

MSC

2.º  
CICLO

FCUP  
2019

U. PORTO

Tracing Maternal Ancestry of Palenque de San  
Basilio Population

Rita Jourdan Rebouta

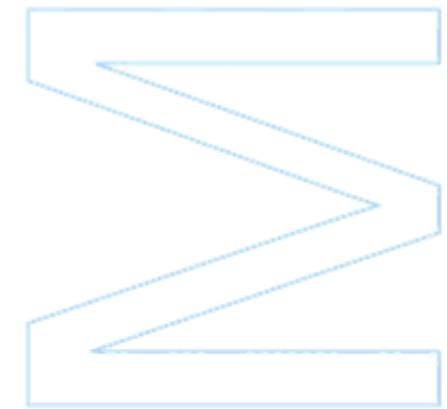
FC

U. PORTO  
FACULDADE DE CIÊNCIAS  
UNIVERSIDADE DO PORTO

# Tracing Maternal Ancestry of Palenque de San Basilio Population

Rita Jourdan Rebouta  
Dissertação de Mestrado apresentada à  
Faculdade de Ciências da Universidade do Porto em  
Biologia  
2019

U. PORTO  
FACULDADE DE CIÊNCIAS  
UNIVERSIDADE DO PORTO





# Tracing Maternal Ancestry of Palenque de San Basilio Population

Rita Jourdan Rebouta

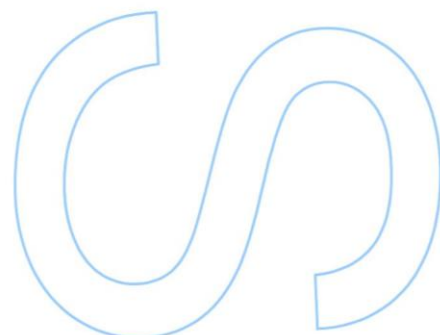
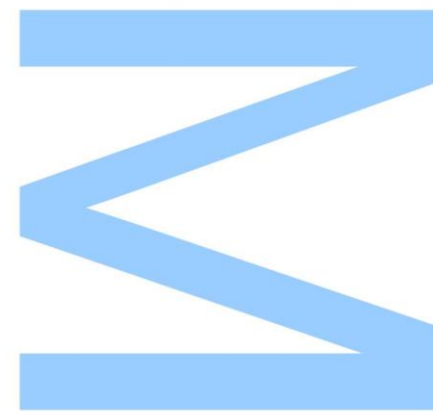
Mestrado em Genética Forense  
Departamento de Biologia  
2019

## **Orientador**

Leonor Gusmão, PhD, UERJ/ IPATIMUP

## **Coorientador**

Verónica Gomes, PhD, IPATIMUP/I3S  
Filipa Simão, PhD, UERJ

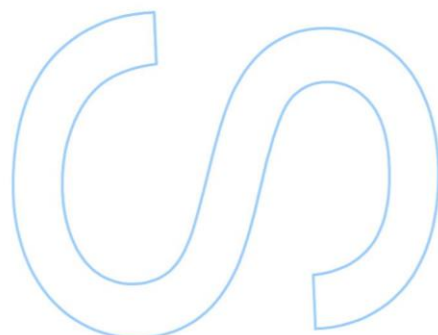
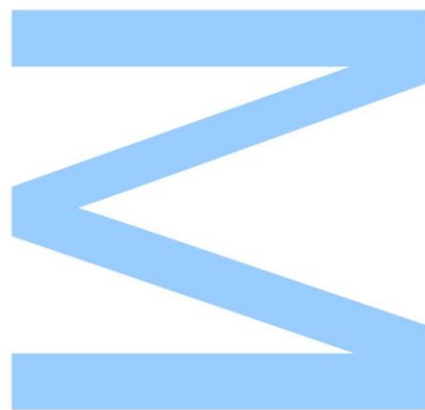




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, \_\_\_\_/\_\_\_\_/\_\_\_\_



“O meu caminho é pelo infinito fora até chegar ao fim!

Se sou capaz de chegar ao fim ou não, não é contigo, deixa-me ir...

(...)

Porque eu (...) posso ser tudo ou posso ser nada, ou qualquer coisa,

Conforme me der na gana...”

Álvaro de Campos

# Agradecimentos

Em primeiro lugar, quero agradecer às minhas orientadoras, pois sem elas este trabalho não existiria. Em particular, à Verónica, pela dedicação a esta área e pela ajuda e paciência quando as coisas pareciam não estar a resultar. À Leonor e à Filipa, que embora a 8 000 km de distância, fizeram tudo o que puderam para me ajudar a concretizar este trabalho. Pelo voto de confiança que me concederam, embora nem sempre merecido, estarei para sempre muito agradecida.

Em segundo lugar, quero também agradecer a todo o pessoal do laboratório, por me ter acolhido. Com uma nota especial à Inês, que enquanto lá esteve, foi sempre um porto seguro dentro do laboratório, sempre prestável a tirar dúvidas e estava sempre disposta a ajudar no que fosse preciso. Agradeço também à Mafalda da sequenciação, que me ajudou quando as coisas não funcionaram no laboratório e me deu dicas preciosas do que poderia fazer diferente.

Não poderia deixar de agradecer aos meus colegas de mestrado pelos almoços e cafés, e desesperos em conjunto. E claro por todas as gargalhadas quando já parecia não haver motivo para sorrir. Com uma especial nota à Carol e à Iris, pelo companheirismo e amizade travada desde o início e que perdura até hoje.

A todos os meus amigos, espalhados por Braga e pelo Porto e até Madrid, do ensino básico à universidade, passando brevemente pelo ensino secundário, e conhecidos que se vieram a tornar as pessoas mais importantes da minha vida. Não quero correr o risco de me esquecer de ninguém por isso não vou enunciar nomes, mas todos vocês, de uma forma ou de outra, fizeram de mim quem eu sou hoje e por isso estou muito agradecida.

À minha mãe, por me ter tido e pela paciência que às vezes lhe falta ter comigo. Ao meu pai, pela ajuda financeira, sem a qual seria impossível alguma vez frequentar este curso e pela tua influência em querer seguir e fazer Ciência. À minha irmã, por me chatear sem fim. Sei que não seria a mesma pessoa se não te tivesse como irmã. Aos meus irmãos pequeninos, por me fazerem sorrir quando mais estou em desespero. Espero um dia conseguir retribuir.

Dedico ainda este trabalho à Ana e à Valeria, pois sem elas sei que não teria conseguido chegar aqui e não sei o que seria da minha sanidade mental sem o vosso acompanhamento. Foram vocês que me fizeram ver que eu conseguia fazer tudo aquilo a que me propusesse, se somente acreditasse que conseguia.

Ao meu melhor amigo, companheiro, namorado, a pessoa com quem eu partilho todos os momentos da minha vida e que espero continuar a partilhá-los no futuro. Gonçalo, obrigada por me fazeres querer continuar e ser cada vez a melhor versão de mim mesma e por me ajudares a levantar sempre que caio (inclusive quando é literalmente). A tua ajuda nesta tese e na minha vida foi mais do que valiosa, foi incalculável e por isso estarei sempre agradecida.

À Kitty e à Nina, que mesmo com os nomes foleiros dados por uma miúda de 12 anos, foram, são e sempre serão as melhores companheiras e ouvintes dos meus desaires. Podem não ser da mesma espécie que eu, mas por vezes entendem-me melhor que os da minha própria espécie e levantam-me o ânimo como ninguém.

E por fim, mas não menos importante, à minha companheira de todas as horas, mesmo quando eu não quero a presença dela. À minha Esclerose Múltipla, que gosta de me lembrar que posso não ser uma lebre na corrida da vida, mas que como a tartaruga, se andar devagarinho, também consigo chegar à meta final e sobreviver.

# Abstract

On the break of the 15<sup>th</sup> century many Iberian colonizers arrived in South America, and with them many abducted imprisoned slaves who were taken from their homes and sold as merchandise to work on the crops of the newly discovered continent.

It is believed that *Palenque de San Basilio* was founded during these colonial times by a group of slaves that managed to escape captivity from *Cartagena de Indias*, then a major city of the Spanish empire. This was considered to be the first free African American town, and it is the last to remain in Colombia in present days.

In previous anthropolinguistic studies, it was reported these fugitive slaves may have been originally from the Congo region due to similarity of the language spoken in both places. Literature reporting genetic data to support this hypothesis is very scarce, however authors have managed to confirm this oral history with support from Non-recombining Y chromosome (NRY) data. The same was not accomplished regarding mtDNA, being that the aim of this work.

The data obtained through this study undoubtedly supports the hypothesis that most individuals came from Africa, when comparing the sequences obtained with data of possible ancestral populations (mainly African and Colombian populations). However, previous genetic and anthropological findings stating a sole origin from the Congo Region could not be confirmed. In fact, the most likely scenario proposed by this work is that the former slaves probably came from not one but many different locations of the African continent. This project also corroborates the theory that Palenque de San Basilio's population, after being founded by slaves, remained a relatively isolated community, for its main haplogroup frequencies were found to be from African ancestry.

The main haplogroups allocated in sub-Saharan Africa belong to the L branch. Indeed, the majority of our samples were classified into haplogroups L0, L1, L2 and L3, which was expectable, according to previous findings. A single origin for African haplotypes could not be established. In fact, there were traces of multiple origins for the observed lineages found in Palenque, connecting them with different African regions, through Network visualization.

## Resumo

No final do século XV, os colonizadores ibéricos chegaram à América do Sul e, com eles, muitos escravos, presos ou sequestrados, foram levados de suas casas e vendidos como mercadorias para trabalhar nas plantações do recém-descoberto continente.

Acredita-se que *Palenque de San Basilio* tenha sido fundado durante esses tempos coloniais por um grupo de escravos que conseguiu escapar do cativeiro em *Cartagena das Índias*, uma das principais cidades do então império espanhol. Esta foi considerada a primeira cidade afro-americana livre e é a última a permanecer na Colômbia nos dias atuais.

