with the origin of replication, heavy strand promoter (HSP), light strand promoter (LSP), heavy strand mitochondrial transcription factor (mtTFH), light strand mitochondrial transcription factor (mtTFL), conserved sequence blocks (CSB) 1-3 and B-F, and the D-loop termination associated sequences (TAS). The numbers at the upper strands represent the nucleotide positions in the reference sequence ANDE (Anderson et al., 1981).

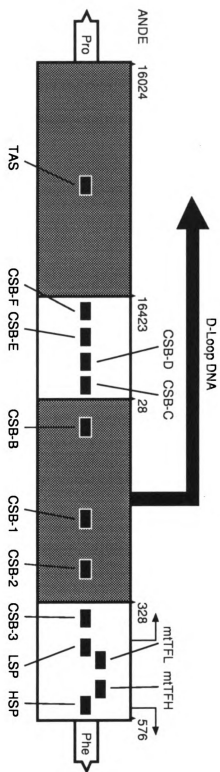


Figure 2

Figure 3. The hypervariable segments of the control region. The two hypervariable segments of the control region are represented by the gray blocks labelled segment 1 and 2. The flanking primers (L15926 and H16498 represented by the two arrows) used in amplification and sequencing are indicated. Proline (pro) and phenylalanine (phe) tRNA genes are noted. The numbers at the upper strand represent the nucleotide positions in the reference sequence ANDE (Anderson et al., 1981).

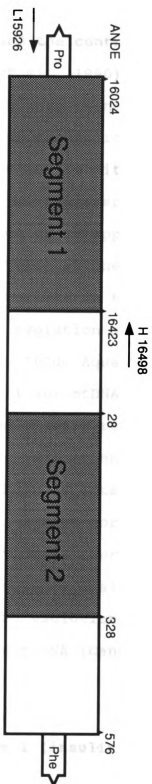


Figure 3

higher frequency of transitions is observed at all codon positions, tRNAs, rRNA, and the control region (Kocher and Wilson, 1991; Stoneking et al., 1990). Such a trend is expected because transversions, even if relatively rare, tend to erase the record of transitions. Two sequential transitions at a given nucleotide site always restores the original base, whereas two transversions at the same site results in either no change or an apparent transition, and a transition plus a transversion at the same site (regardless of the order they occur) results in transversion.

Owing to its easiest evolution and maternal mode of inheritance (Giles et al., 1980; Aquadro and Greenberg, 1983; Horai and Hayasaka, 1990) mtDNA can provide knowledge of genetic relations among closely related individuals (Horai and Hayasaka, 1990; Rienzo and Wilson, 1990; Vigilant et al., 1991). Results obtained from different studies show that there is a high correlation between mtDNA and ethnic origin of individuals (Ferris et al., 1981; Horai and Hayasaka, 1990). So sequence analysis of the control region affords the maximum resolution for distinguishing among very closely related mtDNA (Cann et al., 1987; Vigilant et al., 1991).

### **The Puerto Ricans and type 1 (insulin dependent) diabetes mellitus (IDDM)**

In the Caucasian population class II major histocompatibility complex (MHC II) determinants are known

to be associated with increase susceptibility to IDDM (Lee et al., 1992). The HLA-DQ locus, specifically DQ $\beta$ 57 residue is an important marker in determining susceptibility to IDDM in Caucasians (Rotter et al., 1983; Aparicro et al., 1988; Horn et al., 1988; Sterkers et al., 1988).

The greatest susceptibility to IDDM is presumed to be individuals with both DQ alleles at position 57 encoding non-aspartate amino acids. As high as 96% of the diabetic probands homozygous for a non-aspartate amino acid at DQ $\beta$ 57 have been reported to have IDDM in the Caucasian population (Todd et al., 1987; Morel et al., 1988; Dorman et al., 1991; Penny et al., 1993).

Analysis of the HLA-DQ locus in the Puerto Ricans show that 46.7% of the IDDM patients were homozygous for a non-aspartate amino acid at DQ $\beta$ 57 (Lee et al., 1992), the same study also showed that 13.6% of the non-diabetic Puerto Ricans were also homozygous for non-aspartate amino acid at the same locus.

The difference in the HLA-DQ locus between the Caucasians and the Puerto Ricans could reflect difference in the ethnic origin between the two groups. MtDNA studies have been demonstrated to be a suitable tool for population genetics studies and could be used here to define the ethnic origin of the Puerto Ricans.

## **The Puerto Ricans<sup>1</sup>**

The first human inhabitants of Puerto Rico were believed to have come from North America, probably from Florida. These groups may have arrived any time between 20,000 and 5,000 years ago and they have been called the Archaicos, or Archaics, but their culture was so primitive that no clear-cut signs of it are left. The Archaics were followed by members of the Arwak language family known as the Igneri Indian. The Arawakan people inhabited north South America; the region that extends from coastal Brazil through Venezuela to Colombia, and made their way north to Puerto Rico and the other islands around it. Arawakan are believed to have reached Puerto Rico at about the time of Christ, or perhaps a couple of hundred years earlier. The last major group of indians, the Tainos, are the best known indian group. They lived in Puerto Rico from about A.D. 1000 to the early 1500's when the first Europeans began their settlement on that land.

When Christopher Columbus returned to the West Indies's on his second voyage in 1493 and colonized Puerto Rico, he had seventeen ships that carried a crew and passengers of 1,200 men. There were no women among the would be colonists. By the year 1510, the first smelted gold was being sent to Spain. After that year more Spanish settlers were moving into the new land "Puerto Rico".

---

<sup>1</sup>The Puerto Ricans history was adapted from PUERTO RICO "Island Between Two World" by Lila Perl, 1979.

With the arrival of the first Spanish settlers, radical changes took place in the lives of the Tainos. The Indians were abused under what is called the *repartimiento*, a system of "distribution" under which the Indians were rounded up for the labor bridges and assigned to the building of Spanish forts and residences or put to work in the mines and in the fields. Soon after that the Indians started a series of attacks on the colonists, which were met with harsh retaliation. The Indians realized that despite their superior numbers they were doomed by the weaponry of the Europeans, and began to flee into the mountains of the interior or to set out in small boats for neighboring islands.

The killing and the migration of the Indians led to the rapid decline of the Taino population. Because of the growing demand for labor, the first African slaves began to be brought to the island. In fact, blacks first arrived in Puerto Rico as early as 1509; they arrived as freemen as well as servants of African origin who accompanied the Spaniards on their various expeditions to the West Indies.

With the establishment of the sugar plantations in Puerto Rico around 1520, the need for field labor increased. Black slavery was broadly accepted in the world at the time, so more Africans arrived.

In the year 1897; the Spanish-American War broke out, and Puerto Rico was handed over by Spain to the victorious United States by the year 1898. From that time till now

Puerto Rico is a territory of the United State of America. Because of all the former mentioned groups that have lived in Puerto Rico, it is considered a land of diversity and this makes the Puerto Ricans good candidates for population genetics studies.

**Objectives**

The major objectives of this research are to examine the sequence heterogeneity within the Puerto Rican mt-lineages and to determine the extent of polymorphism in the mitochondrial control region of the Puerto Ricans. By studying the mt-lineages in the Puerto Ricans their predominant ethnic origin can be determined. With this the differences between Caucasian and the Puerto Rican genetic susceptibility to IDDM may be proved.

## **Materials and Methods**

Fifty blood samples were collected randomly from the Puerto Rican population. Ten milliliters (ml) peripheral blood was collected in sodium heparin tubes and samples were centrifuged at 500 x gravity (g) for 20 minutes (min). The buffy coat was aspirated, and the white blood cells (WBCs) were washed once with phosphate buffered saline (PBS) (137 millimolar (mM) NaCl, 2.7 mM KCl, 9.6 mM NaH<sub>2</sub>PO<sub>4</sub>, and 1.5 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4). The pellet was stored with 1 ml PBS at - 70 degrees celsius (°C) for four years.

### **DNA Isolation**

The WBC suspension was thawed at 37°C for 15 min and centrifuged in a fixed angle microcentrifuge for 10 min at 10,000 x g. The pellet was resuspended in 0.5 ml TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0) buffer.

### **Proteinase digestion**

Thirty microgram (ug) proteinase K was added to each sample and incubated at 56°C for 12 - 18 hours. An equal volume of phenol/chloroform:isoamyl alcohol (25:24:1, weight (wt)/ volume (vl)) was added to each tube, followed by vortexing for 1 min, and spinning in a microcentrifuge at 10,000 x g for 4 min. The aqueous phase was transferred to a new tube. A second phenol/chloroform:isoamyl alcohol (25:24:1 wt/vl) extraction was performed followed by

extraction with an equal volume water saturated n-butanol. The aqueous phase was concentrated by centrifugal dialysis using a centricon-100 (Amicon, Beverly, Massachusetts) at 1000 x g for 20-30 minutes in a fixed angle centrifuge. The centrifugal dialysis step was repeated three times using double distilled water to remove salts which may interfere with subsequent steps. The amount and purity of DNA was determined spectrophotometrically (Sambrook et al., 1989).

### **Mitochondrial DNA Amplification**

One hundred nanogram (ng) of total DNA was subjected to 40 cycles of amplification in a 100 microliter (ul) reaction volume. The procedure for setting up a polymerase chain reaction (PCR) was as follows: 1) Addition of 10 ul (100 ng) of sample (or sterile double distilled water for the negative control) to 0.5 ml microcentrifuge tube. 2) Addition of 89.5 ul of reaction cocktail consisting of : i. 10 ul 10X reaction buffer (100 mM Tris-HCl (pH 8.3), 500 mM KCl, 15 mM MgCl<sub>2</sub>, and 0.01% (wt/vl) gelatin). ii. 16 ul of deoxynucleotide 5'-triphosphate (dNTP) mix, which consists of 200 uM of each deoxynucleotide 5'-triphosphate (dNTP), (deoxyadenosine 5'-triphosphate (dATP), deoxycytosine 5'-triphosphate (dCTP), deoxyguanosine 5'-triphosphate (dGTP), deoxythymidine 5'-triphosphate (dTTP)). iii. 10.0 micromolar (uM) of each oligonucleotide primer. 3) Addition of 0.5 ul (2.5 units) *thermus aquaticus* (Taq) DNA polymerase (Perkin Elmer Cetus, Norwalk, Connecticut) to

each tube. Each amplification cycle consisted of denaturation at 94°C for 1 minute, annealing at 56°C for 1 minute, and extension at 72°C for 1 minute.

The two oligonucleotide primers used in the amplification reaction were H16498 (Ward et al., 1991), and L15926 (Kocher et al., 1989; Kocher and Wilson, 1991; Stoneking et al., 1991; Rienzo and Wilson 1990; Vigilant et al., 1991), where the letter indicates the mitochondrial strand, and the numbers identify the base corresponding to the 3' end of the primer. The two oligonucleotide primers were synthesized at the Macromolecular Structure Facility at Michigan State University using the 394 DNA synthesizer (Applied Bioscience, Foster City, California). The two primers were purified using the C<sub>18</sub> sep-pack (Waters Corporation Division, Millipore Corporation, Bedford, Massachusetts) purification procedure (Atkinson and Zoller, 1984). The primer sequences are given in Table 1. Primers A and B defined a 520 base pair (bp) segment of the control (non-coding) region of human mtDNA corresponding to the hypervariable segment 1.

### **Fidelity of Amplification**

The fidelity of PCR was determined by mixing 10 ul of PCR product with 5 ul DNA 5X loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, and 40% (wt/vl) sucrose in water). The mixture was loaded in 8% polyacrylamide gels (16.4 mM acrylamide, 0.2 mM bis

**Table1. Sequences of primers used for amplification and sequencing.**

-----  
Primer A (H16498): 5' CCTGAAGTAGGAACCGAT 3'  
Primer B (L15926): 5' TCAAAGCTTACACCAGTCTTGTA 3'  
-----

(N,N'-methylene-bis-acrylamide)), along with the negative control, as well as a DNA molecular weight marker VIII (Boehringer Mannheim, Indianapolis, Indiana) were loaded in the same gel to enable size estimation. The gel was electrophoresed in 1X TAE buffer (40 mM Tris-Acetate, 1 mM EDTA (pH 8.0)), at 200 volts for 30 minutes. The gel was stained with ethidium bromide (0.5 ug/ml) for 20 min, and destained for another 20 min in distilled water. DNA bands were visualized using trans-illumination with an ultraviolet (UV) light source. The gel was photographed using type 667 film (Polaroid, Cambridge, Massachusetts).

#### **DNA Electroelution**

The remaining amplified product was further purified by electrophoresis in an 8% polyacrylamide gel. The band of interest was cut from the gel and the DNA electroeluted at 200 volts for 4 - 6 hours using a micro-Centrilutor system (Amicon, Beverly, Massachusetts). The electroeluted product was concentrated and washed once with double

distilled water using centricon-100 microconcentrator (Amicon, Beverly, Massachusetts), the microconcentrator was centrifuged in a Sorvall SS34 rotor (Sorvall Instruments; Dupont, Hoffman Estates, Illinois) at 3,000 x g for 30 min at 15-20 °C. DNA was stored at - 20°C to be used for sequencing.

### **Cycling Sequencing**

Three ul of the eluted template, double stranded DNA (dsDNA), was used for sequencing in a linear polymerase chain reaction (LPCR) (Innis et al., 1988; Smith et al., 1990), where 12.5 ul 4X sequencing buffer (40 mM Tris-HCl (pH 8.8), 200 mM KCl, 0.004% (w/v) gelatin, 16 mM MgCl<sub>2</sub>, 8 uM dATP, 20 uM dCTP, 20 uM dGTP, and 20 uM dTTP) (Stratagene, La Jolla, California) was added to the template, with 10 microcurie (uCi) of alpha-<sup>33</sup>P (α-<sup>33</sup>P) dATP (Dupont, Boston, Massachusetts), 147 nM of primer A, or 221 nM of primer B was added to the mix, 2.5 units Taq DNA polymerase (Perkin Elmer Cetus, Norwalk, Connecticut) was added, the total reaction volume was adjusted to 34 ul with sterile double distilled water.