Em estudos anteriores de antropologia linguística, foi relatado que estes escravos fugitivos fundadores poderão ter tido origem na região do Congo devido à semelhança da língua falada no Palenque e uma das línguas faladas na República do Congo. Dados genéticos encontrados na literatura relativos a este assunto são escassos e não são suficientes para apoiar esta hipótese. No entanto, alguns autores conseguiram confirmar a história oral recorrendo a dados do cromossoma Y não-recombinante (NRY). O mesmo não foi conseguido em relação ao DNA mitocondrial (mtDNA), sendo esse o objetivo deste trabalho.

Os dados obtidos no presente estudo corroboram a hipótese, já considerada factual, de que a maioria dos indivíduos fundadores terão tido a sua origem no continente africano, ao comparar as sequências obtidas com os dados de possíveis populações ancestrais (principalmente populações africanas e colombianas). No entanto, não foi possível confirmar conclusões prévias, de dados genéticos e antropológicos, que afirmam uma única origem da região do Congo. De facto, o cenário mais provável, proposto por este trabalho, é que os antigos escravos provavelmente não vieram de um, mas sim de vários locais diferentes do continente africano.

Este projeto corrobora também a teoria de que a população de *Palenque de San Basilio*, após ser fundada por escravos, permaneceu relativamente isolada, pois as principais frequências de haplogrupos encontradas são de ancestralidade africana.

Os principais haplogrupos alocados na África Subsaariana pertencem ao ramo L. De facto, a maioria das nossas amostras foi classificada como pertencente aos haplogrupos L0, L1, L2 e L3, o que era expectável, tendo em conta dados anteriores.

Uma origem regional única para haplótipos africanos não foi estabelecida. Na verdade, há indícios de múltiplas origens para as linhagens observadas encontradas no Palenque, associando-as a diferentes regiões africanas, através da visualização de uma rede filogenética.

## Keywords

South America, Colombia, African ancestry, maternal lineages, mtDNA

## Palavras-Chave

América do Sul, Colômbia, ascendência africana, linhagens maternas, mtDNA

# Table of Contents

|                                                              |      |
|--------------------------------------------------------------|------|
| Agradecimientos                                              | ii   |
| Abstract                                                     | iv   |
| Resumo                                                       | v    |
| Keywords                                                     | vii  |
| Table of Figures                                             | x    |
| Index of Tables                                              | xii  |
| Abbreviations                                                | xiii |
| 1. Introduction                                              | 1    |
| 1.1 Genetics and the dawn of Population Genetics             | 2    |
| 1.2 Uniparental markers                                      | 6    |
| 1.3 Mitochondria: Origin, Structure and Function             | 8    |
| The mitochondrial DNA (mtDNA)                                | 9    |
| Haplotyping and Haplogrouping in mtDNA                       | 12   |
| 1.4 Evolution and Expansion of Humans: “Out-of-Africa” model | 14   |
| Entrance in the Americas                                     | 16   |
| 1.5 The Slave Trade                                          | 18   |
| <i>Palenque de San Basilio: A Brief History</i>              | 20   |
| 2. Aims                                                      | 23   |
| 3. Materials and Methods                                     | 25   |
| 3.1 DNA Sampling Strategy                                    | 26   |
| 3.2 DNA Analysis                                             | 27   |
| 3.3 Data Analysis                                            | 32   |
| 4. Results and Discussion                                    | 35   |
| 4.1 mtDNA diversity in Palenque                              | 36   |
| 4.2 Genetic Diversity Comparisons                            | 43   |
| 4.3 Phylogeography of lineages from Palenque                 | 48   |

|                |    |
|----------------|----|
| 5. Conclusions | 56 |
| References     | 58 |
| Annexes        | 69 |

## Table of Figures

|                                                                                                                                                                                                                                                                                                                |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 1 – Uniparental mode of inheritance. The mtDNA and MSY are passed through generations escaping from meiotic recombination, being transmitted in a single block (haploid) (Adapted from <a href="https://genographic.nationalgeographic.com">https://genographic.nationalgeographic.com</a> ). _____     | 6  |
| Figure 2 - Illustration of the structure of a mitochondrion commonly found in eukaryotic animal cells, with reference to elements enclosed by its walls such as the mitochondrial DNA (Taken from <a href="https://biologydictionary.net/mitochondria">https://biologydictionary.net/mitochondria</a> ). _____ | 8  |
| Figure 3 – Schematic representation of the mitochondrial circular chromosome, highlighting the hyper-variable regions 1 and 2 (HVRI and HVRII) (adapted from Hagelberg et al., 2008). _____                                                                                                                    | 10 |
| Figure 4 – Human mtDNA phylogenetic Tree (Adapted from <a href="http://phylotree.com">phylotree.com</a> ). _____                                                                                                                                                                                               | 12 |
| Figure 5 - Schematic representation of the human first main migrations by the geographic distribution of the major mtDNA haplogroups. Haplogroups are designated by their letter. (Adapted from Stewart and Chinnery, 2015) _____                                                                              | 13 |
| Figure 6 – Phylogenetic Tree of the subfamily Homininae, comprised of the tribes Hominini and Gorillini (Adapted from Almécija et al. 2013). _____                                                                                                                                                             | 14 |
| Figure 7 – The African slave maritime trade routes in the XVII until XIX century (Adapted from UNESCO) _____                                                                                                                                                                                                   | 19 |
| Figure 8 – Palenque de San Basilio shown in a map, alongside Cartagena de Indias (Noguera et al., 2013). _____                                                                                                                                                                                                 | 20 |
| Figure 9 - Schematics of the Hyper-Variable Regions in mtDNA Control Region. ____                                                                                                                                                                                                                              | 27 |
| Figure 10 - Schematics of the Hyper-Variable Regions in mtDNA Control Region. ____                                                                                                                                                                                                                             | 29 |
| Figure 11 –Haplogroup distribution within Palenque population in comparison to the geographical subgroups in Palenque de San Basilio. _____                                                                                                                                                                    | 36 |
| Figure 12 – Haplogroup frequencies of recent African ancestry present in the Americas (adapted from Salas et al., 2004). _____                                                                                                                                                                                 | 39 |

Figure 13 – Network of Palenque’s population. It shows a clear separation between its major haplogroups found. It is also shown that there is low haplotype diversity \_\_\_\_\_ 41

Figure 14 – Amerindian Haplogroups found in Colombia and Palenque (Adapted from Xavier et al., 2015). 1 (Melton et al., 2007); 2 (Xavier et al., 2015); 3 (Salas et al., 2008); 4 (Torres et al., 2006); 5 - Present Study. \_\_\_\_\_ 42

Figure 15 –MMD of Palenque population based on MNPD values. \_\_\_\_\_ 43

Figure 166 – MDS plot based on the  $F_{ST}$  genetic distances, between the 13 populations analysed, using the 74 samples from Palenque with African ancestry. Angola Khoisan population was excluded from this MDS. \_\_\_\_\_ 47

Figure 17 – Network of sub-Haplogroup L1b. \_\_\_\_\_ 49

Figure 18 – Network of sub-Haplogroup L1c. \_\_\_\_\_ 50

Figure 19 - Network of sub-Haplogroup L2a1. \_\_\_\_\_ 52

Figure 20 - Network of sub-Haplogroup L3d. \_\_\_\_\_ 53

Figure 21 - Network of sub-Haplogroup L3e. \_\_\_\_\_ 54

# Index of Tables

|                                                                                                                                                                                                                                                                                                             |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 1 – Sequences of primers utilized in the mentioned reactions. _____                                                                                                                                                                                                                                   | 27 |
| Table 2 – Primers used in the Sanger sequencing reaction. _____                                                                                                                                                                                                                                             | 30 |
| Table 3 – List of populations found in the literature for comparison to this study's population in the African continent. _____                                                                                                                                                                             | 33 |
| Table 4 - – List of populations found in the literature for comparison to this study's population in Amerind populations found in Colombia, South America. _____                                                                                                                                            | 33 |
| Table 5 – Haplogroup frequencies in Palenque discriminated according to sampling region. _____                                                                                                                                                                                                              | 37 |
| Table 6 - Haplotype Diversity (H) and Mean Number of Pairwise Differences of Palenque vs other populations (positions between 16030-16391 and 73-322 were used for minimal resolution). _____                                                                                                               | 44 |
| Table 7 – Matrix of pairwise $F_{ST}$ genetic distances (below the diagonal) and corresponding P-values (above diagonal) obtained for 10100 permutations, for mtDNA control region, between 74 Palenque samples (with African ancestry) and the 12 other African populations found in the literature. _____ | 46 |
| Table 8 – African Haplogroup Frequencies found in 5 different regions of Africa and in the Palenque population. _____                                                                                                                                                                                       | 48 |

# Abbreviations

|        |                                             |
|--------|---------------------------------------------|
| ATP    | Adenosine TriPhosphate                      |
| bp     | base pairs                                  |
| CR     | Control Region                              |
| D-Loop | Displacement Loop                           |
| DNA    | DeoxyriboNucleic Acid                       |
| H      | Haplotype diversity                         |
| HVR    | Hyper-Variable Region                       |
| ISFG   | International Society for Forensic Genetics |
| InDel  | Insertion Deletion                          |
| k      | Number of haplotypes                        |
| MDS    | MultiDimensional Scaling                    |
| MMS    | MisMatch Distribution                       |
| MNPD   | Mean Number of Pairwise Differences         |
| MRCA   | Most Recent Common Ancestor                 |
| MSY    | Male Specific region of the Y chromosome    |
| mtDNA  | mitochondrial DNA                           |
| N      | Number of individuals                       |
| OOA    | Out-Of-Africa                               |
| PAR    | Pseudo-Autosomal Regions                    |
| PCR    | Polymerase Chain Reaction                   |
| PS     | Present Study                               |
| RAO    | Recent African Origin                       |

|        |                                                                  |
|--------|------------------------------------------------------------------|
| rCRS   | revised Cambridge Reference Sequence                             |
| RNA    | RiboNucleic Acid                                                 |
| ROS    | Reactive Oxygen Species                                          |
| RSRS   | Reconstructed Sapiens Reference Sequence                         |
| S      | Number of Segregating Sites                                      |
| sd     | standard deviation                                               |
| UNESCO | United Nations Educational, Scientific and Cultural Organization |

## 1. Introduction

## 1.1 Genetics and the dawn of Population Genetics

The word *genetics* comes from the ancient Greek *genesis* meaning “origin”. As it were, Genetics comprises the biological studies in heredity, particularly the mechanisms of hereditary transmission, patterns of inheritance and the variation of inherited characteristics among similar or related organisms, related to genes and genetic information. Within genetics, there are several ramifications where this field of investigation intersects with others to yield knowledge and insight upon how the world came to be the way it is today (Griffiths *et al.*, 1987; Hartl and Clark, 1997).