Eight ul of the above mixture was added to four termination tubes that contained 2 ul each of dideoxy Nucleotide 5'-triphosphate(ddNTP); (dideoxy Adenosine 5'-triphosphate (ddATP) at 240 uM, dideoxy Cytosine 5'-triphosphate (ddCTP) at 120 uM, dideoxy Guanine 5'-triphosphate (ddGTP) at 20 uM, and dideoxy Thymidine 5'-

triphosphate (ddTTP) at 20  $\mu$ M. The four termination tubes were initially incubated at 95°C for 5 minutes to denature the dsDNA template. After the initial denaturation the termination tubes were subjected to 30 cycles of linear polymerase reaction. Each cycle consisted of denaturation at 95°C for 30 seconds, annealing at 56°C for 30 seconds, and extension at 72°C for 60 seconds. The reaction was stopped by adding 5  $\mu$ l stop dye (95% formamide, 20 mM EDTA (pH 8.0), 0.05% bromophenol blue, and 0.05% xylene cyanol).

### **Sequencing Gel Electrophoresis**

Electrophoresis in 1X TBE (89 mM Tris-borate, 89 mM boric acid, 2 mM EDTA (pH 8.0) ) buffer was through vertical 7.2% polyacrylamide gels, 7 molar (M) urea (800 mM acrylamide, 20 mM bis, 6.8 M urea, 40 mM Tris-borate, 40 mM boric acid, 1 mM EDTA, pH 8.0 ) with dimensions of 40 cm long by 20 cm wide. Gels were prepared by putting together two glass plates with a 0.4 millimeter (mm) thick spacer. A solution of 7.2% acrylamide/7M urea was stored at 4 °C and an aliquot of 90 ml used per gel were warmed to room temperature. A 10% solution of ammonium persulphate (APS) (100  $\mu$ l/90 ml) and a small amount of N,N,N',N'-tetraMethylEthyleneDiamine (TEMED) (Sigma) (50  $\mu$ l/90 ml) were added immediately before pouring. Gels were allowed to polymerize from 30 min to 2 hr. Wells were formed using a double fine comb (Biorad, Hercules, California). The gel was prerun in 1X TBE at 1500 volts for 30 minutes. The four

termination tubes from the previous step were heat-denatured at 70-90°C for 2-5 minutes just before loading the gel: 2.5 ul from each tube was loaded and electrophoresed at 1,400 - 1,600 volts until the 1st dye (xylene Cyanol) ran off the gel. A second set from each sample was then loaded and electrophoresed until the 1st dye of the second load ran off the gel. The gel was transferred to blotting paper, and covered by plastic wrap. The gel was baked at 80°C under vacuum for 40-60 minutes.

### **Autoradiography**

The gel was autoradiographed using X-ray film (X-Omat, Kodak, Rochester, New York) for 12-24 hours. Films were developed using GBX developer and replenisher (Kodak, Rochester, New York), fixed with GBX fixer and replenisher (Kodak, Rochester, New York), washed, and drained at room temperature for 1 hr. A photograph of a sequencing gel autoradiograph can be seen in Figure 4.

### **Sequence Analysis**

A typical gel could provide approximately 250 bp of sequence information. The sequences were read using a gel reader digitizer model GP-7 (International Biotechnologies Incorporation, New Haven, Connecticut). The data generated were analyzed using Genetic Computer Group (GCG) software, Wisconsin. The 500 bp of interest were assembled by sequencing in both direction using primers A and B mentioned

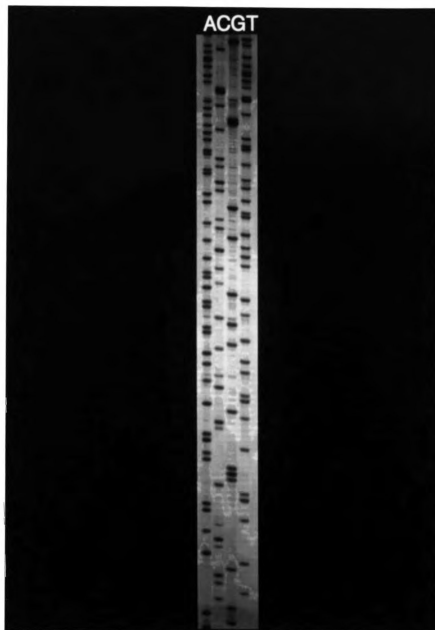


Figure 4. Sequencing gel autoradiograph. Photograph of sequencing gel autoradiograph demonstrating sequences in the hypervariable region 1 of the mtDNA control region. The sequence shown represent one Puerto Rican sample.

in Table 1. The sequence generated by using primer B was reversed and complemented using the reverse program of GCG. The new sequence was aligned to the sequence from primer A using the bestfit program of GCG. An overlap was determined and the two sequences were combined to give 468 bp.

The bestfit program (this program uses the local homology algorithm of Smith and Waterman (Advanced in Applied Mathematics 2; 484-489 (1981)) to find the best segment of similarity between two sequences) of GCG was used to compare all the sequences with each other. The fasta program (this program uses the method of Pearson and Lipman, (1988) to search for similarities between a designated sequence and sequences in the Genbank data base), GCG was used to find a sequence in the Genbank and EMBL data bases, similar to the generated sequence.

The pileup program creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments (simplification of the progressive alignment method of Feng and Doolittle, 1987). It also plots a tree (dendrogram) showing the clustering relationships used to create the alignment. This program of GCG was used to find similarities and differences between different individuals in the group, in order to construct a dendrogram (phylogenetic tree) which best represented the relationship between all individuals compared in the group.

## RESULTS

One hundred ng of the total DNA (see material and methods) was used for PCR amplification as described in material and methods. After 40 cycles of amplification, the products were analyzed on 8% acrylamide gels. Figure 5 shows typical amplification reaction results. The negative control and the molecular weight marker were indicated. The 520 bp fragments of interest were cut and eluted from the gel as described in material and methods. Figure 6 shows a typical gel after fragment purification.

Three ul (approximately 1.0 ug) of the purified PCR product was used in each sequencing reaction. Sequencing was done in a LPCR as described in material and methods. Each sequencing reaction gave 230 - 280 bp, so by using primer A and B, we were able to sequence the 500 bp fragment. The sequence data generated by primer B were reversed and complemented. The new sequence were aligned to the sequence data from primer A (see material and methods). The 468 bp sequence of the mitochondrial control region for each sample was compared with all other samples in the group (total of 50 samples) using the bestfit program of GCG. The primary sequence data showed 95-99% homology with each other.

Five randomly selected samples were used as a template, to search for a homologous sequence in the Genbank and EMBL data bases using the fasta program of the GCG. The 5

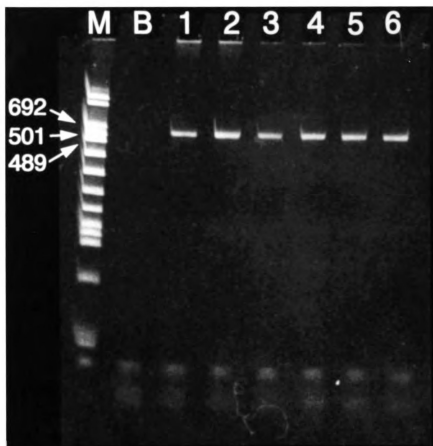


Figure 5. PCR amplification of Mitochondrial DNA. Polyacrylamide gel electrophoresis of the PCR product, after 35 cycles of amplification. lane M is the molecular weight marker, lane B is negative control, lanes 1-6 are different samples from the Puerto Rican populations. Shown above is the 520 bp fragment amplified using primer L15926 and H 16498 specific for the hypervariable segment 1 of the mtDNA control region.

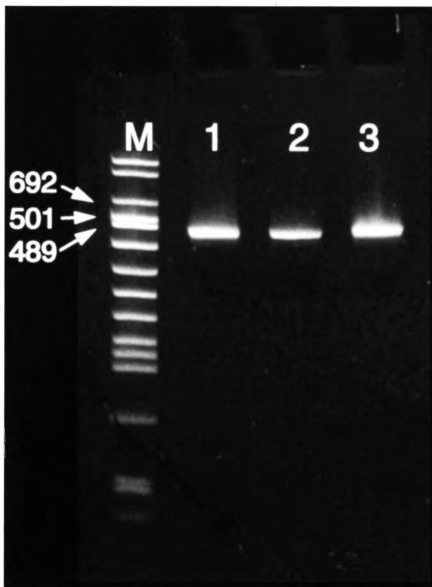


Figure 6. Purified PCR product. Polyacrylamide gel electrophoresis of the purified PCR product, after elution from the gel, and concentration by centricon-100 microconcentrator, lane M is the molecular weight marker, lanes 1-3 are different samples.

separate searches came up with 95-99% homology to the human mitochondrial DNA control region. It also showed different homology with the control region for different ethnic groups.

Comparison of the control region partial sequence from the 50 Puerto Ricans included in this study, identified a total of 266 nucleotide substitutions distributed between 84 sites, and 12 single nucleotide length changes distributed at 11 sites. Table 2 shows the sequence data that will be later used in phylogenetic analysis, and Table 3 lists the 92 variable nucleotide sites and the substitutions found. The system used to number the nucleotide is one in which the first base of our sequence (1) is equivalent to base number 15970, and the last base (468) is equivalent to 16436 in the numbering of the standard reference (Anderson et al., 1981). Table 4 is a list of the variable nucleotide positions of Table 3 according to both numbering systems.

#### **Distribution of mutated sites**

With the exception of nucleotide positions 141, 286, and 332 which apparently underwent both substitution and length mutation events, the remaining mutated sites included either length polymorphism or substitutions. The distribution of the mutations in the control region sequences from 50 Puerto Ricans revealed the polymorphism profile illustrated in Figure 7. The histogram represents the total number of mutations within continuous

**Table 2. Puerto Ricans mitochondrial DNA sequences.**  
**Comparison of all Puerto Rican mitochondrial DNA control region partial nucleotide sequences with the reference sequence "ANDE" (Anderson et al., 1981). The system used to number the nucleotide is one in which the first base is equivalent to 15970 in the reference sequence. Dots (.) indicate identity with the reference sequence, question mark (?) indicate undecided sequence, and the dash sign (-) indicate deletion.**

Table 2. Puerto Rican mitochondrial DNA sequences.

| ID   | 1          |            |            |           | 50         |
|------|------------|------------|------------|-----------|------------|
| ANDE | TTAACTCCAC | CATTAGCACC | CAAAGCTAAG | ATTCTAATT | AAACTATTCT |
| 366  | .....      | .....      | .....      | .....     | .....      |
| 398  | .....      | .....      | .....      | .....     | .....      |
| 416  | .....      | .....      | .....      | .....     | .....      |
| 85   | .....      | .....      | .....      | .....     | .....      |
| 424  | .....      | .....      | .....      | .....     | .....      |
| 425  | .....      | .....      | .....      | .....     | .....      |
| 418  | .....      | .....      | .....      | .....     | .....      |
| 404  | .....      | .....      | .....      | .....     | .....      |
| 394  | .....      | .....      | .....      | .....     | .....      |
| 426  | .....      | .....      | .....      | .....     | .....      |
| 385  | .....      | .....      | .....      | .....     | .....      |
| 417  | .....      | .....      | .....      | .....     | .....      |
| 413  | .....      | .....      | .....      | .....     | .....      |
| 401  | .....      | .....      | .....      | .....     | .....      |
| 415  | .....      | .....      | .....      | .....     | .....      |
| 368  | .....      | .....      | .....      | .....     | .....      |
| 391  | .....      | .....      | .....      | .....     | .....      |
| 107  | .....      | .....      | .....      | .....     | .....      |
| 369  | .....      | .....      | .....      | .....     | .....      |
| 387  | .....      | .....      | .....      | .....     | .....      |
| 163  | .....      | .....      | .....      | .....     | .....      |
| 410  | .....      | .....      | .....      | .....     | .....      |
| 377  | .....      | .....      | .....      | .....     | .....      |
| 363  | .....      | .....      | .....      | .....     | .....      |
| 397  | .....      | .....      | .....      | .....     | .....      |
| 376  | .....      | .....      | .....      | .....     | .....      |
| 408  | .....      | .....      | .....      | .....     | .....      |
| 409  | .....      | .....      | .....      | .....     | .....      |
| 393  | .....      | .....      | .....      | .....     | .....      |
| 382  | .....      | .....      | .....      | .....     | .....      |
| 362  | .....      | .....      | .....      | .....     | .....      |
| 371  | .....      | .....      | .....      | .....     | .....      |
| 390  | .....      | .....      | .....      | .....     | .....      |
| 399  | .....      | .....      | .....      | .....     | .....      |
| 92   | .....      | .....      | .....      | .....     | .....      |
| 389  | .....      | .....      | .....      | .....     | .....      |
| 420  | .....      | .....      | .....      | .....     | .....      |
| 392  | .....      | .....      | .....      | .....     | .....      |
| 400  | .....      | .....      | .....      | .....     | .....      |
| 364  | .....      | .....      | .....      | .....     | .....      |
| 87   | .....      | .....      | .....      | ..C.....  | .....      |
| 83   | .....      | .....      | .....      | .....     | .....      |
| 406  | .....      | .....      | .....      | .....     | .....      |
| 414  | .....      | .....      | .....      | .....     | .....      |
| 110  | .....      | .....      | .....      | .....     | .....      |
| 113  | .....      | .....      | .....      | .....     | .....      |
| 421  | .....      | .....      | .....      | .....     | .....      |
| 422  | .....      | .....      | .....      | .....     | .....      |
| 419  | ....T..... | .....      | ....G....  | .....     | .....      |
| 360  | .....      | .....      | .....      | .....     | .....      |

Table 2 (cont'd).

| ID   | 51         |            |           |            | 100        |
|------|------------|------------|-----------|------------|------------|
| ANDE | CTGTTCTTTC | ATGGGGAAGC | AGATTGGGT | ACCACCCAAG | TATTGACTCA |
| 366  | .....      | .....      | .....     | .....      | .....      |
| 398  | ..A.....   | .....      | .....     | .....      | .....      |
| 416  | .....      | .....      | .....     | .....      | .....      |
| 85   | .....      | .....      | .....     | .....      | .....      |
| 424  | ..A.....   | .....      | .....     | .....      | .....      |
| 425  | .....      | .....      | .....     | .....      | .....      |
| 418  | .....      | .....      | .....     | .....      | .....      |
| 404  | .....      | .....      | .....     | .....      | .....      |
| 394  | .....      | .....      | .....     | .....      | .....      |
| 426  | .....      | .....      | .....     | .....      | .....      |
| 385  | .....      | .....      | .....     | .....      | .....      |
| 417  | .....      | .....      | .....     | .....      | .....      |
| 413  | .....      | .....      | .....     | .....      | .....      |
| 401  | .....      | .....      | .....     | .....      | .....      |
| 415  | .....      | .....      | .....     | .....      | .....      |
| 368  | .....      | .....      | .....     | .....      | .....      |
| 391  | .....      | .....      | .....     | .....      | .....      |
| 107  | .....      | .....      | .....     | .....      | .....      |
| 369  | .....      | .....      | .....     | .....      | .....      |
| 387  | .....      | .....      | .....     | .....      | .....      |
| 163  | .....      | .....      | .....     | .....      | .....      |
| 410  | .....      | .....      | .....     | .....      | .....      |
| 377  | .....      | .....      | .....     | .....      | .....      |
| 363  | .....      | .....      | .....     | .....      | .....      |
| 397  | .....      | .....      | .....     | .....      | .....      |
| 376  | .....      | .....      | .....     | .....      | .....      |
| 408  | .....      | .....      | .....     | .....      | .....      |
| 409  | .....      | .....      | .....     | .....      | .....      |
| 393  | .....      | .....      | .....     | .....      | .....      |
| 382  | .....      | G.....     | .....     | .....      | .....      |
| 362  | .....      | .....      | .....     | .....      | .....      |
| 371  | ..?.....   | .....      | .....     | .....      | .....      |
| 390  | .....      | .....      | .....     | .....      | .....      |
| 399  | .....      | .....      | .....     | .....      | .....      |
| 92   | .....      | .....      | .....     | .....      | .....      |
| 389  | .....      | .....      | .....     | .....      | .....      |
| 420  | .....      | .....      | .....     | .....      | .....      |
| 392  | .....      | .....      | .....     | .....      | .....      |
| 400  | .....      | .....      | .....     | .....      | .....      |
| 364  | .....      | .....      | .....     | .....      | .....      |
| 87   | .....      | .....      | .....     | .....      | .....      |
| 83   | .....      | .....      | .....     | .....      | .....      |
| 406  | .....      | .....      | .....     | .....      | .....      |
| 414  | .....      | .....      | .....     | .....      | .....      |
| 110  | .....      | .....      | .....     | .....      | .....      |
| 113  | .....      | .....      | .....     | .....      | .....      |
| 421  | .....      | .....      | .....     | .....      | .....      |
| 422  | .....      | .....      | .....     | .....      | .....      |
| 419  | .....      | T.....     | .....     | .....      | .....      |
| 360  | .....      | .....      | .....     | .....      | .....      |

Table 2 (cont'd)

| ID   | 101        |            |            |            | 150        |
|------|------------|------------|------------|------------|------------|
| ANDE | CCCATCAACA | ACCGCTATGT | ATTTCGTACA | TTACTGCCAG | CCACCATGAA |
| 366  | .....      | .....      | .....      | .....      | .....      |
| 398  | .....      | .....      | .....G..   | .....      | .....      |
| 416  | .....      | .....      | .....      | .....      | T.....     |
| 85   | .....      | .....      | .....      | .....      | .....      |
| 424  | .....      | .....      | .....      | .....      | .....      |
| 425  | .....      | .....      | .....      | .....      | .....      |
| 418  | .....      | .....      | .....      | .....      | .....      |
| 404  | .....      | .....      | .....      | -.....     | .....      |
| 394  | .....      | .....      | .....      | .....      | .....      |
| 426  | .....      | .....      | .....      | .....A     | .....      |
| 385  | .....      | .....      | .....      | .....      | .....      |
| 417  | .....      | .....      | .....      | .....      | .....      |
| 413  | .....      | .....      | .....      | .....      | .....      |
| 401  | .....      | .....      | .....      | .....      | .....      |
| 415  | .....      | .....      | .....      | .....      | .....      |
| 368  | .....      | .....      | .....      | .....      | .....      |
| 391  | .....      | .....      | .....      | .....      | .....      |
| 107  | .....      | .T..?      | .....?     | .....      | T?.....    |
| 369  | .....      | .T..?      | .....      | .....      | T.....     |
| 387  | .....      | .T..?      | .....      | .....      | T?.....    |
| 163  | .....      | .....      | .....      | .....      | T.....     |
| 410  | .....      | .....      | .....?     | .....      | T?.....    |
| 377  | .....      | .....      | .....?     | .....      | T?.....    |
| 363  | .....      | .....      | .....      | .....      | T.....     |
| 397  | .....      | .....      | .....CG..  | .....      | T.....     |
| 376  | .....      | .....      | .....?     | .....      | -?.....    |
| 408  | .....      | .....C..   | .....      | .....      | .....      |
| 409  | .....      | .....      | .....      | .....      | .....      |
| 393  | .....      | .....      | .....      | .....      | .....      |
| 382  | .....      | .....      | .....      | .....      | .....      |
| 362  | .....      | .....      | .....      | .....      | .....      |
| 371  | .....      | .....      | .....      | .....      | .....      |
| 390  | .....      | .....      | .....      | .....      | .....      |
| 399  | .....      | .....?     | .....?     | .....      | ?.....     |
| 92   | .....      | .....      | .....      | .....      | .....      |
| 389  | .....      | .....      | .....      | .....      | .....      |
| 420  | .....      | .....      | .....C..   | .....      | .....      |
| 392  | .....      | .....      | .....      | .....      | .....      |
| 400  | .....      | .....      | .....      | .....      | T.....     |
| 364  | .....      | .....      | .....?     | .....      | .....      |
| 87   | .....      | .....      | .....      | .....      | .....      |
| 83   | .....      | .....      | .....      | .....      | .....      |
| 406  | .....      | .....      | .....      | .....      | TT.....    |
| 414  | .....      | .....      | .....      | .....      | .....      |
| 110  | .....      | .....      | .....      | .....      | T.....     |
| 113  | .....      | .....      | .....      | .....      | .....      |
| 421  | .....      | .....      | .....      | .....      | .....      |
| 422  | .....      | .....      | .....T.C.. | .....      | .....      |
| 419  | .....      | .....      | .....      | .....      | .....      |
| 360  | .....      | .....      | .....      | .....      | .....      |

Table 2 (cont'd)

| ID   | 151        |            |            |            | 200        |
|------|------------|------------|------------|------------|------------|
| ANDE | TATTGTACGG | TACCATAAAT | ACTTGACCAC | CTGTAGTACA | TAAAAACCCA |
| 366  | .....      | .....      | .....      | .....      | .....      |
| 398  | .....      | .....      | .....      | .....      | .....      |
| 416  | .....      | .....      | .....      | .....      | .....      |
| 85   | .....A.    | .....      | .....      | .....      | .....      |
| 424  | .....A.    | .....      | .....      | .....      | .....      |
| 425  | .....A.    | .....      | .....      | .....      | .....      |
| 418  | .....      | .....      | .....      | .....      | .....      |
| 404  | .....A.    | .....      | .....      | .....      | .....      |
| 394  | .....C.    | .....      | .....      | .....      | .....      |
| 426  | .....      | .....      | .....A.    | .....      | .....      |
| 385  | .....C.    | .....?     | .....      | .....      | .....      |
| 417  | ...C.C.    | .....      | .....      | .....      | .....      |
| 413  | .....C.    | .....      | .....      | .....      | .....      |
| 401  | .....C.    | .....      | .....      | .....      | .....      |
| 415  | .....      | .....      | .....      | .....      | .....      |
| 368  | .....A.    | .....      | .....      | .....      | .....      |
| 391  | .....A.    | .....?     | .....      | .....      | .....      |
| 107  | .....C.    | .....      | .....      | .....      | .....      |
| 369  | .....      | .....      | .....      | .....      | .....      |
| 387  | .....      | .....      | .....      | .....      | .....      |
| 163  | .....      | .....      | .....      | .....      | .....      |
| 410  | .....      | .....      | .....      | .....      | .....      |
| 377  | .....      | .....      | .....      | .....      | .....      |
| 363  | .....      | .....      | .....      | .....      | .....      |
| 397  | .....      | .....?     | .....      | .....      | .....      |
| 376  | .....      | .....      | .....      | .....      | .....      |
| 408  | .....      | .....?     | .....      | .....      | .....      |
| 409  | .....      | .....?     | .....      | .....      | .....      |
| 393  | .....      | .....      | .....      | .....      | .....      |
| 382  | .....C.    | .....      | .....      | .....      | .....      |
| 362  | .....      | .....      | .....      | .....      | .....      |
| 371  | .....      | .....      | .....      | .....      | .....      |
| 390  | .....      | .....      | .....      | .....      | .....      |
| 399  | .....      | .....      | .....      | .....      | .....      |
| 92   | ...C.      | .....      | .....      | .....      | .....      |
| 389  | .....      | .....?     | .....      | .....      | .....      |
| 420  | ...C.      | .....      | .....      | ...C.      | .....      |
| 392  | ...C.      | .....      | .....      | .....      | ...G.      |
| 400  | .....      | .....      | .....      | .....      | .....      |
| 364  | .....      | ...-       | .....      | .....      | .....      |
| 87   | .....      | .....      | ...T.      | .....      | .....      |
| 83   | .....      | .....      | .....      | .....      | .....      |
| 406  | ..C.       | .....      | .....      | .....      | .....      |
| 414  | ...C.      | .....      | .....      | .....      | .....      |
| 110  | .....      | .....      | .....      | .....      | .....      |
| 113  | .....      | .....      | .....      | .....      | .....      |
| 421  | .....A.    | .....      | .....      | .....      | .....      |
| 422  | .....      | .....      | .....      | .....      | .....      |
| 419  | .....      | ...C.      | .....      | .....      | .....      |
| 360  | .....      | .....      | .....      | .....      | .....      |

Table 2 (cont'd)

| ID   | 201         |            |             |             | 250         |
|------|-------------|------------|-------------|-------------|-------------|
| ANDE | ATCCACATCA  | AAACCCCCTC | CCCATGCTTA  | CAAGCAAGTA  | CAGCAATCAA  |
| 366  | .....       | .....      | .....       | .....       | .....       |
| 398  | .....       | .....      | .....       | .....       | .....       |
| 416  | .....T.     | .....      | .....       | .....       | .....       |
| 85   | .....       | .....      | .....       | .....?      | .....       |
| 424  | .....?      | .....      | .....       | .....       | .....A..... |
| 425  | .....       | .....      | .....       | .....       | .....A..... |
| 418  | .....       | .....      | .....       | .....       | .....       |
| 404  | .....       | .....      | .....A.     | .....       | .....       |
| 394  | .....       | .....      | .....       | .....       | .....       |
| 426  | .....       | .....      | .....       | .....       | .....       |
| 385  | .....       | .....      | .....       | .....       | .....       |
| 417  | .....       | .....      | .....       | .....       | .....       |
| 413  | .....       | .....T.C.  | .....G..... | .....       | .....       |
| 401  | .....       | .....      | .....       | .....       | .....       |
| 415  | .....       | .....      | .....?      | .....?      | .....       |
| 368  | .....       | .....      | .....       | .....C.     | .....       |
| 391  | .....       | .....      | .....       | .....C.     | .....       |
| 107  | .....       | .....      | .....       | .....       | .....       |
| 369  | .....       | .....      | .....       | .....       | .....       |
| 387  | .....       | .....      | .....       | .....       | .....       |
| 163  | .....       | .....      | .....       | .....       | .....       |
| 410  | .....       | .....      | .....       | .....       | .....       |
| 377  | .....       | .....      | .....?      | .....       | .....       |
| 363  | .....       | .....      | .....       | .....       | .....       |
| 397  | .....       | .....      | .....A.     | .....       | .....       |
| 376  | .....       | .....      | .....       | .....       | .....G      |
| 408  | .....       | .....      | .....       | .....       | .....G.     |
| 409  | .....C..... | .....      | .....       | .....       | .....G.     |
| 393  | .....       | .....      | .....       | .....       | .....       |
| 382  | .....       | .....      | .....       | .....       | .....       |
| 362  | .....       | .....      | .....       | .....       | .....       |
| 371  | .....       | .....      | .....?      | .....       | .....       |
| 390  | .....       | .....      | .....       | .....       | .....       |
| 399  | .....       | .....      | .....       | .....?      | .....       |
| 92   | .....       | .....      | .....       | .....       | .....       |
| 389  | .....       | .....      | .....?      | .....C.     | .....       |
| 420  | .....       | .....      | .....       | .....       | .....       |
| 392  | .....       | .....      | .....       | .....       | .....       |
| 400  | .....       | .....      | .....       | .....       | .....       |
| 364  | .....       | .....      | .....       | .....       | .....       |
| 87   | .....T..... | .....      | .....       | .....G..... | .....       |
| 83   | .....       | .....      | .....       | .....       | .....       |
| 406  | .....       | .....      | .....       | .....-      | .....       |
| 414  | .....       | .....      | .....       | .....       | .....       |
| 110  | .....       | .....      | .....       | .....       | .....       |
| 113  | .....       | .....      | .....       | .....?      | .....       |
| 421  | .....       | .....      | .....       | .....       | .....       |
| 422  | .....       | .....      | .....       | .....       | .....       |
| 419  | .....C..... | .....      | .....A.C.   | .....C.     | .....       |
| 360  | .....       | .....      | .....       | .....       | .....       |