In 1859, Charles Darwin wrote “On the Origin of the Species”, the most famous theory of Evolution, giving birth to a new scientific field and becoming its father. At the same time Wallace reached the same conclusion as Darwin, but with a more radical view than Darwin’s according to his peers at the time due to his complete disregard of *Lamarckism* and claim for *Mutationism* instead. Simultaneously and unaware of these studies, Gregor Mendel in 1866 proceeded with his work on pea-plants, formulating the Laws of Inheritance, granting him the posthumous title of father of modern genetics. Mendel’s work in the Theory of Heredity (1866) is appointed as the first in a field that would soon become a giant within Biology and cemented the way for genetics in the following century to thrive, when in a few decades later it was brought back into the spotlight.

The beginning of the 20th century saw Hugo de Vries and Carl Correns independently rediscover Gregor Mendel's work. In 1905 William Bateson, along with Reginald Punnett and Edith Saunders, deepened Mendel's work with pea plants, noticing there wasn't always an independent assortment of genes, thus introducing the term linkage. Bateson is also attributed to have first used the term Genetics. Biometricians Karl Pearson and Walter Weldon, co-authors of *Biometrika*, greatly opposed this newly improved theory arguing that mutations, as Mendelians would suggest, couldn't be the explanation for evolution due to their empirical observation of a continuous variation, as the Darwinist view implied.

Later on, circa 1910, Thomas Morgan released a study regarding fruit flies as a subject study, where he not only helped to explain Bateson and colleagues' work but also begun to join the two apparent diverging ideas of Mendel and Darwin. He concluded that evolution could be caused by mutation but also that variation was continuous instead of a drastic shift.

Only around 20 years later was this scientific dispute finally settled with the works of Ronald Fisher, Sewall Wright and JBS Haldane, delivering a resolution for what seemed a conflict of ideas (Fisher, 1930; Wright, 1931; Haldane, 1932). The combined works of Fisher, Haldane and Wright managed to illustrate how Mendel's laws of inheritance (Mendelian genetics) could correlate with the Darwin's idea of continuous Evolution driven by Natural Selection. This was the critical first step in developing a unified theory of evolution. These authors provided its first theoretical foundations and paved the way for the discipline of Population Genetics to emerge, thus initiating the period of "Evolutionary Synthesis" (Millstein and Skipper, 2006).

Finally, in 1937, with Theodosius Dobzhansky's work "Genetics and the Origin of Species", a bridge between these two opposing groups, population geneticists and field naturalists, was finally established. Dobzhansky adapted the conclusions reached by Fisher, Haldane and Wright in their highly mathematical papers to the natural world, turning them into a more practical view, and thus accepted to those who studied it.

To summarize, population genetics was born out of evolutionary biology's need to answer the most pressing question of how to quantify in biology. Even though the main accepted dogma nowadays was changed when new and improved information was added, these ideas still prevail and are put into practice by biologists. It was of utter importance for biology to grow as a science to be paired up with statistics and be able to turn it into a quantifiable science such as physics or chemistry. As such, population genetics is a more objective science than its relatives due to this fact.

## Human Population Genetics

The main purpose of Population Genetics as a scientific field is to develop ways to study genetic variation, defined by discrete genetic changes across the human genome, within and between populations, by studying their genotypic differences at both intra- and inter-population levels. Additionally, it aims to grasp the fundamental mechanisms behind this differentiation process, which consist of mutations, which are the primary source of diversity and generate variation, and many evolutionary processes that shape the variation among populations. Factors such as natural selection, genetic drift or gene flow (migration) are among the most notorious, and are highly influenced by historical and demographic circumstances (Chakravarti, 1999(Hartl and Clark, 1997; Relethford and Harding, 2001)

In sum, Population genetics strives to interpret and understand how genetic diversity (allelic diversity) accumulated over time, leading to the distribution seen today in the current human landscape, and which were the events responsible for those patterns (Hartl and Clark, 1997; Relethford and Harding, 2001; Cavalli-Sforza and Feldman, 2003).

Throughout the last 30 years, in simultaneous with the discovery of new genetic markers, new techniques and softwares for data analysis were developed, giving important contributions to unveil patterns of genetic variation in populations and underlying processes (Hedrick, 2009).

## Applications of Genetics: A Trade of Information

Ever since humanity has found the ability to record its history that there were great thinkers who wondered where they came from. And they all started with a need to answer the most pressing question: where do we come from and how did we get here? In present day, one field that studies this question is Anthropology (Roberts, 2011; Stoneking, 2017).

Anthropology is comprised of the ancient Greek words *Anthropos* and *Logos* and it stands for the study of the human being. As such, this field focuses in the investigation of the human species in its multiple dimensions, from society to culture to Biology. In the specific case of the latter, there is a subfield titled Bioanthropology, or as the current designation Human Biology, dedicated to study every biological aspect of humans. An example of this is Genetics that, being a part of Biology, can also be paired up with Anthropology creating the subdiscipline of Genetical Anthropology. It takes the knowledge and theory of Anthropology and complements with the practical work of Genetics. Evolutionary Anthropology, with the sight of Evolution applied to the known human history, comes together with Genetical Anthropology to give the most accurate theory as possible to the truth (Roberts, 2011; Anastasiou and Mitchell, 2013; Jobling, 2013; Stoneking, 2017).

Another important subject is Paleogenetics, which deals with the study of ancient DNA found in fossil skeletal remains, and is also a crucial tool providing data with new insight to hypothesise new theories or support theories in vogue (Green *et al.*, 2010; Roberts, 2011; Anastasiou and Mitchell, 2013).

From Medicine to the Legal aspect, there is a wide spectrum of areas in human life where Genetics can be of service. It can shed light into questions not answered before, in fields such as medical genetics or forensic genetics (Goodwin, Linacre and Hadi, 2007; Parson *et al.*, 2014).

Forensic genetics is the application of genetics to legal disputes. Databases containing the alleles or haplotypes found (depending on the type of marker) were created with the purpose of being able to calculate the frequencies at which a given genetic profile may appear within a population. Therefore, forensic genetics gives usage to population genetics postulates to investigate individuals in present day populations (Budowle *et al.*, 2003; Goodwin, Linacre and Hadi, 2007; Behar *et al.*, 2012; Parson *et al.*, 2014; Simão *et al.*, 2018).

## 1.2 Uniparental markers

The mitochondrial DNA (mtDNA) and the Y Chromosome are the commonly designated uniparental markers, accounted for their particularity of being transmitted from only one of the progenitors. While the Y chromosome is transmitted from male individuals exclusively to their male descendants, the mtDNA is transmitted from the female individual to both their male and female offspring. However, unlike the Y chromosome, only the daughters have the ability to transmit it to the subsequent generation. Consequently, a paternal and maternal lineage, respectively, can be formed as seen in figure 1 (Pakendorf and Stoneking, 2005; Charlesworth, 2009; Jobling, 2013).

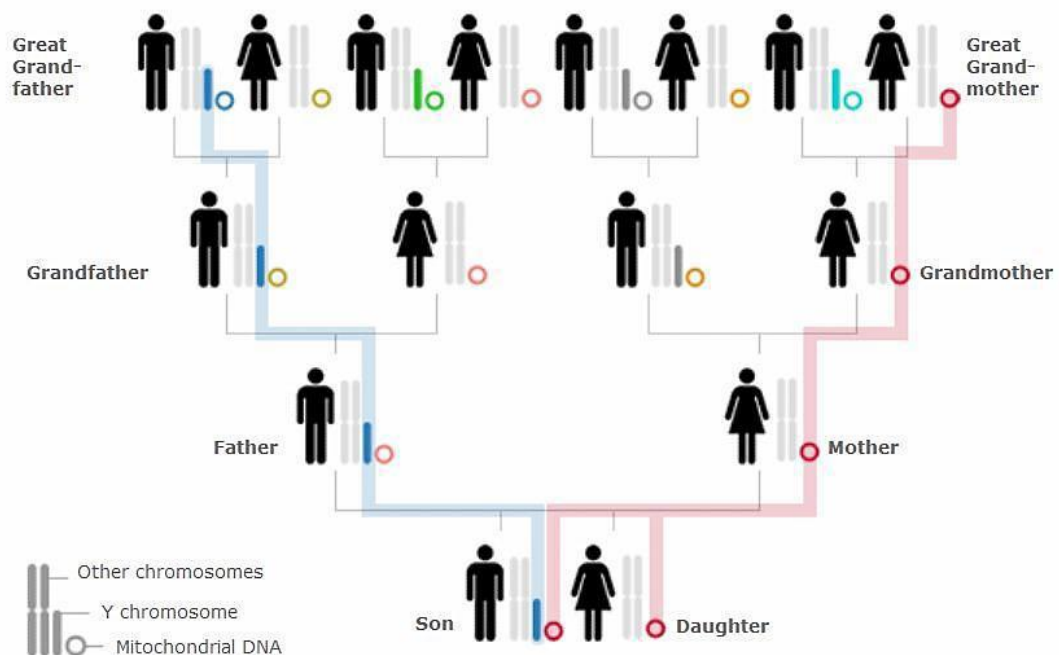


Figure 1 – Uniparental mode of inheritance. The mtDNA and MSY are passed through generations escaping from meiotic recombination, being transmitted in a single block (haploid) (Adapted from <https://genographic.nationalgeographic.com>).

Given that the mitochondrial DNA has a maternal inheritance and that the mitochondria are haploid cellular organelles, which means that for every fragment of information there is only one source, they can be powerful tools and particularly useful in studies of population genetics as well as in certain cases of legal affairs, falling in the

department of Forensic Genetics (Goodwin, Linacre and Hadi, 2007; Behar *et al.*, 2012; Parson *et al.*, 2014).