Table 2 (cont'd)

| ID   | 251        |            |            |            | 300        |
|------|------------|------------|------------|------------|------------|
| ANDE | CCCTCAACTA | TCACACATCA | ACTGCAACTC | CAAAGCCACC | CCTCACCCAC |
| 366  | ..T.....   | .....      | .....      | .....      | .....      |
| 398  | ..T.....   | .....      | .....      | .....      | .....      |
| 416  | ..T.....   | .....T.    | .....      | .....      | .....      |
| 85   | ..T.....   | .....?     | .....      | .....      | .....      |
| 424  | ..T.....   | .....      | .....      | .....      | .....      |
| 425  | ..T.....   | .....      | .....      | .....      | .....      |
| 418  | ..T.....   | .....      | .....      | .....      | .....      |
| 404  | ..T.....   | .....      | .....      | .....      | .....      |
| 394  | ..T.....   | .....?     | .....      | .....      | .....T     |
| 426  | ..T.....   | .....      | .....      | .....      | .....      |
| 385  | ..T.....   | .....      | .....      | T.....     | .....T     |
| 417  | ..T.....   | .....      | .....      | .....      | .....T     |
| 413  | ..T.....   | .....      | .....      | .....      | .....T     |
| 401  | ..T.....   | .....?     | .....      | .....      | .....T     |
| 415  | ..T.....   | .....?     | .....      | .....      | .....      |
| 368  | ..T.T..... | .....G     | .....      | .....      | .....      |
| 391  | ..T.....   | .....??    | ?.....     | .....      | .....      |
| 107  | ..T.....   | .....?     | .....      | .....T?    | .....      |
| 369  | ..T.....   | .....      | .....      | .....T     | .....      |
| 387  | ..T.....   | .....      | .....      | .....T     | .....      |
| 163  | ..T.....   | .....      | .....      | .....      | .....      |
| 410  | ..T.....   | .....?     | .....      | .....      | .....      |
| 377  | ..T.....   | .....?     | .....      | .....      | .....      |
| 363  | ..T.....   | .....      | .....      | .....      | .....      |
| 397  | ..T.....   | .....?     | .....      | .....      | .....      |
| 376  | ..T.....   | .....?     | .....      | .....      | .....      |
| 408  | .....      | .....?     | .....      | .....      | .....      |
| 409  | .....      | .....?     | .....      | .....      | .....      |
| 393  | .....      | .....?     | .....      | .....      | .....T     |
| 382  | .....      | .....      | .....      | .....      | .....      |
| 362  | ..T.....   | .....      | .....      | .....      | .....      |
| 371  | ..T.....   | .....?     | .....      | .....      | .....      |
| 390  | ..T.....   | .....?     | .....      | .....      | .....      |
| 399  | ..T.....   | .....      | .....      | .....      | .....      |
| 92   | ..T.....   | .....      | .....      | .....      | .....      |
| 389  | ..T.....   | .....?     | .....      | .....      | .....      |
| 420  | ..T.....   | .....      | .....      | .....      | .....      |
| 392  | .....      | .....?     | .....      | .....      | .....      |
| 400  | ..T.....   | .....?     | .....      | .....      | .....      |
| 364  | .....      | .....?     | .....      | .....      | .....      |
| 87   | ..T.....   | .....      | .....      | .....      | .....      |
| 83   | ..T.....   | .....      | .....      | .....      | .....      |
| 406  | ..T.....   | .....?     | .....      | .....      | .....      |
| 414  | .....      | .....?G    | .....      | .....--    | .....      |
| 110  | .....      | .....G     | .....      | .....?     | .....      |
| 113  | .....      | .....?     | .....      | .....A.    | .....      |
| 421  | .....      | .....      | .....C.    | .....      | .....      |
| 422  | ..T.....   | .....      | .....      | .....-     | .....      |
| 419  | ..T.....   | .....      | .....      | .....      | .....      |
| 360  | ..T.....   | .....      | .....      | .....      | .....      |

Table 2 (cont'd)

| ID   | 301        |          |   |            |            | 350        |
|------|------------|----------|---|------------|------------|------------|
| ANDE | TAGGATACCA | ACAAACCT | A | CCCACCCTTA | ACAGTACATA | GTACATAAAG |
| 366  | .....      | .....C   | . | .....C.    | .....      | .....      |
| 398  | .....      | .....    | . | .....      | .....      | .....      |
| 416  | .....      | .....    | . | .....      | .....      | .C.....    |
| 85   | .....      | .....    | . | T.....     | .....      | .....A     |
| 424  | .....      | .....    | . | .....      | .....      | .....      |
| 425  | .....T..   | .....    | . | .....      | .....      | .C.....    |
| 418  | .....T..   | .....    | . | ....T....  | .....G     | .....      |
| 404  | .....      | .....    | . | T.....     | .....      | .....      |
| 394  | .....T..   | .....    | . | ..G.....   | .T.....    | .C.....    |
| 426  | .....      | .....    | . | ..G.....   | .....      | .....      |
| 385  | .....T..   | .....    | . | ..G.....   | .....      | .C.....    |
| 417  | .....T..   | .....    | . | .....      | .....      | .....      |
| 413  | .....T..   | .....    | . | ..G.....   | .....      | .C.....    |
| 401  | .....T..   | .....C   | . | ..G.....   | .....      | .C.....    |
| 415  | .....      | .....C   | . | .....      | .....      | .....      |
| 368  | .....      | .....    | . | ..T..T...  | .....      | .C.....    |
| 391  | .....      | .....    | . | ..T..T...  | .....      | .C.....    |
| 107  | .....G..   | .....    | . | T.....     | .....      | .....?..A  |
| 369  | .....      | .....    | . | T.....     | .....      | .....A     |
| 387  | .....      | .....    | . | T.....     | .....      | .....A     |
| 163  | .....      | .....    | . | T.....     | .....      | .....A     |
| 410  | .....      | .....    | . | T.....     | .....      | .....?..A  |
| 377  | .....      | .....    | . | T.....?    | .....      | .....?..A  |
| 363  | .....      | .....    | . | T.....     | .....      | .....A     |
| 397  | .....      | .....    | . | T.....     | .....      | .....A     |
| 376  | .....      | .....    | . | T.....     | .....      | .....?..A  |
| 408  | .....T..   | .....    | . | .....      | .....      | .....      |
| 409  | .....T..   | .....    | . | .....      | .....      | .....      |
| 393  | .....      | .....    | . | .....      | .....      | .....      |
| 382  | .....      | .....    | . | .....      | .....      | .....      |
| 362  | .....      | .....    | . | .....      | .....      | .....      |
| 371  | .....      | .....    | . | .....      | .....      | .....      |
| 390  | .....      | .....    | . | .....C.    | .....      | .....      |
| 399  | .....      | .....    | . | .....C.    | .....      | .....      |
| 92   | .....      | .....    | . | .....C.    | .....      | .....      |
| 389  | .....      | .....    | . | .....      | .....      | .....      |
| 420  | .....      | .....    | . | .....      | .....      | .....      |
| 392  | .....      | .....    | . | .....      | G.....     | .....G..A  |
| 400  | .....T..   | .....    | . | ..GT....   | .....      | .C.....    |
| 364  | .....      | .....    | . | .....?     | .....      | .....      |
| 87   | .....      | .....    | . | .....      | .....      | ..G.....   |
| 83   | .....      | .....    | . | .....C.    | .....      | .....      |
| 406  | .....      | .....    | . | .....      | .....      | .....      |
| 414  | .....T..   | .....    | . | .....      | .....      | .C.....    |
| 110  | .....      | .....    | . | .....      | -.....     | .....      |
| 113  | ....-?..   | .....    | . | .....      | .....      | .....?     |
| 421  | .....T..   | .....    | . | ....T....  | .....      | .C.....    |
| 422  | .....T..   | .....    | . | ....T....  | .....G     | .....      |
| 419  | .....      | .....    | . | .....      | .....      | .....      |
| 360  | .....      | .....    | . | .....?     | .....      | .....      |

Table 2 (cont'd)

| ID   | 351        |            |            |            | 400        |
|------|------------|------------|------------|------------|------------|
| ANDE | CCATTTACCG | TACATAGCAC | ATTACAGTCA | AATCCCTTCT | CGTCCCCATG |
| 366  | .....C.T.. | .....      | .....      | .....      | .....      |
| 398  | .....T..   | .....      | .....?     | .....      | .....      |
| 416  | T.....     | .....      | .....?     | .....      | ?..C.....  |
| 85   | .....      | .....      | .....?     | .....      | ?..C.....  |
| 424  | .....      | .....      | .....?     | .....      | .....      |
| 425  | .....      | .....      | .....?     | .....      | ?..C.....  |
| 418  | .....      | .....      | .....?     | .....      | .....      |
| 404  | .....      | .....      | .....?     | .....      | ..C.....   |
| 394  | .....      | .....      | .....T.    | .....      | .....      |
| 426  | .....      | .....      | .....?     | .....      | ..C.....   |
| 385  | .....      | .....      | .....?     | .....      | ?.....     |
| 417  | .....      | .....      | .....?     | .....      | ?.....     |
| 413  | .....      | .....      | .....?     | .....      | .....      |
| 401  | .....      | .....      | ..G.....?  | .....?     | ?.....     |
| 415  | T.....     | .....      | .....?     | .....      | .....      |
| 368  | .....      | .....      | .....?     | .....      | .....      |
| 391  | .....      | .....      | .....?     | .....      | ?.....     |
| 107  | .....      | .....      | .....?     | .....      | ?..C.....  |
| 369  | .....      | .....      | .....T.    | .....      | ?..C.....  |
| 387  | .....      | .....      | .....      | .....      | ..C.....   |
| 163  | .....      | .....      | .....?     | .....      | ?..C.....  |
| 410  | .....      | .....      | .....?     | .....      | ?..C.....  |
| 377  | .....      | .....      | .....?     | .....?     | ?..?       |
| 363  | .....      | .....T.    | .....?     | .....C.    | ?..C.....  |
| 397  | T.....     | .....      | .....?     | .....T.    | ?..C.....  |
| 376  | .....      | .....      | .....?     | .....?     | ?..C.....  |
| 408  | .....      | .....      | .....?     | .....      | .....      |
| 409  | .....      | .....      | .....?     | .....      | .....      |
| 393  | T.....     | .....      | .....?     | .....      | ?.....     |
| 382  | .....      | .....      | .....?     | .....      | .....      |
| 362  | .....T.    | .....      | .....?     | .....      | .....      |
| 371  | .....T.    | .....      | .....?     | .....      | .....      |
| 390  | .....C.T.  | .....      | .....?     | .....      | .....      |
| 399  | .....C.T.  | .....      | .....?     | .....      | ?.....     |
| 92   | .....C.T.  | .....      | .....?     | .....      | .....      |
| 389  | .....C.T.  | .....      | .....?     | .....      | ?.....     |
| 420  | .....      | .....      | .....?     | .....      | ?.....     |
| 392  | .....      | .....      | ..G.....   | .....      | .....      |
| 400  | .....      | .....      | .....?     | .....      | ?.....     |
| 364  | .....C.T.  | .....      | .....?     | .....      | ?.....     |
| 87   | .....T.    | .....      | .....?     | .....      | .....      |
| 83   | .....C.T.  | .....      | .....?     | .....-     | .....      |
| 406  | .....      | .....      | .....?     | .....      | .....      |
| 414  | .....      | .....      | .....?     | .....      | ?..C.....  |
| 110  | .....      | .....      | .....      | .....      | .....      |
| 113  | .....      | .....      | .....?     | .....      | ?.....     |
| 421  | .....      | .....      | .....?     | .....      | T.....     |
| 422  | .....      | .....      | .....      | .....      | .....      |
| 419  | T.....     | .....      | .....?     | .....      | .....      |
| 360  | .....C.T.  | .....      | .....?     | .....      | .....      |

Table 2 (cont'd)

| ID   | 401        |            |            |            | 450        |
|------|------------|------------|------------|------------|------------|
| ANDE | GATGACCCCC | CTCAGATAGG | GGTCCCTTGA | CCACCATCCT | CCGTGAAATC |
| 366  | ...?.....  | .....      | ...?.....  | .....      | .....      |
| 398  | .....      | .....      | ...?.....  | .....      | .....      |
| 416  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 85   | .....      | .....      | ...?.....  | .....      | .....      |
| 424  | .....      | .....      | A.....     | .....      | .....      |
| 425  | .....      | .....      | A.....     | .....      | .....?     |
| 418  | .....      | .....      | A.....     | .....      | .....?     |
| 404  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 394  | .....      | .....      | .....      | .....      | .....T     |
| 426  | .....      | .....      | A.....G    | .....      | .....      |
| 385  | ...?.....  | .....      | ...?.....  | .....      | ...?..?    |
| 417  | ...?.....  | ...?.....  | ...?.....  | .....      | ...?..T    |
| 413  | .....      | .....      | ...?.....  | .....      | .....      |
| 401  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 415  | .....      | .....      | ...?.....G | .....      | .....      |
| 368  | .....      | .....      | .....      | .....      | .....      |
| 391  | ...?.....  | .....      | ...?.....  | .....      | ...?..?    |
| 107  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 369  | .....      | .....      | .....      | .....      | .....      |
| 387  | .....      | .....      | .....      | .....      | ...A.....  |
| 163  | .....      | ..A.....   | ...?.....  | .....      | .....      |
| 410  | .....      | .....      | ...?.....  | .....      | .....?     |
| 377  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 363  | .....      | .....      | .....      | .....      | .....      |
| 397  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 376  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 408  | ...?.....  | .....      | .....      | .....      | .....      |
| 409  | .....      | .....      | ...?.....  | .....      | .....      |
| 393  | .....      | .....      | ...?.....  | .....      | .....      |
| 382  | .....      | .....      | ...?.....  | .....      | .....?     |
| 362  | .....      | .....      | ...?.....  | .....      | .....      |
| 371  | .....      | .....      | ...?.....  | .....      | .....      |
| 390  | .....      | .....      | .....      | .....      | .....      |
| 399  | ...?.....  | .....      | ...?.....  | .....      | .....      |
| 92   | .....      | .....      | ...?.....  | .....      | .....      |
| 389  | ...?.....  | .....      | .....      | .....      | .....?     |
| 420  | ...?.....  | .....      | ...?.....  | .....      | .....      |
| 392  | .....      | .....      | ...?.....  | .....      | .....      |
| 400  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 364  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 87   | .....      | .....      | .....      | .....      | .....      |
| 83   | .....      | .....      | .....      | .....      | .....      |
| 406  | ...?.....  | .....      | ...?.....  | .....      | .....      |
| 414  | ...?.....  | .....      | ...?.....  | .....      | ...?..?    |
| 110  | .....      | .....      | .....      | .....      | .....      |
| 113  | ...?.....  | .....      | ...?.....  | .....      | .....?     |
| 421  | .....      | .....      | ...?.....  | .....      | .....?     |
| 422  | .....      | .....      | A.....     | .....      | .....      |
| 419  | .....      | .....      | ...?.....  | .....      | .....G     |
| 360  | ...?.....  | ..?.....   | ...?.....  | .....      | .....      |

Table 2 (cont'd)

| ID   | 451        | 468      |
|------|------------|----------|
| ANDE | AATATCCCGC | ACAAGAGT |
| 366  | .....      | .....    |
| 398  | .....      | .....    |
| 416  | .....      | .....    |
| 85   | .....      | .....    |
| 424  | .....      | .....    |
| 425  | .....      | .....    |
| 418  | .....      | .....    |
| 404  | .....      | .....    |
| 394  | .....T.A.  | .....    |
| 426  | .....      | .....    |
| 385  | .....      | .....    |
| 417  | .....      | .....    |
| 413  | .....      | .....    |
| 401  | .....      | .....    |
| 415  | .....      | .....    |
| 368  | .....      | .....    |
| 391  | .....      | .....    |
| 107  | .....      | .....    |
| 369  | .....      | .....    |
| 387  | .....      | .....    |
| 163  | .....      | .....    |
| 410  | .....      | .....    |
| 377  | .....-     | .....    |
| 363  | .....      | .....    |
| 397  | .....      | .....    |
| 376  | .....      | .....    |
| 408  | .....      | .....    |
| 409  | .....      | .....    |
| 393  | .....      | .....    |
| 382  | .....      | .....    |
| 362  | .....      | .....    |
| 371  | .....      | .....    |
| 390  | .....      | .....    |
| 399  | .....      | .....    |
| 92   | .....      | .....    |
| 389  | .....      | .....    |
| 420  | .....      | .....    |
| 392  | .....      | .....    |
| 400  | .....      | .....    |
| 364  | .....      | .....    |
| 87   | .....      | .....    |
| 83   | .....      | .....    |
| 406  | .....      | .....    |
| 414  | .....      | .....    |
| 110  | .....      | .....    |
| 113  | .....      | .....    |
| 421  | .....      | .....    |
| 422  | .....      | .....    |
| 419  | .....      | .....    |
| 360  | .....      | .....    |

**Table 3.** Base substitution and length mutations.  
Base substitutions and length mutations at 92 nucleotide positions found in a survey of 50 Puerto Rican. The nucleotide positions represents those in table 3-1. Dots (.) indicate identity with the reference sequence "ANDE" (Anderson et al., 1981) shown at the top of each page. Unidentified data indicated by Question mark (?), and deletions indicated by a dash sign (-).

**Table 3. Base substitution and length mutations.**

[illegible]



**Table 3 (cont'd).**

|      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|      | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| ID   | 8 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| ANDE | C | C | A | A | C | T | C | C | A | C | C | T | A | C | A | T | A | A | G | C | T | C | C | C | C | C | C | A |
| 366  | . | . | . | . | . | C |   | . | . | . | . | C | . | . | . | . | . | . | . | . | C | T | . | . | . | . | . | . |
| 398  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| 416  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | C | . | . | . | T | . | . | . | . | . | . | . | . |
| 85   | . | . | . | . | . | . |   | T | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . | . | . |
| 424  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 425  | . | . | . | . | T | . |   | . | . | . | . | . | . | . | G | C | . | . | . | . | . | . | . | . | . | . | . | . |
| 418  | . | . | . | . | T | . |   | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 404  | . | . | . | . | . | . |   | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 394  | . | T | . | . | T | . |   | . | . | G | . | . | . | . | T | . | C | . | . | . | . | . | . | . | . | . | . | . |
| 426  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . |
| 385  | . | T | . | . | T | . |   | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . |
| 417  | . | T | . | . | T | . |   | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . |
| 413  | . | T | . | . | T | . |   | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . |
| 401  | . | T | . | . | T | . | C | . | . | G | . | . | . | . | . | . | C | C | . | . | . | . | . | . | . | . | . | G |
| 415  | . | . | . | . | . | C |   | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . |
| 368  | . | . | . | . | . | . |   | . | T | . | . | T | . | . | . | . | C | . | . | . | T | . | . | . | . | . | . | . |
| 391  | . | . | . | . | . | . |   | . | T | . | . | T | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . |
| 107  | ? | . | . | G | . | . |   | T | T | . | . | . | . | . | . | . | . | . | ? | A | . | . | . | . | . | . | . | . |
| 369  | . | . | . | . | . | . |   | T | T | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . |
| 387  | . | . | . | . | . | . |   | T | T | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . |
| 163  | . | . | . | . | . | . |   | T | T | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . | . |
| 410  | . | . | . | . | . | . |   | T | T | . | . | . | . | . | . | . | . | . | ? | A | . | . | . | . | . | . | . | . |
| 377  | . | . | . | . | . | . |   | T | T | . | . | . | . | ? | . | . | . | . | . | A | . | . | . | . | . | . | . | . |
| 363  | . | . | . | . | . | . |   | T | T | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | T | . |
| 397  | . | . | . | . | . | . |   | T | T | . | . | . | . | . | . | . | . | . | . | A | T | . | . | . | . | . | . | . |
| 376  | . | . | . | . | . | . |   | T | . | . | . | . | . | . | . | . | . | . | ? | A | . | . | . | . | . | . | . | . |
| 408  | . | . | . | . | T | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 409  | . | . | . | . | T | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 393  | . | T | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . |
| 382  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 362  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . |
| 371  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T | T | . | . | . | . | . |
| 390  | . | . | . | . | . | . |   | . | . | . | . | . | C | . | . | . | . | . | . | . | . | C | T | T | . | . | . | . |
| 399  | . | . | . | . | . | . |   | . | . | . | . | . | C | C | . | . | . | . | . | . | . | C | C | T | . | . | . | . |
| 92   | . | . | . | . | . | . |   | . | . | . | . | . | C | . | . | . | . | . | . | . | . | C | C | T | . | . | . | . |
| 389  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | C | T | . | . | . | . |
| 420  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 392  | . | . | . | . | . | . |   | . | . | . | . | . | . | G | . | . | . | . | . | G | A | . | . | . | . | . | . | G |
| 400  | . | . | . | . | T | . |   | . | . | G | T | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . |
| 364  | . | . | . | . | . | . |   | . | . | . | . | . | ? | . | . | . | . | . | . | . | . | C | T | . | . | . | . | . |
| 87   | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | T | T | . | . | . | . |
| 83   | . | . | . | . | . | . |   | . | . | . | . | . | C | . | . | . | . | . | . | . | . | C | T | T | . | . | . | . |
| 406  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 414  | . | . | . | . | T | . |   | . | . | . | . | . | . | . | . | . | C | . | . | . | . | . | . | . | . | . | . | . |
| 110  | ? | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 113  | A | - | ? | . | . | . |   | . | . | . | . | . | . | . | - | . | . | . | ? | . | . | . | . | . | . | . | . | . |
| 421  | . | . | . | . | T | . |   | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 422  | . | . | . | . | T | . |   | . | . | . | T | . | . | . | . | . | G | . | . | . | . | . | . | . | . | . | . | . |
| 419  | . | . | . | . | . | . |   | . | . | . | . | . | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . |
| 360  | . | . | . | . | . | . |   | . | . | . | . | . | ? | . | . | . | . | . | . | . | . | C | T | . | . | . | . | . |

Table 3 (cont'd).

|      |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|      | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| ID   | 7 | 8 | 8 | 8 | 9 | 9 | 0 | 1 | 1 | 2 | 2 | 3 | 4 | 5 | 5 | 5 | 5 |
| ANDE | C | T | T | C | C | T | G | C | A | G | C | A | G | C | C | C | G |
| 366  | . | . | . | . | . | . | ? | . | . | . | ? | . | . | . | . | . | . |
| 398  | ? | . | . | . | . | . | . | . | . | . | ? | . | . | . | . | . | . |
| 416  | ? | . | . | . | ? | C | ? | . | . | . | ? | . | . | ? | . | . | . |
| 85   | ? | . | . | . | ? | C | . | . | . | . | ? | . | . | . | . | . | . |
| 424  | ? | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . |
| 425  | ? | . | . | . | ? | C | . | . | . | A | . | . | . | ? | . | . | . |
| 418  | ? | . | . | . | . | . | . | . | . | A | . | . | . | ? | . | . | . |
| 404  | ? | . | . | . | . | C | ? | . | . | . | ? | . | . | ? | . | . | . |
| 394  | T | . | . | . | . | . | . | . | . | . | . | . | . | T | . | T | A |
| 426  | ? | . | . | . | . | C | . | . | . | A | . | G | . | . | . | . | . |
| 385  | ? | . | . | . | ? | . | ? | . | . | . | ? | . | ? | ? | . | . | . |
| 417  | ? | . | . | . | ? | . | ? | . | ? | . | ? | . | ? | T | . | . | . |
| 413  | ? | . | . | . | . | . | . | . | . | ? | . | . | . | . | . | . | . |
| 401  | ? | . | . | ? | ? | . | ? | . | . | . | ? | . | . | ? | . | . | . |
| 415  | ? | . | . | . | . | . | . | . | . | . | ? | G | . | . | . | . | . |
| 368  | ? | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 391  | ? | . | . | . | ? | . | ? | . | . | . | ? | . | ? | ? | . | . | . |
| 107  | ? | . | . | . | ? | C | ? | . | . | . | ? | . | . | ? | . | . | . |
| 369  | T | . | . | . | ? | C | . | . | . | . | . | . | . | . | . | . | . |
| 387  | . | . | . | . | . | C | . | . | . | . | . | . | A | . | . | . | . |
| 163  | ? | . | . | . | ? | C | . | A | . | . | ? | . | . | . | . | . | . |
| 410  | ? | . | . | . | ? | C | . | . | . | . | ? | . | . | ? | . | . | . |
| 377  | ? | . | . | ? | ? | ? | ? | . | . | . | ? | . | . | ? | - | . | . |
| 363  | ? | . | C | . | ? | C | . | . | . | . | . | . | . | . | . | . | . |
| 397  | ? | . | . | T | ? | C | ? | . | . | . | ? | . | . | ? | . | . | . |
| 376  | ? | . | . | ? | ? | C | ? | . | . | . | ? | . | . | ? | . | . | . |
| 408  | ? | . | . | . | . | . | ? | . | . | . | . | . | . | . | . | . | . |
| 409  | ? | . | . | . | . | . | . | . | . | . | ? | . | . | . | . | . | . |
| 393  | ? | . | . | . | ? | . | . | . | . | . | ? | . | . | ? | . | . | . |
| 382  | ? | . | . | . | . | . | . | . | . | . | ? | . | . | . | . | . | . |
| 362  | ? | . | . | . | . | . | . | . | . | . | ? | . | . | . | . | . | . |
| 371  | ? | . | . | . | . | . | . | . | . | . | ? | . | . | . | . | . | . |
| 390  | ? | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 399  | ? | . | . | . | ? | . | ? | . | . | . | ? | . | . | . | . | . | . |
| 92   | ? | . | . | . | . | . | . | . | . | . | ? | . | . | . | . | . | . |
| 389  | ? | . | . | . | ? | . | ? | . | . | . | . | . | . | ? | . | . | . |
| 420  | ? | . | . | . | ? | . | . | . | . | . | ? | . | . | . | . | . | . |
| 392  | . | . | . | . | . | . | . | . | . | . | ? | . | . | . | . | . | . |
| 400  | ? | . | . | . | ? | . | ? | . | . | . | ? | . | . | ? | . | . | . |
| 364  | ? | . | . | . | ? | . | ? | . | . | . | ? | . | . | ? | . | . | . |
| 87   | ? | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 83   | ? | - | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 406  | ? | . | . | . | . | . | ? | . | . | . | ? | . | . | . | . | . | . |
| 414  | ? | . | . | . | ? | C | ? | . | . | . | ? | . | ? | ? | . | . | . |
| 110  | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| 113  | ? | . | . | . | ? | . | ? | . | . | . | ? | . | . | ? | . | . | . |
| 421  | ? | . | . | . | T | . | . | . | . | . | ? | . | . | ? | . | . | . |
| 422  | . | . | . | . | . | . | . | . | . | A | . | . | . | . | . | . | . |
| 419  | ? | . | . | . | . | . | . | . | . | . | ? | . | . | G | . | . | . |
| 360  | ? | . | . | . | . | . | ? | ? | . | . | ? | . | . | . | . | . | . |