Moreover, as recombination is absent and the haploid genome is theoretically transmitted unaltered along generations, its effective population size is  $\frac{1}{4}$  relatively to autosomes, making it more sensitive to demographic population events such as bottleneck, genetic drift and founding effect (Avice, 1991; Mishmar *et al.*, 2003; Pakendorf and Stoneking, 2005).

Additionally, to qualify as a uniparental marker, it cannot go through recombination, making it possible to trace the unaltered transmission of a particular haplotype through various generations. Of course, mutation still occurs in both markers, even if smaller in comparison to other markers, making it an exception to the theoretical rule (Wallace, 1994; Santibanez-Koref *et al.*, 2019).

Another exception is the pseudo-autosomal regions (PAR) present in the Y chromosome, needed for the meiosis, making the marker called specifically Male Specific region of Y chromosome Y (MSY). Its previous denomination was Non-recombining Y (NRY) until recent years when inter-chromosomal recombination was described making this designation erroneous (Karafet *et al.*, 1999; Goodwin, Linacre and Hadi, 2007; Gomes *et al.*, 2010; Battaglia *et al.*, 2013; Jobling, 2013).

In the last few decades, Y chromosome and mtDNA analysis have become crucial tools to investigate a number of questions involving the human origin and the history of human populations (Nei, 1995; Budowle *et al.*, 2003; Torroni *et al.*, 2006; van Oven and Kayser, 2009; Auton *et al.*, 2015).

### 1.3 Mitochondria: Origin, Structure and Function

The origin of mitochondria is thought to have happened when a primordial bacterial cell with an ability of aerobic respiration established a symbiotic relationship with a larger prokaryotic cell. This is known as the Theory of Endosymbiosis or Symbiogenesis (Margulis, 1981; Anderson *et al.*, 1981; Andrews *et al.*, 1999; Lodish *et al.*, 2003; Alberts *et al.*, 2014) .

Mitochondria are a type of cellular organelle found in almost all eukaryotic cells. They are composed of two membranes: the outer membrane, which is permeable to small molecules, and the inner membrane, which is more selective, and has several invaginations, named mitochondrial ridges, that increase its surface area, as is illustrated in Figure 2. Between the two membranes there is a space called the inter-membrane space, which also forms the cristae. The shape of cristae in the mitochondria is highly variable between species and the number can even be variable between cell types, depending on its ATP demand. Inside the mitochondrion, within the inner membrane it is present a mitochondrial matrix, where enzymes, ribosomes, RNA and DNA are dispersed (Lodish *et al.*, 2003; Torroni *et al.*, 2006; Alberts *et al.*, 2014).

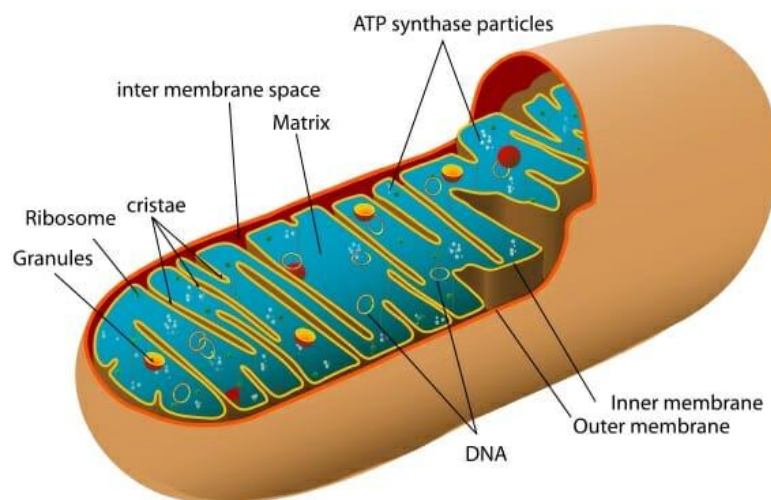


Figure 2 - Illustration of the structure of a mitochondrion commonly found in eukaryotic animal cells, with reference to elements enclosed by its walls such as the mitochondrial DNA (Taken from <https://biologydictionary.net/mitochondria>).

The key purpose of the mitochondria within the cell is to provide the energy needed for cellular activity, in the form of Adenosine Triphosphate (ATP), through the transformation of monosaccharides via a process known as Oxidative Phosphorylation (Saraste 1999; Jobling et al. 2013) or the Krebs cycle. Additional functions include participation in apoptosis (programmed cell death) signalling, cell homeostasis, calcium ions storage and immunity response (Voet *et al.* 2013). Lastly, some mitochondrial functions are specific to certain cell types. For example, mitochondria present in the brain was found to have estrogen receptors, meaning they are sensitive to hormonal signalling (Klinge, 2008). Mitochondrial diseases are often the result of mutations in the genes regulating these functions (Jobling, 2013; Stewart and Chinnery, 2015; Burr, Pezet and Chinnery, 2018).

Typically, in mammals, there is more than one mitochondrion per cell, being able to reach hundreds within the same cell. The number of mitochondria per cell typically varies between cell types. Organs such as liver and lungs usually have the highest count of these organelles whereas the red blood cells, or hematocytes, have none. They also vary between species. For example, in human lung fibroblasts it was found to have an average of approximately 300 mitochondria per cell, but in homologous rabbit cells it was found to have more than double that number (Robin and Wong, 1988).

## The mitochondrial DNA (mtDNA)

Each mitochondrion encloses inside its walls a genome of its own, a circular chromosome called mitochondrial DNA (mtDNA), with each chromosome considered a single locus. It is found that to a single mitochondrion can be attributed more than one mtDNA copy, with a believed to be maximum number of ten copies per organelle, which can then be translated as thousands of mtDNA copies per cell (Robin and Wong, 1988). Mitochondrial DNA is therefore known to replicate by binary fission, as opposed to during the meiotic process alone (Griffiths *et al.*, 1987).

In particular, the human mitochondrial DNA is 16,569 base pairs (bp) in length and is composed by two strands, the heavy strand and the light strand, differing in the amount of guanine and cytosine present in each (Anderson *et al.*, 1981; Jobling, 2013). The mtDNA is comprised as well of two regions: the coding region and the non-coding region. The coding region represents the majority of mtDNA and encodes 37 genes

(Wallace, 1994). The non-coding region, also known as Control Region or Displacement loop (D-loop), contains the hyper-variable regions (HVR), which are more susceptible to mutation than the rest of the mtDNA and therefore contain sequences that are more variable within the population (figure 3) (Walberg and Clayton, 1981; Meyer, Weiss and Haeseler, 1999). The majority of variability regarding the mtDNA comes from these small regions (Budowle *et al.*, 2003).

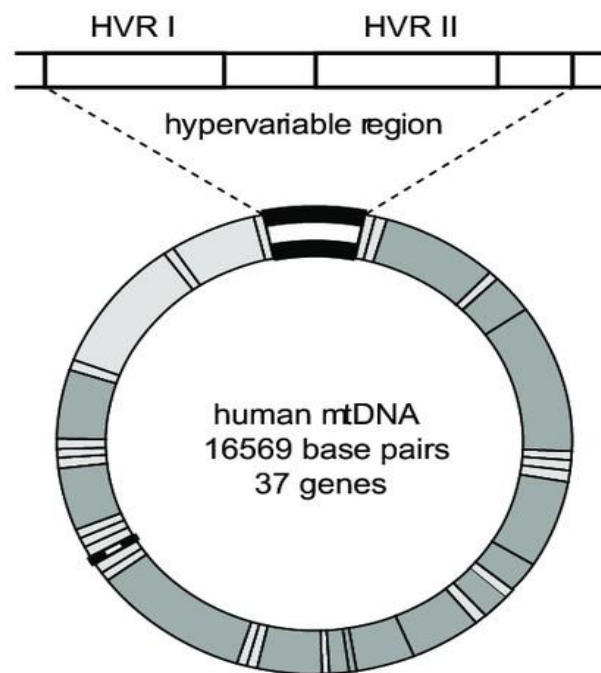


Figure 3 – Schematic representation of the mitochondrial circular chromosome, highlighting the hyper-variable regions 1 and 2 (HVR I and HVR II) (adapted from Hagelberg *et al.*, 2008).

Since recombination, theoretically, does not occur in this haploid genome, the mtDNA sequence is highly conserved within lineages, leading to a Uniparental inheritance. Opposingly, the mutation rate is nonetheless higher in comparison to nuclear DNA (Parsons *et al.*, 1997). This fact is explained due to the high levels of free oxygen radicals or reactive oxygen species (ROS) as a result of the ATP production, the low fidelity of mtDNA polymerase and the scarcity of mtDNA repair mechanisms associated with the absence of protective proteins, like histones (Griffiths *et al.*, 1987; Jobling, 2013; Li *et al.*, 2013). These mutations accumulated over time and led to a partition of the human population into several discrete haplogroups (Stewart and Chinnery, 2015).

These traits are particularly advantageous when studying the history and evolution of human populations (Pakendorf and Stoneking, 2005; Torroni *et al.*, 2006; Roberts, 2011; Stoneking and Krause, 2011).

## Haplotyping and Haplogrouping in mtDNA

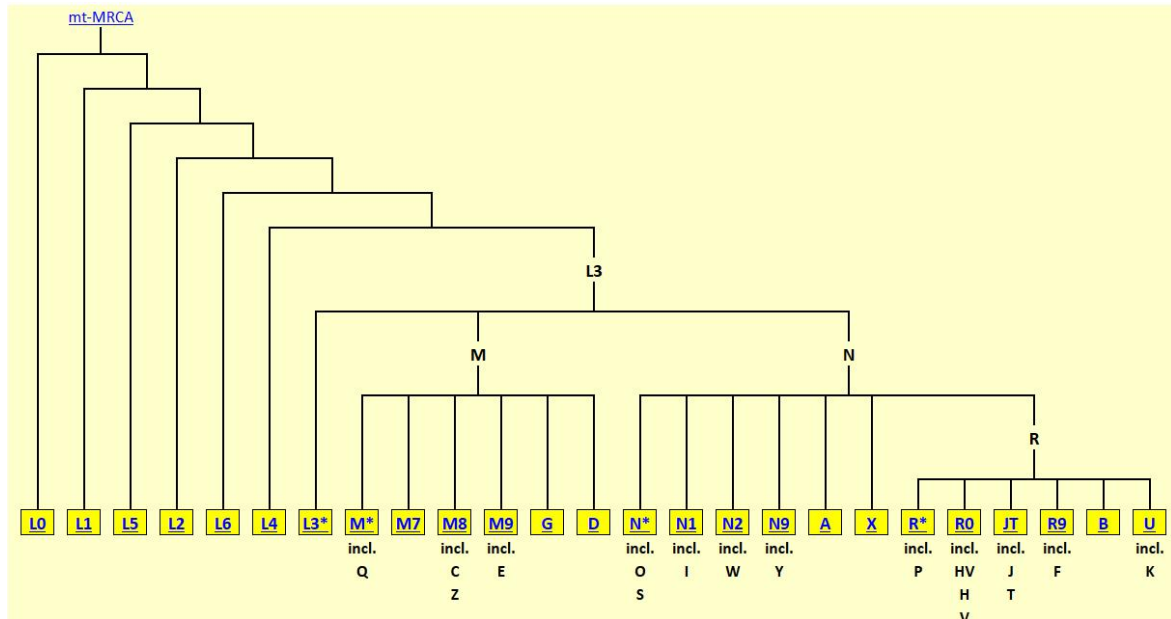


Figure 4 – Human mtDNA phylogenetic Tree (Adapted from phylotree.com).

The revised Cambridge Reference Sequence (rCRS) refers to the mtDNA reference sequence used in studies conducted regarding the aforementioned marker, regarding present human populations, published by Andrews *et al.* in the year of 1999, and previously agreed upon by a consortium of geneticists specialized in the subject. This was a revision of the original reference sequence, first published in 1981 by Anderson *et al.*, in which reading errors were corrected as well as some polymorphisms were replaced. Recently a replacement of this revision was proposed, called Reconstructed Sapiens Reference Sequence (RSRS) (Behar *et al.*, 2012). However, most researchers believe the proposed switch to RSRS would inevitably lead to notational chaos, mistakes and misinterpretations in the literature and therefore, it was preferred to maintain the employment of the rCRS as the present reference sequence to report mtDNA profiles (Bandelt *et al.*, 2014).

Human populations are very diverse regarding mtDNA as shown by Figure 4. As previously mentioned, mtDNA is not invulnerable to mutation, being that the cause for the observed variability between individuals.

African populations are known to have the highest genetic diversity in the globe and as such the mtDNA lineages found there fall into several distinctive paraphyletic clusters and monophyletic haplogroups (Tishkoff and Verrelli, 2003).

There is a clear separation between lineages found North and South of the Saharan Desert since the most commonly found haplogroups in sub-Saharan Africa belong to the L branch, whereas in Northern Africa it is found to have haplogroups outside the L-branch, such as U and N (Salas *et al.*, 2004, 2008; Ansari-pour *et al.*, 2016). Due to this fact it is believed that the main lineages of sub-Saharan Africa at the time of the first human migration were L0, L1, L2 and L3 and that they most likely evolved there as well, as seen in Figure 5 (van Oven and Kayser, 2009; Stewart and Chinnery, 2015).

As modern humans migrated out of Africa and populated the rest of the globe, more haplogroups emerged from the African root, believed to be L3 the haplogroup from which the rest of the non-African haplogroups arose, as Figure 4 demonstrates (Mishmar *et al.*, 2003; Gomes *et al.*, 2015; Stewart and Chinnery, 2015).

In the Americas, where diversity within native population is the smallest comparatively to other human populations, the most commonly found haplogroups are A, B, C, D and X, as shown in figure 5 (Torroni *et al.*, 2006; van Oven and Kayser, 2009). They belong to the Asian branch due to their ascendance in northern Asians who eventually crossed the Bering strait to populate the American continent (Flegontov *et al.*, 2019; Stone, 2019).

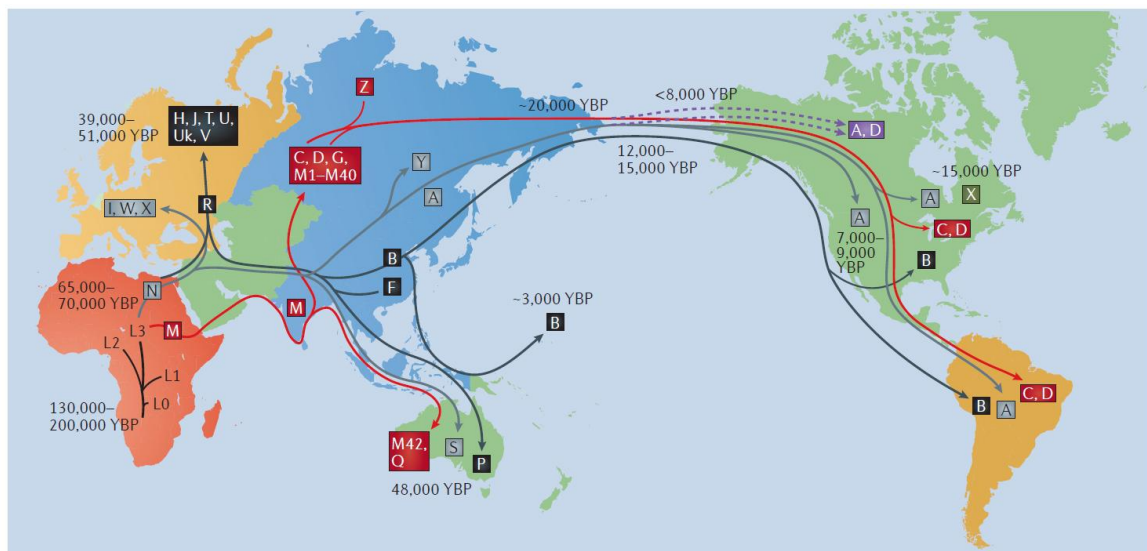


Figure 5 - Schematic representation of the human first main migrations by the geographic distribution of the major mtDNA haplogroups. Haplogroups are designated by their letter. (Adapted from Stewart and Chinnery, 2015)

## 1.4 Evolution and Expansion of Humans: “Out-of-Africa” model

Fossil evidence shows that East Africa is most likely the location where modern human ancestors first evolved from their Most Recent Common Ancestor (MRCA) shared with the chimpanzees (*Pan* genus) as depicted in figure 6. Posterior genetic analysis, with a calibrated molecular clock, indicated that the subtribes *Hominina* and *Panina* diverged between 4 and 6 million years ago. Despite nowadays not being the most ancient remains to ever be encountered, fossilised remains of the first example of an early hominine species *Australopithecus afarensis* were found there, giving momentum to the aforementioned theory (Diamond, 1991; Roberts, 2011; Behar *et al.*, 2012).

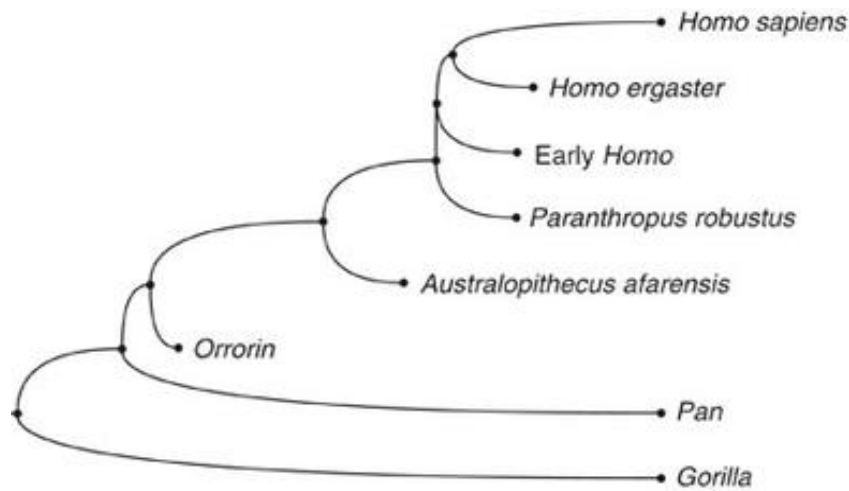


Figure 6 – Phylogenetic Tree of the subfamily *Homininae*, comprised of the tribes *Hominini* and *Gorillini* (Adapted from Almécija *et al.* 2013).

There were two hypotheses as to how modern humans first colonized the globe. The first, dubbed the model of “Out-Of-Africa” (OOA) or the “recent African origin” model, hypothesised that the first *Homo sapiens* originated in Africa between 200,000 to 100,000 years ago, and, from that point on, moved out of Africa approximately 100,000 years ago and colonized the rest of the continents, stating that all modern human populations outside sub-Saharan Africa are primarily descendants of a population of *H. sapiens* that left Africa after that time (Diamond, 1991; Nei, 1995; Stringer, 2003; Roberts, 2011).

The other counter hypothesis, called the Multiregional theory of human evolution, states that the modern humans evolved gradually from *Homo erectus*

worldwide with the effects of gene flow and natural selection, having this ancestral species move out of Africa and spread to various parts of the world around 1 million years (MY) ago and then suffer a paralleled evolution. However, genetic analysis proved this to be highly unlikely. The world population of present-day shares a great number of genetic similarities that could not be explained by a convergent evolution (Nei, 1995; Stringer, 2003).

Nowadays, the most widely accepted model is the OOA or the recent African origin model (RAO), reporting to the fact that other species of hominins such as *Homo erectus* also exited Africa prior to modern humans. However, when *Homo sapiens* emerged in Africa, they had higher fitness adaption to harsher environments and as such a replacement of their sister species and subspecies occurred (Diamond, 1991; Roberts, 2011).

Nonetheless this replacement was not complete as some recent studies suggest. In the last decade, population genetics studies have revealed evidence of breeding between *H. sapiens* and archaic humans in Eurasia and Oceania, meaning that to a smaller degree these populations also descended from archaic humans found in said locations (Green *et al.*, 2010; Behar *et al.*, 2012).

Genetic evidence has also been able to show how the extinct and extant hominids are related within the subclades as well as to modern humans. This data obtained from DNA comparisons between all species formed a cladogram with the human evolution and phylogeny (Roberts, 2011; Stoneking and Krause, 2011; Anastasiou and Mitchell, 2013; Jobling, 2013; Stoneking, 2017).