**Table 4. Nucleotide positions in the reference sequence.  
Nucleotide positions corresponding to the numbering system  
of the standard reference sequence (Anderson et al., 1981).**

---

|           |                 |
|-----------|-----------------|
| 5=15974   | 279=16248       |
| 36=16005  | 281=16250       |
| 43=16012  | 285=16254       |
| 53=16022  | 286=16255       |
| 60=16029  | 287=16256       |
| 61=16030  | 300=16269       |
| 113=16082 | 305=16274       |
| 116=16085 | 307=16276       |
| 123=16092 | 308=16277       |
| 125=16094 | 318=16287       |
| 126=16095 | 319="insertion" |
| 127=16096 | 321=16290       |
| 128=16097 | 323=16292       |
| 131=16100 | 324=16293       |
| 136=16105 | 325=16294       |
| 141=16110 | 326=16295       |
| 142=16111 | 329=16298       |
| 153=16122 | 331=16300       |
| 154=16123 | 332=16301       |
| 156=16125 | 340=15309       |
| 159=16128 | 341=16310       |
| 165=16134 | 342=16311       |
| 166=16135 | 343=16312       |
| 168=16137 | 347=16316       |
| 175=16144 | 350=16319       |
| 185=16154 | 351=16320       |
| 194=16163 | 356=16325       |
| 202=16171 | 358=16327       |
| 206=16175 | 368=16337       |
| 209=16178 | 373=16342       |
| 216=16185 | 379=16348       |
| 218=16187 | 383=16351       |
| 222=16191 | 388=16357       |
| 226=16195 | 389=16358       |
| 228=16197 | 391=16360       |
| 229=16198 | 393=16363       |
| 231=16200 | 404=16373       |
| 235=16204 | 413=16382       |
| 236=16205 | 415=16384       |
| 239=16208 | 421=16390       |
| 242=16211 | 424=16393       |
| 249=16218 | 430=16399       |
| 250=16219 | 445=16414       |
| 253=16222 | 450=16419       |
| 255=16224 | 456=16425       |
| 269=16238 | 457=16426       |
| 270=16239 | 459=16428       |

---

**Figure 7. Distribution of mutations.**  
The histogram shows the total number of mutations within blocks of 20 bases. Three hypervariable domains (I, II, and III) are apparent in the region (hypervariable segment 1).

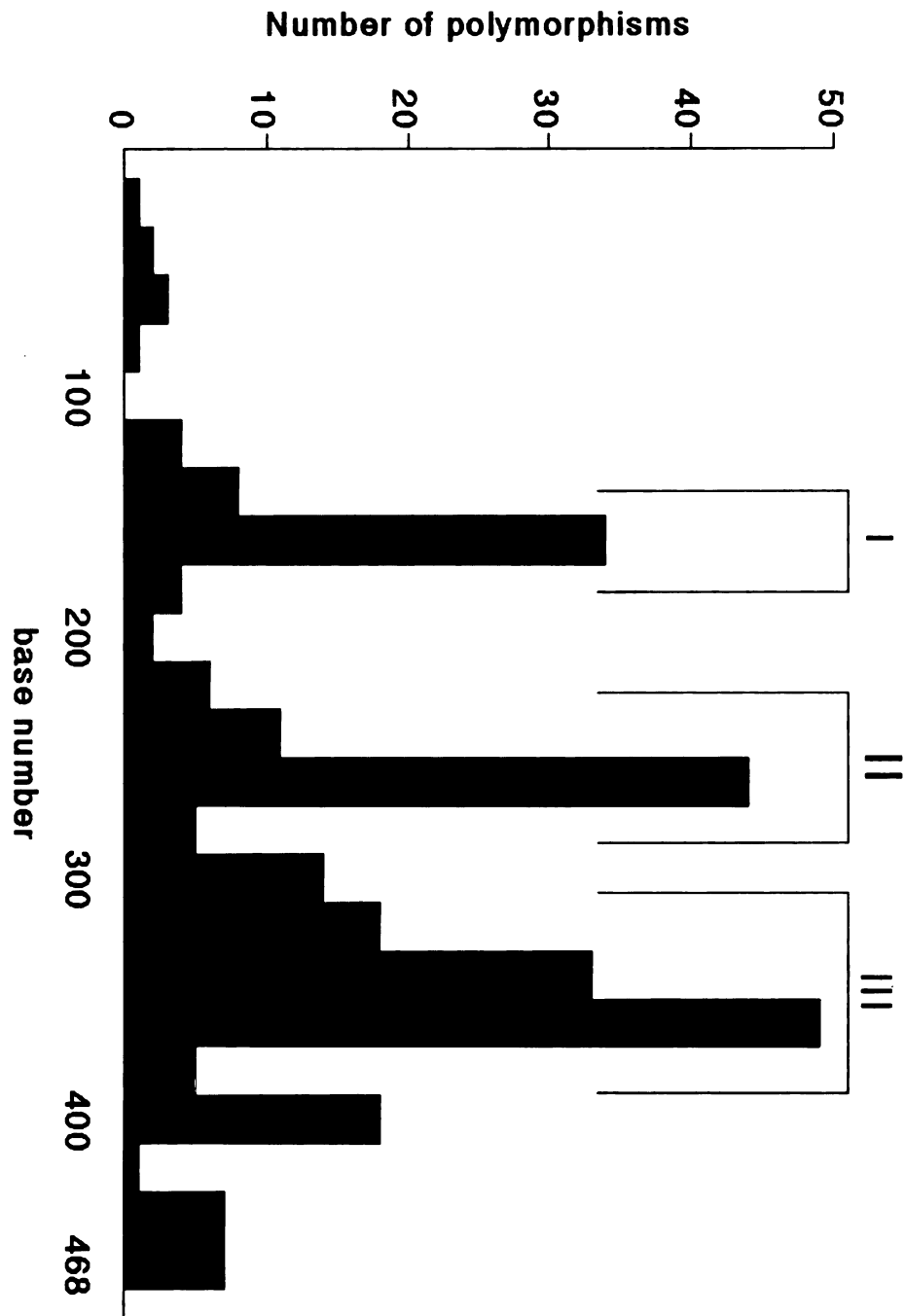


Figure 7

nonoverlapping blocks of 20 bases. Figure 7 shows that the most highly variable sequence lies in three main domains (I, II, and III). The first domain (I) lies between nucleotide 140 and 160, the second domain (II) lies between nucleotide 240 and 260, and the third domain (III) lies between nucleotide 340 and 380. These three domains are within the hypervariable region 1, and this type of polymorphism has been documented before (Walberg and Clayton, 1981; Greenberg et al., 1983; Brown et al., 1986; and Vigilant et al., 1989).

#### **Substitutions versus length mutations**

The majority of polymorphism observed consisted of single base substitutions rather than insertions or deletions of bases. The ratio of length mutation to substitution events was 1:8 as shown in Table 5. Of the 92 polymorphic sites 11 were length mutations.

#### **Transitions versus transversions**

The sequence information in Table 3 was used to calculate the number of transitions (purine to purine or pyrimidine to pyrimidine) and transversions (purine to pyrimidine or pyrimidine to purine) type substitutions. Of the 266 point mutations shown in Table 5, 259 were transitions, and 9 was transversion. Therefore the ratio of transitions:transversions is 28.8:1. A bias towards transition type events is the norm in mammalian mtDNA

**Table 5. Analysis of Nucleotide Substitutions, Deletions, and Insertions.**

| Type of Mutation        | No. Observed |
|-------------------------|--------------|
| Deletion:               |              |
| A → -                   | 2            |
| C → -                   | 5            |
| G → -                   | 2            |
| T → -                   | 2            |
|                         | --           |
| Total:                  | 11           |
| Insertion:              |              |
| - → C                   | 1            |
|                         | --           |
| Total length mutations: | 12           |
| Substitution:           |              |
| Transition:             |              |
| T → C                   | 69           |
| C → T                   | 134          |
| A → G                   | 24           |
| G → A                   | 32           |
|                         | ----         |
| Total:                  | 259          |
| Transversion:           |              |
| A → C                   | 1            |
| G → T                   | 1            |
| C → A                   | 2            |
| C → G                   | 3            |
| T → A                   | 2            |
|                         | --           |
| Total:                  | 9            |
|                         | ----         |
| Total substitutions:    | 266          |
| -----                   |              |
| Total mutations:        | 278          |
| -----                   |              |

(Brown et al., 1982).

### **Sequence diversity**

The most highly variable 20 positions were picked from Table 3 in order to establish a phylogenetic relationship between the different mt-lineages in this study group (Ward et al., 1991). Table 6 shows that by using these highly variable positions it is possible to define 33 Puerto Ricans mt-lineages, the most frequent lineage occurred in 6 individuals, and 26 of the 33 lineages occurred just once, giving an estimated diversity value,  $h$  (Nei, 1987) of 96.1%. Twelve of the 20 variable nucleotide positions in Table 6 show a similar polymorphism to Nuu-Chah-Nulth mt-lineages (Ward et al., 1991). Comparison of the 33 mt-lineages identified the 4 main clusters (A, B, C, and D) presented in Table 6.

### **Phylogenetic analysis**

The 33 lineages data presented in Table 6 were used to determine the relationship between these different mt-lineages. The pileup program of GCG was used to construct the dendrogram presented in Figure 8. Careful inspection of the dendrogram in Figure 8 suggested that the majority of lineages fall into four clusters (A, B, C, and D) as predicted from Table 6.

Table 6. Mitochondrial-lineages.

Definition of 33 Puerto Rican mitochondrial lineages from 50 samples in terms of the 20 variable positions within the control region. Position 141 corresponds to position 16110 in the reference sequence "ANDE" (Anderson et al., 1981). Dots (.) indicate identity with the reference sequence. The number of individuals carrying a specific sequence is indicated in the right-hand column.

|      | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | No. of<br>indiv-<br>iduals |   |
|------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----------------------------|---|
| ID   | 4 | 5 | 5 | 5 | 3 | 4 | 4 | 5 | 0 | 0 | 2 | 2 | 2 | 4 | 5 | 5 | 5 | 5 | 9 | 2                          |   |
| ANDE | C | T | T | G | T | G | A | C | C | C | C | C | T | T | G | C | T | C | T | G                          |   |
| 28)  | . | . | . | A | . | . | . | T | . | . | T | . | . | . | . | . | . | C | . | 1                          | B |
| 29)  | . | . | . | A | . | . | . | T | . | . | T | . | . | . | A | . | . | C | . | 1                          |   |
| 33)  | . | . | . | . | . | . | . | T | . | . | T | . | . | . | A | . | . | C | . | 1                          |   |
| 30)  | T | . | . | . | . | . | . | T | . | . | T | . | . | . | A | . | . | C | . | 6                          |   |
| 31)  | T | . | C | . | . | . | . | T | . | . | T | . | . | . | A | . | . | C | . | 1                          |   |
| 32)  | T | . | . | . | . | . | . | T | . | . | T | . | . | . | A | T | . | C | . | 1                          |   |
| 23)  | T | . | . | . | . | . | . | T | . | . | . | . | . | . | C | T | . | C | . | 1                          | A |
| 26)  | . | . | . | A | . | A | . | T | . | T | . | . | . | . | C | . | . | C | A | 1                          |   |
| 27)  | . | . | . | A | . | A | . | T | . | . | . | . | . | . | . | . | . | A | . | 1                          |   |
| 9)   | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | C | A | 1                          |   |
| 8)   | . | . | . | A | C | . | . | T | . | . | . | . | . | . | C | . | . | . | . | 2                          |   |
| 16)  | . | . | . | . | . | . | . | T | . | . | . | . | . | . | C | . | C | T | . | 4                          | D |
| 17)  | . | C | . | . | . | . | . | T | . | . | . | . | . | . | C | . | C | T | . | 1                          |   |
| 13)  | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | T | . | 4                          |   |
| 15)  | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | C | T | . | 1                          |   |
| 18)  | . | . | . | . | C | . | . | T | . | . | . | . | . | . | . | . | C | T | . | 1                          |   |
| 14)  | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C | T | . | 1                          |   |
| 11)  | . | . | . | . | . | . | . | T | . | . | . | . | . | . | . | T | . | . | . | 1                          | C |
| 12)  | . | . | . | . | C | . | . | T | . | . | . | . | . | . | . | T | . | . | . | 1                          |   |
| 10)  | . | . | . | . | . | . | . | . | T | . | . | . | . | . | . | T | . | . | . | 1                          |   |
| 1)   | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 1                          |   |
| 2)   | . | . | C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 1                          |   |
| 3)   | T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | 1                          |   |
| 7)   | T | . | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | 1                          | C |
| 5)   | . | C | . | . | . | . | . | . | . | . | . | . | . | . | A | . | . | . | . | 1                          |   |
| 6)   | . | C | . | . | . | . | . | T | . | . | . | . | . | . | . | . | . | . | . | 1                          |   |
| 4)   | . | . | . | . | . | G | . | . | T | . | . | . | . | . | . | . | . | . | . | 2                          |   |
| 24)  | . | . | C | . | . | . | . | T | T | T | . | . | . | . | C | . | . | . | . | 4                          |   |
| 5)   | . | C | C | . | . | . | . | T | T | T | . | . | . | . | . | . | . | . | . | 1                          |   |
| 19)  | . | . | . | . | . | . | . | T | . | T | . | T | . | . | . | . | . | . | A | 2                          | C |
| 20)  | . | . | . | . | . | . | . | T | . | T | . | T | . | C | . | . | . | . | . | 1                          |   |
| 21)  | . | . | . | A | . | . | . | . | T | . | T | . | C | . | . | . | . | . | . | 1                          |   |
| 22)  | . | C | . | . | . | . | . | . | T | . | . | . | C | . | . | . | . | C | . | 1                          |   |

**Figure 8. Phylogenetic tree.**  
**The dendogram represent the relationship between the 33 mt-**  
**lineages in the Puerto Ricans. Shown in the dendogram the**  
**four main clusters (A-D), which count for 96% of the group**  
**studied.**

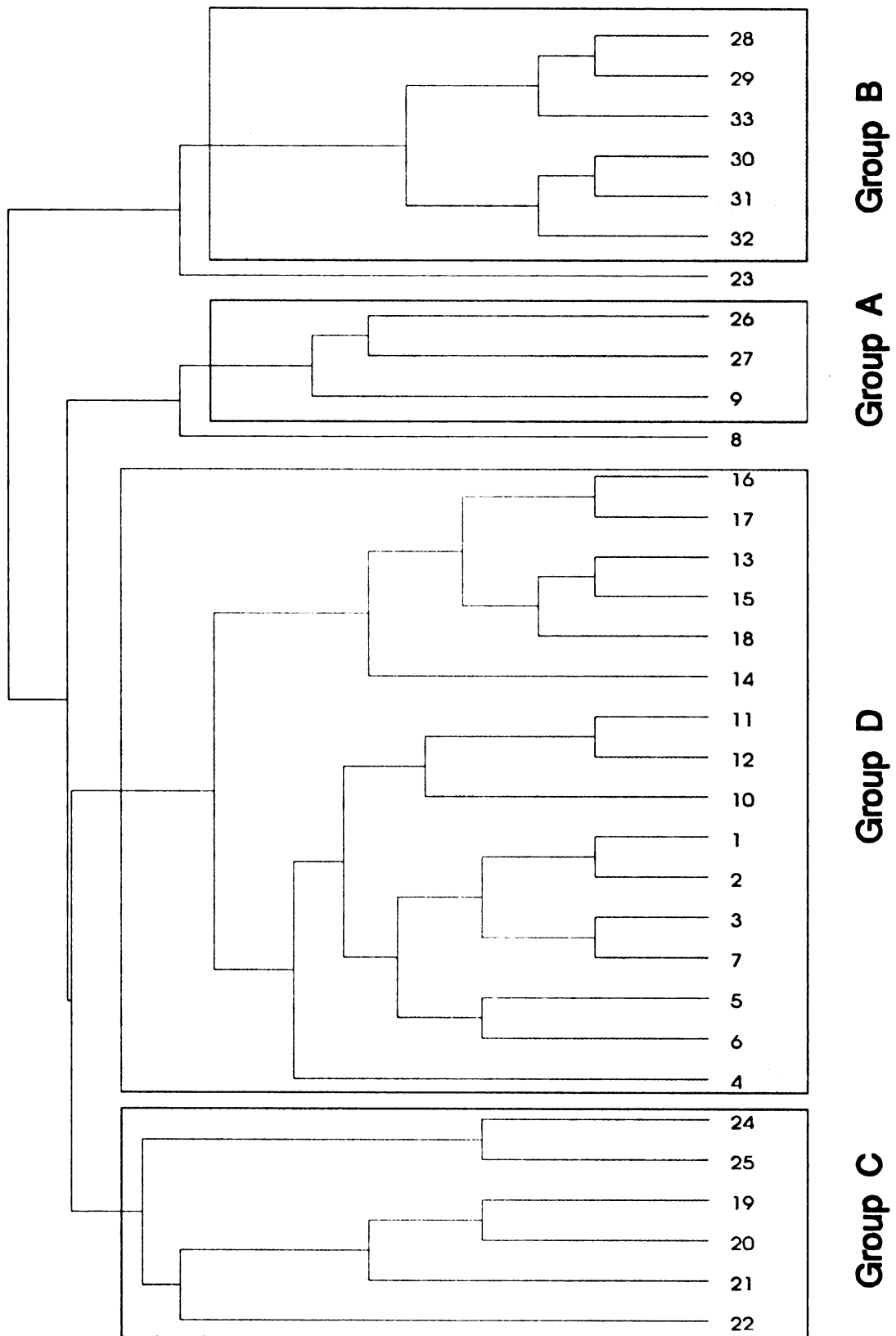


Figure 8

## Discussion

Using high molecular weight DNA and mitochondrial specific oligonucleotide primers (Kocher et al., 1989; Rienzo and Wilson, 1990; Kocher and Wilson, 1991; Stoneking et al., 1991; Vigilant et al., 1991; Ward et al., 1991), the PCR was successfully used to amplify and sequence the hypervariable segment 1 of the mitochondrial control region of 50 Puerto Ricans. The size of the amplified fragment (500 bp), and the sequencing data generated were in agreement with reference sequence (Anderson et al., 1981).

The hypervariable segment 1 was chosen in this study, because it is the most variable segment in the mt-genome (Kocher and Wilson, 1991; Stoneking et al., 1991; Vigilant et al., 1991). Figure 7 shows the three variable domains (I, II, and III) identified within the hypervariable segment 1. The domains in this study group are the same hypervariable sequences identified previously in different population studies (Vigilant et al., 1989; Ward et al 1991). It also shows that the sequence between 180-200; which associated with the only functional element(TAS) in that region (Walber et al., 1981), is the least hypervariable.

### Comparison of all sequences

The sequence were aligned as shown in Table 2 with the human reference sequence (ANDE) (Anderson et al., 1981). In the alignment shown, there are 468 sites, of which 372 are

identical in all 50 sequences. Eleven small gaps are indicated, ten being attributed to the loss of a single base and one to the gain of one base. In addition there are 84 sites of base substitution. There is an eight fold excess of substitutions to length mutation events in the hypervariable segment 1. This is consistent with patterns observed from previous studies of control region sequences (Greenberg et al., 1983) and restriction surveys of the entire mtDNA genome (Stoneking et al., 1986a). All of the eleven positions experiencing length mutations include a repeat of a single nucleotide or lie next to a dinucleotide repeat unit. This is in accordance with theories of frameshift mutagenesis, which predict that the occurrence of frameshift mutations at a particular site depends upon the base sequence at the site (Streisinger et al., 1966). The model to explain the length mutations postulates that the frameshifts arise as a result of a mispairing event in a region of repeated bases or base doublets (Streisinger et al., 1966; Albertini et al., 1982).

The substitutions observed obey the expected bias towards transition rather than transversion type events. By considering all the mutated sites in the first hypervariable segment a transition:transversion ratio of 28.8 : 1 was obtained. The 28.8 : 1 ratio obtained in this study is in agreement with the results obtained for the human mtDNA control region documented elsewhere (Brown et al., 1982; Aquadro and Greenberg, 1983; Vigilant et al., 1989).

### **Sequence diversity and phylogenetic analysis**

As shown in table 6 and figure 8, thirty three mt-lineages were identified in the Puerto Rican population. These 33 mt-lineages were found to be clustered in four main groups (A, B, C, and D). Sixteen of the 33 mt-lineages were clustered in one group designated as group D which counts for 46% (23 of the 50 samples) of the study group. The second largest group was group B which has 6 of the 33 mt-lineages. This group counts for 22% (11 of the 50 samples) of the study group. The third group found was group C which also included 6 of the mt-lineages. Even though this group has only 6 mt-lineages it accounted for 20% (10 of the 50 samples) of the study group. The last and the smallest group was group A which has 3 of the 33 mt-lineages and accounts for 6% (3 of the 50 samples) of the study group. The remaining 6% (3 of the 50 samples) were found to define 2 of the 33 mt-lineages.

Comparison of the four main groups identified in this study with the four main groups (I, II, III, IV) identified in the Nuu-Chah-Nulth (an Amerindian of the Pacific Northwest) mt-lineages (Ward et al., 1991) revealed very high similarity. Groups D and B which accounts for 68% (34 of the 50 samples) show 80-90% homology with groups II and III of the Nuu-Chah-Nulth respectively. The other two groups (A and C) which has 9 of the 33 mt-lineages accounting for 26% (13 of the 50 samples) of the study group, were found to have sequence similarity to the Kung

(people of southern Africa) mt-lineages (Vigilant et al., 1989). Most of the similarities seen in these two groups were mainly within group C which accounts for 20% of the study group as mentioned earlier.

Puerto Rican history finds that the first human inhabitant was an Amerindian. The Amerindians came from North America between 20,000 and 5,000 years ago. After that and between the years 1493 - 1510 the European (Spanish) and African people arrived to Puerto Rico in their voyage to the West Indies's.

Therefore, taking into consideration the maternal mode of inheritance of mtDNA and the role it plays in population genetic studies, these results can be explained. It's obvious that the mother of most Puerto Ricans was an Amerindian. Sixty eight percent of the mt-lineages were similar to Amerindian mt-lineages. The other 26% of African origin may arise from the African people who arrived to Puerto Rico at around 1510. Based on blood group frequencies Hanies et al., 1991 reported that the Puerto Ricans are a mixed population of Amerindian, African, and Spanish.

The remaining 6% (3 of the 50 samples) represent the white people of Puerto Rico. It's obvious that this is a very small percentage of the population and this can be explained by the fact that when the Europeans first arrived to Puerto Rico they didn't have any women with them. Most probably the 68% Amerindian mt-lineages are mixed lineages

due to the maternal mode of inheritance of the mtDNA from indigenous Amerindian women.

The main outcome of this study is the definition of the 33 mt-lineages in the Puerto Ricans. Sixty eight percent of which is of an Amerindian origin, 26% of African origin, and 6% of European origin. These data can by itself explain the different pattern of IDDM susceptibility at the HLA-DQ locus between the Caucasians and the Puerto Ricans. It is obvious that the Puerto Ricans have a different ethnic origin than the Caucasians.

**Future prospective**

It will be of interest to study the HLA-DQ $\beta$ 57 locus in this group of Puerto Ricans, in order to define the similarities between this group and the published data about the HLA-DQ $\beta$ 57. It will be of special interest to study the polymorphism at the HLA-DQ $\beta$ 57 locus in different Amerindian groups and to compare those to the Puerto Rican population. This kind of study will emphasize the ethnic origin of the Puerto Ricans and the relationship between them and Amerindians.

# **APPENDIX**

## List of references

- Albertini, A. M., M. hofer, M. P. Calos and J. H. Miller (1982). On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions. *Cell* 29: 319-328.
- Anderson. S., A. T. Bankier, B. G. de Bruijn, M. H. L. de Bruijn, A. R. Roe, F. Sanger, P. H. Screier, A. J. H. Smith, R. Staden and I. G. Young (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290 : 457-465.
- Anderson, S., M. H. L. de Bruijn, A. R. Coulson, I. C. Eporn, and I. G. Young (1982). Complete sequence of bovine mitochondrial DNA: Conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* 156: 683-717.
- Aparicro, J. M., A. Wakisaka and A. Takada (1988). HLA-DQ system and insulin dependent diabetes mellitus in Japanese: does it contribute to the development of IDDM as it does in Caucasians? *Immunogenetics* 28: 240-246.
- Aquadro, C.F. and B. D. Greenberg (1983). Human mitochondrial DNA variation and evolution : Analysis of nucleotide sequences from seven individuals. *Genetics* 103: 287-312.
- Attradi, G. (1985). Animal mitochondrial DNA: An extreme example of genetic economy. In *International Review of Cytology*, Vol. 93: *Genome Evolution in Prokaryotes and Eukaryotes*, eds. D. C. Reanny and P. Chambon, Academic Press, New York, pp. 93-145.
- Atkinson, T. and M. Zoller (1984). In Oligonucleotide Synthesis: A practical approach, M. J. Gait, Editor, IRL, Press, Oxford, England. pp. 35-81.
- Avis, J. C., B. W. Bowen and T. Lamb (1989). DNA fingerprints from hypervariable mitochondrial genotypes. *Mol. Biol. Evol.* 6: 258-269.
- Avis, J. C., R. M. Ball and J. Arnold (1988). Current versus historical population sizes in vertebrate species with high gene flow: A comparison based on mtDNA lineages and inbreeding theory for neutral mutation. *Biol. Evol.* 5(4): 331-344.
- Bogenhagen, D. and Clayton, D. A. (1974). The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human hela cells. *The Journal of Biological Chemistry* 249: 7991-7995.

Brown, W. M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. USA*. 77: 3605-3609.

Brown, W. M., 65E. M. Prager, A. Wang and A. C. Wilson (1982). Mitochondrial DNA sequence65 of primates: Tempo 65and mode of evolution. *J. Mol. Evol.* 18: 225-239.

Brown, G. G., G. Gadaleta, G. Pepe, C. Saccone and E. Sbisà (1986). Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. *J. Mol. Biol.* 192: 503-511.

Brown, W. M., J. Shine, and H. M. Goodman (1978). Human mitochondrial DNA : Analysis of 7S DNA from the origin of replication. *Proc. Natl. Acad. Sci. USA* 75: 735-739.

Brown, W. M., M. George and A. C. Wilson (1979). Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA*. 76: 1967-1971.