## Entrance in the Americas

The last continents to be colonized by humans were the Americas. The entrance of anatomically modern humans into the Americas is still to this day a highly discussed topic, especially regarding the timing, routes used and ancestry of the migrants. Nevertheless, the predominant hypothesis most researchers agree on is that approximately 25 to 15 thousand years ago migrants entered the continent, beginning the peopling process of the Americas. The routes used, however, still have many dispersed opinions, with some defending it occurred via the Pacific coast, others supporting the use of interior ice-free passages, and others believing that migrants might have used both routes (Potter *et al.*, 2018). Still, it is consensual that they initially reached the Siberian continent and then crossed the strait of Bering. Since this event most likely happened during the last glacial era, the area was then land, making the crossing of these modern humans possible (Flegontov *et al.*, 2019; Sikora *et al.*, 2019; Stone, 2019).

It is theorized that Alaska was reached c. 16,000 years ago from Northeast Asia via the Bering Sea land bridge, but further progress was barred until the continental ice sheets began to retreat c. 14,000 years ago. By this time, other groups had already bypassed the ice sheets by migrating along the Pacific coast, reaching South America by 14,500 years ago (Mishmar *et al.*, 2003; Flegontov *et al.*, 2019; Sikora *et al.*, 2019; Stone, 2019).

The idea that the Native American populations were likely migrants from Asian populations was first thought of around the sixteenth century, when the Spanish Jesuit José de Acosta, while travelling, studied these populations and wrote “Natural and moral history of the Indies (de Acosta *et al.*, 2002). Indigenous peoples of the Americas have been linked to Siberian populations first by linguistic factors and appearance only. Later on, by blood type distribution and genetic composition, with whom they share a distribution pattern (Wang *et al.*, 2007; O’Rourke and Raff, 2010; Sikora *et al.*, 2019).

Native American populations went through successive bottlenecks while dispersing through the continent, that coupled with a strong founder effect and not enough time of differentiation to regain the diversity lost, resulted in a low genetic diversity that is still present today (Wang *et al.*, 2007; O’Rourke and Raff, 2010; Xavier *et al.*, 2015; Posth *et al.*, 2018).

In recent historical times, the colonization of the “New World” by Europeans, initiated in the late XV century, as well as the transatlantic slave-trade with destination

to the Americas, in the following centuries, dramatically reshaped the population scenario of the entire continent at all levels (Salas *et al.*, 2004; Simão *et al.*, 2018).

## 1.5 The Slave Trade

The turn of XV century marked the beginning of an era that would later become known as the Maritime Discoveries. It was an unsettling time in Europe, with the threat of the Ottoman empire at its door, and its kingdoms just recovering from an epidemic of the black plague. This meant they were eager change the course in their favour in order to not be subdued by Arabian commerce and as such a Portuguese king decided to find a way to reach India and trade directly with them. Thus, began the Portuguese maritime expansion through the uncharted African coast. At first, they traded European goods for natural resources like gold and soon after other European kingdoms followed. But eventually, these resources included slaves as well, slaves that were sold by other African men. And once Europeans reached the Americas, the colonizers would need for someone to work the land (Wheat, 2011; Newson, 2012).

The idea of human beings enslaved by other humans was not novelty in the XV century, it was a notion that has been present throughout all Human History, in various formats. Almost every ancient society, from Ancient Greece and Ancient Egypt to the Persian and Roman Empires, there have always been human slaves. While this concept may not have been new, these new unmapped lands and its resident people were. The landscapes found there were nothing like they had seen before and the people had many different features, like the tone of their skin (UNESCO).

There was no immediate big scale capture of African slaves, as depicted in figure 7. As time passed, and once Europeans conquered and established its first settlers in the American continent, plantation industries began to thrive. In particular, the sugar industry became the most important raw material to be exported from the Americas, with tobacco coming only in second. With the ever-higher demand of increasing production, the colonials needed more workers, without diminishing profit. And around mid-sixteenth century, the transatlantic slave trade, or the triangular trade, was established (Thomas, 1997; Newson and Minchin, 2007 Almeida Mendes, 2008).

The majority of the slaves came from the West African coast. But at some point, the demand from Europeans was so high, African traders started to buy slaves from other traders further inland and even raiding small towns and villages, seizing captives far from the coast (Newson and Minchin, 2004).

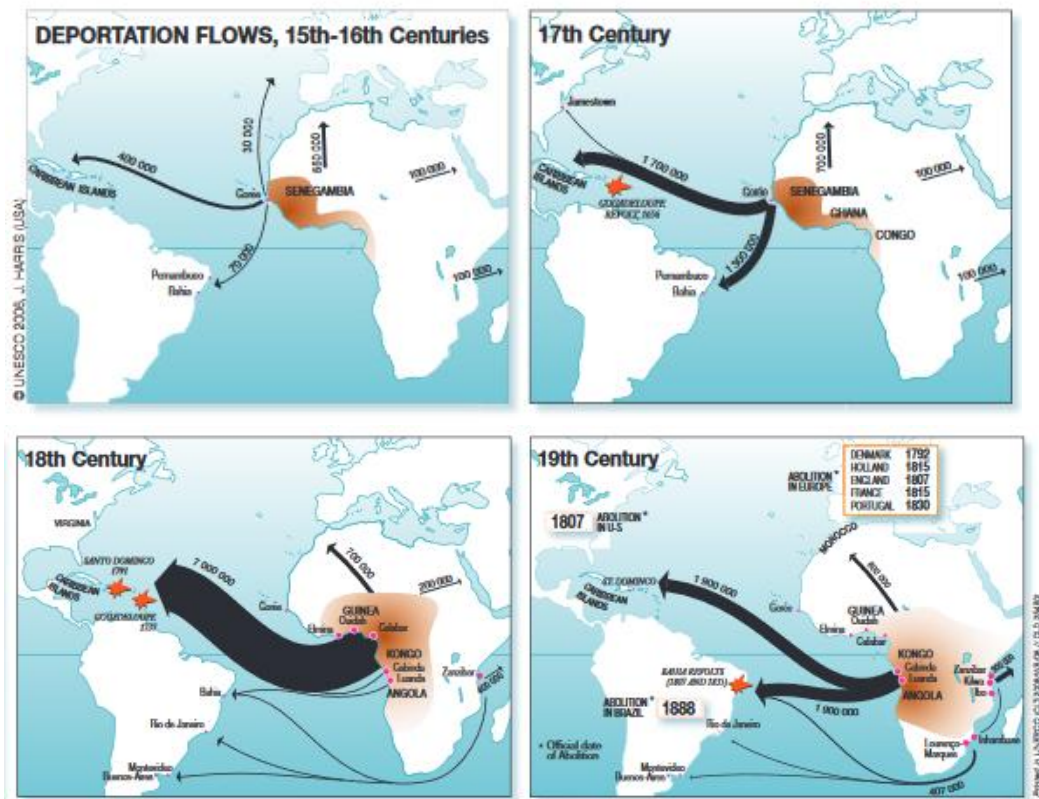


Figure 7 – The African slave maritime trade routes in the XVII until XIX century (Adapted from UNESCO)

The traffic of African slaves reached its highest number in the 18<sup>th</sup> century, with a movement of approximately 8 million people. It is estimated that between 25 to 30 million people were taken from their homes and sold as slaves. It is also estimated that between 5 and 15% of Africans in cargo ships did not survive the journey (Cohn, 1982).

By the end of the 18<sup>th</sup> century, slave abolition laws were beginning to emerge, after many revolutions in Central America occurred. In Colombia, its abolition law was passed in 1851 (UNESCO)

As a result, nowadays, the presence of genes of African descent within many American populations is common (e.g. Salas *et al.*, 2004, 2008; Noguera *et al.*, 2013; Xavier *et al.*, 2015; Martínez *et al.*, 2017; Simão *et al.*, 2018).

## *Palenque de San Basilio: A Brief History*

*Palenque de San Basilio* or *San Basilio de Palenque* refers to a village in the Municipality of Mahates, in the department of Bolivar, located in northern Colombia (Figure 8).

The history of the origin of (Palenque) has remained a mystery until recent years due to the lack of records and the fact that even the population itself could not name the origin of the population, aside from the oral history of its founder, *Benkós Biohó* (Ansari-pour *et al.*, 2016; Schwegler, 2016). It is thought that he arrived in Cartagena de Indias around the year 1596 as a slave, and presumably established the maroon community of Palenque between the end of the 16<sup>th</sup> century and the beginning of the 17<sup>th</sup> century.

Not long ago, Palenque was a secluded community near Cartagena (Figure 8), outlandish to its vicinities.



Figure 8 – Palenque de San Basilio shown in a map, alongside Cartagena de Indias (Noguera *et al.*, 2013).

*Palenques* was the name given to these communities of fugitive slaves that managed to escape captivity from Spanish controlled cities. In Portuguese territory, these towns were called *Quilombos*.

During colonial times, many imprisoned slaves were taken from the African continent and brought to the Americas against their will, by European colonizers. In particular, in Spanish and Portuguese controlled South America, it was not uncommon for slaves to try to escape from their captivity into the near and abundant forests. Once they managed to flee, they would build free towns in which they could defend themselves from the landowners chasing them to bring them back to the plantations. These fortified towns successfully resisted many attacks from the Iberian military (Wheat, 2009).

Nowadays Palenque is the only one that remains in Colombia and was considered the first free African town in Castilian-owned American soil (Schwegler *et al.*, 2011; Ansari-pour *et al.*, 2016).

What is believed by the general public is that Palenque was founded by a group of male slaves who escaped captivity from *Cartagena de Indias* (Figure 8), then the most important city for the South-American Hispanic slave trade (Navarrete, 2012; Schwegler, 2016). In addition, the population of Palenque speaks a creole Spanish called *Lengua* or Palenquero. Anthropological and linguistic studies conducted in Palenque demonstrated that the Kikongo group of languages are the probable source of origin for more than 200 words in Palenquero (Schwegler *et al.*, 2011; Schwegler, 2016). Then it led to the theory that Yombe, belonging to the Kikongo group of languages, spoken in the region of Mayombe, in the Republic of Congo, is the most likely source of the Palenquero (Schwegler, 2016). This belief has been also corroborated by Ansari-Pour *et al.* (2016), with Y chromosome analysis conducted.

## Population Genetics in Palenque

As mentioned earlier, Palenque is thought to have been built by runaway slaves, which was indicative that its present-day population probably descended from them, without any contributions from European ancestors. As such, the people residing in this area would most likely be genetically closer to African populations than to other South American Colombian populations (Schwegler *et al.* 2016).

From the moment *Palenque de San Basilio* was declared as Masterpiece of the Oral and Intangible Heritage of Humanity, in 2005, and Intangible Cultural Heritage of Humanity, in 2008, by UNESCO due to its history, traditions, ethnicities, culture and language, many anthropological and linguistic studies about Palenque were conducted (Schwegler *et al.* 2016). But studies in the literature regarding genetic data, based on mtDNA, to support the aforementioned hypothesis on the origins of Palenque are insufficient. Nevertheless, in the existing limited literature, authors have managed to confirm the oral history of Palenque being founded by fugitive slaves from the Congo region, with support from NRY data (Ansari-Pour *et al.* 2016). However, the same was not accomplished regarding mtDNA. In Ansari-Pour *et al.* (2016) the results for mtDNA were homogenous when compared to the rest of the African populations, leaving uncertainties about the place of origin of its founder population.

Due to the limited amount of studies regarding mtDNA in this population, all of them with either a low sample number or low resolution, the conclusions obtained by these studies may be superficial.

## 2. Aims

The main goal of this study was to further the existent investigation upon the maternal ancestry of Palenque population and trace the origin of the lineages found to the corresponding region of Africa, especially since genetic studies of this population group focused on mitochondrial DNA are scarce and restricted to HVR I.

Accordingly, we aimed to:

- Characterize the full control region mtDNA haplotypes present in Palenque;
- Determine this population's founding haplogroups and geographical ancestral origin or origins of the latter, if possible;
- Observe if a pattern of admixture is present within the population, from ancestral European, Native-American and African populations.

### 3. Materials and Methods

### 3.1 DNA Sampling Strategy

During this study 81 samples were analyzed. The samples correspond to maternally unrelated individuals, randomly selected, living in Palenque de San Basilio, in Colombia.

All blood samples collected were preserved in FTA™ cards (Whatman®, BioSciences). The DNA of the samples was later extracted from the bloodstains with the aid of the chelating resin Chelex®, using the Biorad standard protocols (Lareu *et al.*, 1994) and/or through the process of salting-out as described by Miller, Dykes, and Polesky (1988) and afterwards quantified in NanoDrop (Thermo Scientific).

The samples were collected after informed consent according to the requirements. In addition, approval was obtained from the ethics committee of the University of Cartagena, Colombia.

Throughout sample collection, information about the subjects was gathered to allow the division between two sub-groups: “Palenque rural” (*Palenqueros* living and working in the community) and “Palenque urbano” (*Palenqueros* working and living in Cartagena during the week, and having the family house in the community of Palenque).

## 3.2 DNA Analysis

Over the course of this work, the purpose was to sequence the total Control Region of mtDNA, between positions 16024 and 576, through Sanger Sequencing method.

### Amplification of the mtDNA control region

The CR of the mtDNA present in the samples was amplified in a single Polymerase Chain Reaction (PCR) with the pair of primers F15900/R639, as described in table 1.

Table 1 – Sequences of primers utilized in the mentioned reactions.

|        | Primer Sequence                             | Reference                    |
|--------|---------------------------------------------|------------------------------|
| F15900 | 5'- TAA ACT AAT ACA CCA GTC TTG TAA ACC -3' | (Simão <i>et al.</i> , 2018) |
| R639   | 5'- GGG TGA TGT GAG CCC GTC TA -3'          | (Simão <i>et al.</i> , 2018) |

The result amplified fragments consisted of the full CR sequence between position 16024 and 576, reaching a total length of approximately 1200 bp, for each sample.

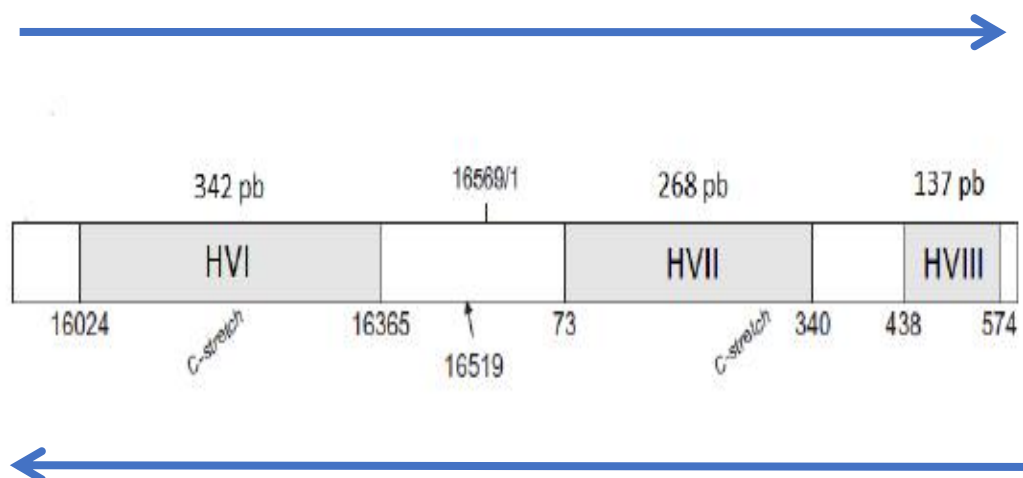


Figure 9 - Schematics of the Hyper-Variable Regions in mtDNA Control Region.

A primer solution (forward and reverse) was prepared resorting to a mother solution of each primer with an initial concentration of 100  $\mu\text{M}$ . The final concentration of each primers in the primer mix was 2  $\mu\text{M}$ .

An amplification with Qiagen Kit Amplification PCR Mixture was performed with 3.5  $\mu\text{L}$  of  $\text{H}_2\text{O}$ , 1  $\mu\text{L}$  of Primer Mix, 5  $\mu\text{L}$  of 2x Qiagen and 0.5  $\mu\text{L}$  of sample DNA, reaching a total volume of 10  $\mu\text{L}$ . A negative control was always performed. If the sample's quality was deemed low, a higher volume of extracted DNA than the one stated in the protocol would be used for best results.

The conditions set in the thermocycler for amplification were a Hot Start of 95°C for 15 minutes; 35 cycles of 94°C for 30 seconds for the Denaturation, 60°C for 90 seconds during the annealing phase and 72°C for 60 seconds for the extension step. Lastly, for the final extension, the settings were 72°C for 10 minutes.

## Electrophoresis

A posterior gel electrophoresis to confirm the amplification of the desired fragments was always performed. The preferred method was polyacrylamide gel, with a subsequent silver nitrate staining method (Orita *et al.*, 1989; Stellwagen, 2009).

This was an important step in order to be certain that the following were not performed vainly, as a confirmation that there was in indeed DNA present to be sequenced. As previously mentioned, errors in the PCR reaction could occur and as such if no DNA bands appeared in their corresponding wells, that could be a sign that either the reaction did not occur, or not enough DNA was amplified, resulting in a very light band.

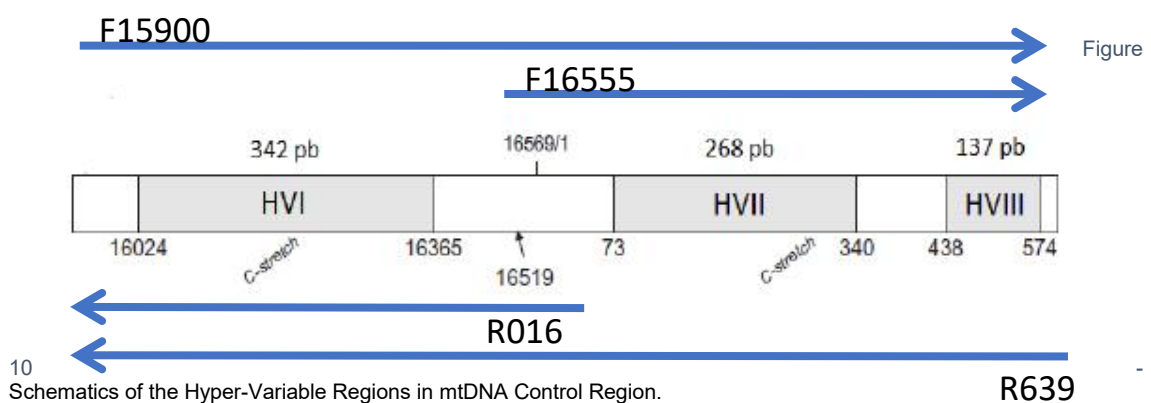
## Purification by ExoSAP

A purification by ExoSAP-IT prior to the sequencing itself was required in order to remove all the remaining dNTPs and primers from our PCR product. A total of 2 $\mu\text{L}$  of ExoSAP-IT enzyme was added to each 5 $\mu\text{L}$  of amplified product. The mixture was placed in the thermocycler at 37°C for 15 minutes (for enzymes activation) and at 80°C for 15 minutes (for enzymes inactivation).

## Sequencing through Sanger Method

The PCR products were sequenced for the full mtDNA control region, using the pairs of primers as described in table 2.

In the sequencing reaction, 8 different primers were selected, two of which were previously used in the amplification. Depending on the quality of a given sample, single coverage could suffice, usually the forward primers F15900 and F16555 were enough. However, if a given nucleotide position, in the sequence under study, seemed dubious, whether it was a mutation in the sequence causing confusion in the software analysis or if it was a sequencing error altogether, other primer would be used (table 2). Furthermore, to guarantee double coverage as to lessen the probability of committing a base assignment error, a wide variation of reverse and forward primers from the listed in table 2 could be used to sequence the vicinity of the base or bases in question (Parson *et al.*, 2014). In addition, as these sets usually are used for samples that display heteroplasmies of length and/or position, it provided the double coverage required in such occasions.



As previously stated in the amplification reaction, the primers used in a sequencing reaction also had the required concentration of 2,5  $\mu$ M.