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

Cann, R. L., W. M. Brown and A. C. Wilson (1984). Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* 106: 479-499.

Case, J. T. and D. C. Wallace (1981). Maternal inheritance of mitochondrial DNA polymorphisms in cultured human fibroblasts. *Som. Cell. Genet.* 7: 103-108.

65

Chang, D. D., and D. A. Clayton (1984). Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. *Cell* 36: 635-643.

Chang, D. D. and D. A. Clayton (1987a). A mammalian mitochondrial RNA processing activity contains nucleus-encoded RNA. *Science* 235: 1178-1184.

Chang, D. D. and D. A. Clayton (1987b). A novel endoribonuclease cleaves at a priming site of mouse mitochondrial DNA replication. *EMBO J.* 6:409-417.

Chang, D. D., J. H. Hixson and D. A. Clayton (1986). Minor transcription initiation events indicate that both human mitochondrial promoters function bidirectionally. *Mol. Cell. Biol.* 6: 294-301.

Chomyn, A., P. Mariottini, M. Cleeter, F. Ragan, A. Matsuno-Yagi, Y. Hatefi, R. Doolittle and G. Attardi (1985). Six unidentified reading frames of human mitochondrial DNA encode components of the respiratory-chain NADH dehydrogenase. *Nature* 314: 592-597.

Chomyn, A., W. A. Cleeter, C. I. Ragan, M. Riley, F.R. Doolittle and G. Attardi (1986). URF6, last unidentified reading frame of human mtDNA, codes for an NADH dehydrogenase subunit. *Science* 234: 614-618.

Clark, A. (1990). Mitochondrial genome: defects, disease, and evolution. *J. Med. Genet* 27: 451-456.

Clayton, D. A. (1982). Replication of animal mitochondrial DNA. *Cell* 28: 693-705.

Clayton, D. A. (1984). Transcription of mammalian mitochondrial genome. *Ann. Rev. Biochem.* 53: 573-594.

Clayton, D. A. (1991). Replication and transcription of vertebrate mitochondrial DNA. *Annu. Rev. Cell Biol.* 7: 453-478.

Cote, C., D. Boulet, and J. Poirier (1990). Expression of the mammalian mitochondrial genome. *J. Biol. Chem.* 265: 7532-7538.

Dorman, J., R. Laporte, R. Stone and M. Trucco (1991). World wide differences in the incidence of type 1 diabetes are associated with amino acid variation at position 57 of the HLA-DQ beta chain. *Proc. Natl. Acad. Sci. USA.* 87: 7330-7339.

Ernster, l.; Schatz, G. (1981). Mitochondria: a historical review. *J. Cell Biol.* 91: 2275-2285.

Feng, D. F. and R. F. Doolittle (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Molec. Evol.* 25: 351-360.

Ferris, S. D., A. C. Wilson and W. M. Brown (1981). Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA.* 78: 2432-2436.

Fisher, R. P., J. N. Topper and D. A. Clayton (1987). Promoter selection in human mitochondrial involves binding of a transcription factor to orientation-independent upstream regulatory elements. *Cell* 50: 247-258.

- Foran, D. R., J. E. Hixson and W. M. Brown (1988). Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. *Nuc. Acids Res.* 16: 5841-5861.
- Giles, R. E., H. Blance, H. M. Cann, and D. C. Wallace (1980). Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. USA* 77 : 6715-6719.
- Giles, R. E., I. Stroynowski and D. C. Wallace (1980). Characterization of mitochondrial DNA in chloramphenicol-resistant interspecific hybrids and a cybrid. *Soma. Cell. Genet.* 6:543-554.
- Greenberg, B.D., J. E. Newbold, and A. Sugino (1983). Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* 21: 33-49.
- Hall, H. G. and D. R. Smith (1991). Distinguishing African and European honeybee matrilineages using amplified mitochondrial DNA. *Proc. Natl. Acad. Sci. USA.* 88: 4548-4552.
- Hanis, C. L., D. Hewett-Emmett, T. K. Bertin and W. J. Schull (1991). Origins of US Hispanic: Implications for diabetes. *Diabetes Care* 14: 618-627.
- Hedges, S. B., S. Kumar, K. Tamura and M. Stoneking (1991). Human origins and analysis of mitochondrial DNA sequences. *Science* 255: 737-739.
- Hixson, J. E. and D. A. Clayton (1985). Initiation of transcription from each of the two human mitochondrial promoters requires unique nucleotides at the transcriptional start sites. *Proc. Natl. Acad. Sci. USA.* 82: 2660-2664.
- Horai, S. and E. Matsunaga (1986). Mitochondrial DNA polymorphism in Japanese. *Hum. Genet.* 72: 105-117.
- Horai, S. and K. Hayasaka (1990). Intraspecific nucleotide sequence differences in the major noncoding region of human mitochondrial DNA. *Am. J. Hum. Genet.* 46: 828-842.
- Horn, G. T., T. L. Bugawan and C. M. Long (1988). Allelic sequence variation of the HLA-DQ loci: relation to serology and to insulin dependent diabetes susceptibility. *Proc. Natl. Acad. Sci. USA.* 85:6012-6016.
- Innis, M. A., K. B. Mayambo, D. H. Gelfand and M. A. Brow (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci. USA.* 85: 9436-9440.

Johnson, M. J., D. C. Wallace, S. D. Ferris, M. C., Rattazzi and L. L. Cavalli-Sforza (1983). Radiation of human DNA types analyzed by restriction endonuclease cleavage patterns. *J. Mol. Evol.* 19: 255-271.

King, T. C. and R. L. Low (1987). Mapping of control elements in the displacement loop region of bovine mitochondrial DNA. *J. Biol. Chem.* 262: 6204-6213.

Kocher, T. D. and A. C. Wilson (1991). Sequence evolution of mitochondrial DNA in human and chimpanzees: Control region and a protein-coding region. In Evolution of life - fossils, molecules and culture, S. Osawa and T. Honjo (eds.), Springer-Verlag, Tokyo, pp. 391-413.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Paabo, F. X. Villblnce and A. C. Wilson (1989). Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA.* 86: 6196-6200.

Lee, G., F. N. Shamma, M. P. Diamond and J. T. D. Lee (1992). HLA-DQ $\beta$ 57 in Hispanic patients with insulin-dependent diabetes mellitus. *Am. J. Obstet. Gynecol.* 167: 1565-1570.

Low, R. L., J. M. Buzan and C. L. Couper (1988). The preference of the mitochondrial endonuclease for a conserved sequence block in mitochondrial DNA is highly conserved during mammalian evolution. *Nucleic Acids Res.* 16(14): 6427-6445.

Maniatis, T., E. F. Fritsch and R. Sambrook (1989). Molecular cloning. A laboratory Manual. Cold Spring Harbor laboratory, New York.

Montoya, J., G. L. Gaines, and G. Attardi (1983). The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. *Cell* (Cambridge, Mass.) 34:151-159.

Montoya, J., T. Christianson, D. Levens, M. Rabinowitz, Ndd G. Attardi (1982). Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 79: 7195-7199.

Morel, P. A., J. S. Dorman and J. A. Todd (1988). Aspartic acid at position 57 of the HLA-DQ $\beta$  chain protects against type 1 diabetes. *Proc. Natl. Acad. Sci. USA.* 85: 8111-8115.

Nei, M. (1987). Molecular Evolution Genetics (Columbia Univ. Press, New York), pp 178-179.

Palca, J. (1990). The other human genome. Science 249: 1104-1105.

Pearson, W. R. and D. J. Lipman (1988). Improved tools for biological sequence analysis. Proc. Natl. Acad. Sci. USA. 85: 2444-2448.

Penny, M. A., C. H. Mijovic, D. A. Cavan, K. H. Jacobs, D. Jenkins, J. Fletcher and A. H. Barnett (1993). An investigation of the association between HLA-DQ heterodimers and type 1 (insulin-dependent) diabetes mellitus in five racial groups. Hum. Immun. 38: 179-183.

Perl, L. Puerto Rico "Island between two worlds" (1979). William Morrow and company, New York.

Rienzo, A. D. and A. C. Wilson (1991). Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc. Natl. Acad. Sci. USA. 88: 1597-1901.

Rotter, J. I. (1983). An HLA genotype study of IDDM. Diabetes 32: 169-174.

Schurr, T. G., S. W. Ballinger, Y-Y. Gan, J. A. Hodge, D. A. Merriwether, D. N. Lawrence, W. C. Knowler, K. M. Weiss and D. C. Wallace (1990). Amerindian mitochondrial DNAs have rare Asian mutations at high frequencies, suggesting they derived from four primary maternal lineages. Am. J. Hum. Genet. 46: 613-623.

Smith, D. P., E. M. Johnstone, S. P. Little and H. M. Hsiung (1990). Direct DNA sequencing of cDNA inserts from plaques using the linear polymerase chain reaction. Biotechniques 9: 48-54.

Smith, T. F. and M. S. Waterman (1981). Comparison of bio-sequences. Advanced in Applied Mathematics 2: 482-489.

Southern, S. O., P. J. Southern and A. E. Dizon (1988). Molecular characterization of a cloned dolphin mitochondrial genome. J. Mol. Evol. 28: 32-42.

Spuhler, J. N. (1988). Evolution of mitochondrial DNA in monkeys, apes and humans. Yearbook of phys. Anthro. 31: 15-48.

Sterkers, G., P. Zeliszewski and A. M. chaussee (1988) HLA-DQ rather than HLA-DR region might be involved in dominant nonsusceptibility to diabetes. Proc. Natl. Acad. Sci. USA. 85: 6473-6477.

Stoneking, M., D. Hedgecock, R. G. Higuchi, L. Vigilant and H. A. Erlich (1991). Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet.* 48: 370-382.

Stoneking, M., K. Bahtia and A. C. Wilson (1986a). Mitochondrial DNA variation in eastern highlanders of Papua New Guinea. In Genetic Variation and its Maintenance, eds. D. F. Roberts and G. F. DeStefano, Cambridge University Press, Cambridge, pp. 87-100.

Stoneking, M., K. Bahtia and A. C. Wilson (1986b). Rate of sequence divergence estimated from restriction maps of mitochondrial DNAs from Papua New Guinea. In Cold Spring Harbor Symposia on Quantitative Biology, Volume L1, Cold Spring Harbor Laboratory, Cold Spring Harbor, pp. 433-439.

Stoneking, M., L. B. Jorde, K. Bhatia and A. C. Wilson (1990). Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124: 717-733.

Stoneking, M. and R. L. Cann (1989). African origin of human mitochondrial DNA. In The human Revolution, eds. P. Mellars and C. Stringer, Edinburgh University Press, Edinburgh, pp. 17-30.

Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi and M. Inouye (1966). Frameshift mutations and the genetic code. In Cold Spring Harbor symposia on quantitative biology, volume XXXI, Cold Spring Harbor laboratory of quantitative biology, Cold Spring Harbor, pp. 77-84.

Templeton, A. R. (1991). Human origins and analysis of mitochondrial DNA sequences. *Science* 255: 737.

Todd, J. A., J. I. Bell and H. O. McDevitt (1987). HLA-DQ $\beta$  gene contributes to susceptibility and resistance to insulin dependent diabetes mellitus. *Nature* 329: 599-604.

Vigilant, L., M. Stoneking and A. C. Wilson (1988). Conformational mutation in human mtDNA detected by direct sequencing of enzymatically amplified DNA. *Nucl. Acids Res.* 16: 5945-5955.

Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes and A. C. Wilson (1991). African population and the evolution of human mitochondrial DNA. *Science* 253: 1503-1507.

Vigilant, L., R. Pennington, H. Harpending, T. D. Kocher and A. C. Wilson (1989). Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Nat. Acad. Sci. USA.* 86: 9350-9354.

Walberg, M. W. and D. A. Clayton (1981). Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. *Nucl. Acids Res.* 9: 5411-5421.

Wallace, D. C. (1989). Report of the committee on human mitochondrial DNA. *Cytogenet. Cell. Genet.* 51: 612-621.

Wallace, D. C. (1992). Mitochondrial genetics: A paradigm for aging and degenerative diseases. *Science* 256: 628-632.

Ward, R. H., B. L. Frazier, K. Dew-Jager and S. Paabo (1991). Extensive mitochondrial diversity within a single Amerindian tribe. *Proc. Natl. Acad. Sci. USA.* 88: 8720-8724.

Wayne, R. K., A. Meyer, N. Lehman, B. V. Valkenburgh, P. W. Kat, T. K. Fuller, D. Girman and S. J. O'brien (1990). Large sequence divergence among mitochondrial DNA genotypes within populations of eastern African black-backed jackles. *Proc. Natl. Acad. Sci. USA.* 87: 1772-1776.

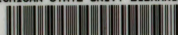
Wilson, A. C., M. Stoneking, R.L. Cann, E. M. Prager, S. D. Ferris, L. A. Wrishcnik and R. G. Higuchi (1987). Mitochondrial clans and the age of our common mother. In Human Genetics, eds. F. Voegt and K. Sperling, Springer-Verlag, Berlin, pp. 158-164.

Wilson, A. C., R. L. Cann, S. M. Carr, M. George, U. B. Gyllensten, K. M. Helm-Bychowski, R. G. Higuchi, S. R. Palumbi, R. D. Sage and M. Stoneking (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* 26: 375-400.

Wolstenholme, D. R. and K. W. Jeon (1992). Mitochondrial genomes. Academic Press, San Diego, California, pp. 178-179.

Wrishcnik, L. A., R. G. Higuchi, M. Stoneking, H. A. Erlich, N. Arnheim and A. C. Wilson (1987). Length mutations in human mitochondrial DNA: Direct sequencing of enzymatically amplified DNA. *Nuc. Acids Res.* 15: 529-542.

MICHIGAN STATE UNIV. LIBRARIES



31293010208324