Table 2 – Primers used in the Sanger sequencing reaction.

|               | Primer sequence                             |
|---------------|---------------------------------------------|
| <i>F15900</i> | 5'- TAA ACT AAT ACA CCA GTC TTG TAA ACC -3' |
| <i>F15997</i> | 5'- CAC CAT TAG CAC CCA AAG CT -3'          |
| <i>F16268</i> | 5'- CAC TAG GAT ACC AAC AAA CC -3'          |
| <i>F16555</i> | 5'- CCC ACA CGT TCC CCT TAA AT -3'          |
| <i>R016</i>   | 5'- CCC GTG AGT GGT TAA TAG GGT - 3'        |
| <i>R599</i>   | 5'- TTG AGG AGG TAA GCT ACA TA -3'          |
| <i>R639</i>   | 5'- GGG TGA TGT GAG CCC GTC TA -3'          |
| <i>R649</i>   | 5'- TTT GTT TAT GGG GTG ATG TGA -3'         |

The PCR Sequencing Mixture was comprised of, per sample, 0.5 µL Kit Big Dye, 1 µL Buffer, 0.5 µL *Primer* (with a concentration at 2.5 µM), 1.5 µL purified sample and 1.5 µL H<sub>2</sub>O, reaching a total volume of 5 µL. An independent reaction for each primer had to be performed as using Sanger method it must be presented in a single strand.

Sequencing reaction conditions started with 96°C for 2 minutes. For the duration of 35 cycles the apparatus was set at 96°C for 15 seconds for the denaturation, and 60°C for 2 minutes for the extension. For the final step of the reaction, the final extension, settings were 60°C for 10 minutes. However, due to difficulties of balancing primer pairing poses during the annealing phase, these temperatures could occasionally vary among samples. This happens because the temperature must be low enough to allow the primer to pair with the targeted sequence but also high enough for it to pair with the intended sequence alone. As a result, it would be common for the sequences to exhibit a blurry appearance, causing at times some difficulty in identifying which base was the correct one, at a given position. Additionally, when a reaction using the reverse primers for the HVRII was performed, the aforementioned reagents quantities and PCR conditions would be different than the rest due to its higher working

specificity. For example, there would be a third step within the cycle, omitted from the rest of the reactions.

## Final purification with Sephadex

A final purification of the sequenced product with Sephadex columns was needed prior to the capillary electrophoresis, for best results in the upcoming data analysis, as to remove the non-incorporated ddNTPs.

The Sephadex solution was prepared by weighing 10 g of Sephadex and diluting in 150 ml of water. Then, 750  $\mu$ L of Sephadex solution filled a proper column with filter, which had been previously placed in an Eppendorf tube of 2 mL and centrifuged for 4 minutes at 4400 rotations per minute (rpm). Afterwards, the column was moved into a new, well-identified Eppendorf tube, into which the purified product would soon be transferred into. The total volume of 5  $\mu$ L of sequenced product was transferred onto the center of the Sephadex column; and centrifuged for 4 minutes at 4400 rpm. To the purified product in the Eppendorf tube, 8  $\mu$ L of formamide was added and the resulting mixture was inserted on the plate of the ABI device to run the capillary electrophoresis.

The results of the capillary electrophoresis were then visualized using the Sequencing Analysis software (Applied Biosystems Sequencing Analysis Software version 6.0)

### 3.3 Data Analysis

The mtDNA sequences obtained by Sanger method were assembled and aligned against the revised Cambridge Reference Sequence – rCRS - (Andrews *et al.*, 1999), using the software *Geneious Pro 5.5.8*<sup>®</sup> (Drummond *et al.*, 2011).

The differences detected between them, named polymorphisms, led to a description of the haplotypes of each individual. This analysis was performed using the phylogenetic approach suggested by Bandelt and Parson (2008) and the haplotypes followed the nomenclature described in Parson *et al.*, (2014), as the International Society for Forensic Genetics (ISFG) recommendations for haplotyping.

With the aid of the *Haplogrep* software (Kloss-Brandstätter *et al.*, 2011), the haplotypes found were classified into haplogroups according to the most recently updated version of the mtDNA phylogenetic tree, *Phylotree 2018* (van Oven and Kayser, 2009), using the latter for confirmation of the haplotype as well.

The haplotypes in this study were later submitted to the EMPOP mtDNA database (<https://empop.online>). The sequences were analysed and checked as well confirmed for the haplogroup classifications. EMPOP mtDNA database (Parson *et al.*, 2004) follows the criteria for the nomenclature rules and phylogenetic approach established by the International Society for Forensic Genetics (Bandelt and Parson, 2008; Carracedo *et al.*, 2000). EMPOP was created to fulfil the need of a high quality mtDNA database for forensic and population studies.

Haplotype frequencies were estimated by direct counting. *Haplosearch* ( <https://haplosearch.com>; Fregel and Delgado, 2011) was then used in order to obtain a single file containing the aligned sequences due to the requirement of an upload of the full aligned sequences by the *DnaSP* software. The *DnaSP* program (currently version 6.12.01, Rozas, 2009; Rozas *et al.*, 2017) was used to estimate diversity indices within the population sample.

In order to do a comparative analysis with our study group, sequences were collected from publicly available literature on mtDNA (HVRI, HVRII and complete Control Region) from North, South, East and West African (Table 3) populations as well as Colombian Amerindian populations (Table 4).

Table 3 – List of populations found in the literature for comparison to this study's population in the African continent.

| <b>African Region</b> | <b>Country</b>   | <b>N</b> | <b>Analysed Region</b>                | <b>Authors, Year</b>                 |
|-----------------------|------------------|----------|---------------------------------------|--------------------------------------|
| North                 | Tunisia          | 64       | HVRI (16024-16400) + HVRII (056-360)  | (Turchi <i>et al.</i> , 2009)        |
|                       | Morocco          | 56       | HVRI (16024-16400) + HVRII (056-360)  | (Turchi <i>et al.</i> , 2009)        |
|                       | Morocco          | 508      | CR (16024-576)                        | (Aboukhalid <i>et al.</i> , 2013)    |
| East                  | Kenya            | 100      | CR (16024-576)                        | (Brandstätter <i>et al.</i> , 2004)  |
|                       | Uganda           | 261      | CR (16024-576)                        | (Gomes <i>et al.</i> , 2015)         |
| Central West          | Guinea Bissau    | 79       | CR (16024-576)                        | (Carvalho <i>et al.</i> , 2011)      |
|                       | Ghana            | 192      | CR (16024-576)                        | (Fendt, Röck, <i>et al.</i> , 2012)  |
| South West            | Angola           | 365      | HVRI (16024-16400) + HVRII (073-340)  | (Coelho <i>et al.</i> , 2009)        |
|                       | Angola           | 42       | HVRI (16024-16391) + HVRII (063-322)  | (Plaza <i>et al.</i> , 2004)         |
|                       | Angola           | 105      | HVRII (16030-16401) + HVRII (048-408) | (Cerezo <i>et al.</i> , 2016)        |
|                       | Angola (Khoisan) | 129      | CR (16024-576)                        | (Fendt, Huber, <i>et al.</i> , 2012) |
| South East            | Mozambique       | 147      | CR (16024-576)                        | (Cerezo <i>et al.</i> , 2016)        |

Table 4 - – List of populations found in the literature for comparison to this study's population in Amerind populations found in Colombia, South America.

| <b>Continental Region</b> | <b>Colombian Region</b> | <b>N</b> | <b>Analysed Region</b> | <b>Author, Year</b>           |
|---------------------------|-------------------------|----------|------------------------|-------------------------------|
| South America             | North                   | 188      | HVRI                   | (Melton <i>et al.</i> , 2007) |
| South America             | Central/South           | 98       | CR (16024-576)         | (Xavier <i>et al.</i> , 2015) |
| South America             | South                   | 98       | HVRI                   | (Salas <i>et al.</i> , 2008)  |
| South America             | Colombia                | 64       | HVRI + HVRII           | (Torres <i>et al.</i> , 2006) |

A total of 2048 mtDNA sequences from African populations and 448 of Native American were compiled, of geographical relevance as previously stated. The haplogroup frequencies were calculated by direct counting as well.

Data collected from the literature was organized under a criterion relating to geographic region, as seen in Tables 3 and 4. All analyses that included our data and data from literature considered only the mtDNA region shared by all sequences. Diversity indices were calculated for comparison populations also using the software *DnaSP* v.5. Accordingly, we calculated number of shared haplotypes within populations or the number of different haplotypes (K), for a corrected calculation of haplotype diversity (H), Mean Number Pairwise Differences (MNPD) and the number of segregating sites (S) (Nei, 1987).

For comparative purposes, pairwise  $F_{ST}$  genetic distances between populations were calculated by *Arlequin* v.3.5 software (currently version 3.5.2.2; Excoffier and Lischer 2010), resorting to data exported from *DnaSP*.

Phylogenetic networks were constructed using the *NETWORK*® software (version 5.0.1.1, Fluxus Technology Ltd 2004-2019) by applying consecutively the method of the reduced median algorithm and the median-joining algorithm (Bandelt *et al.*, 1995; Bandelt *et al.*, 1999), in order to minimize reticulation. In all networks, the positions 16193.xC, 309.xC, 315.xC and 573.xC were excluded due to high mutation rates and to not being considered for haplogroup discrimination. This software helped investigate genetic relationships by checking for haplotype sharing within the studied population and between other related populations.

To have a more clear visual of the pairwise genetic distances based on the calculated  $F_{ST}$  values, a MultiDimensional Scaling (MDS) analysis was performed, with a two-dimensional representation spatial plan, included in the StatSoft, Inc. (2007) program STATISTICA (data analysis software system), ver.8.0 ([www.statsoft.com](http://www.statsoft.com)).

## 4. Results and Discussion

## 4.1 mtDNA diversity in Palenque

In the present study, 81 mtDNA control region samples were identified and its haplotypes deposited in the EMPOP mtDNA database. More detailed information on the haplotypes and corresponding haplogroups attributed to each sample can be obtained in the appendix.

The lineages found were classified according to their continental ancestry within African (91,4%), and Native American (8,6%).

Haplogroup distribution and frequencies within Palenque population and in the two sub-samples of "Palenque rural" (*Palenqueros* living and working in the community) and "Palenque urbano" (*Palenqueros* working and living in Cartagena during the week, and having the family house in the community of Palenque) are presented in Figure 11 and Table 5.

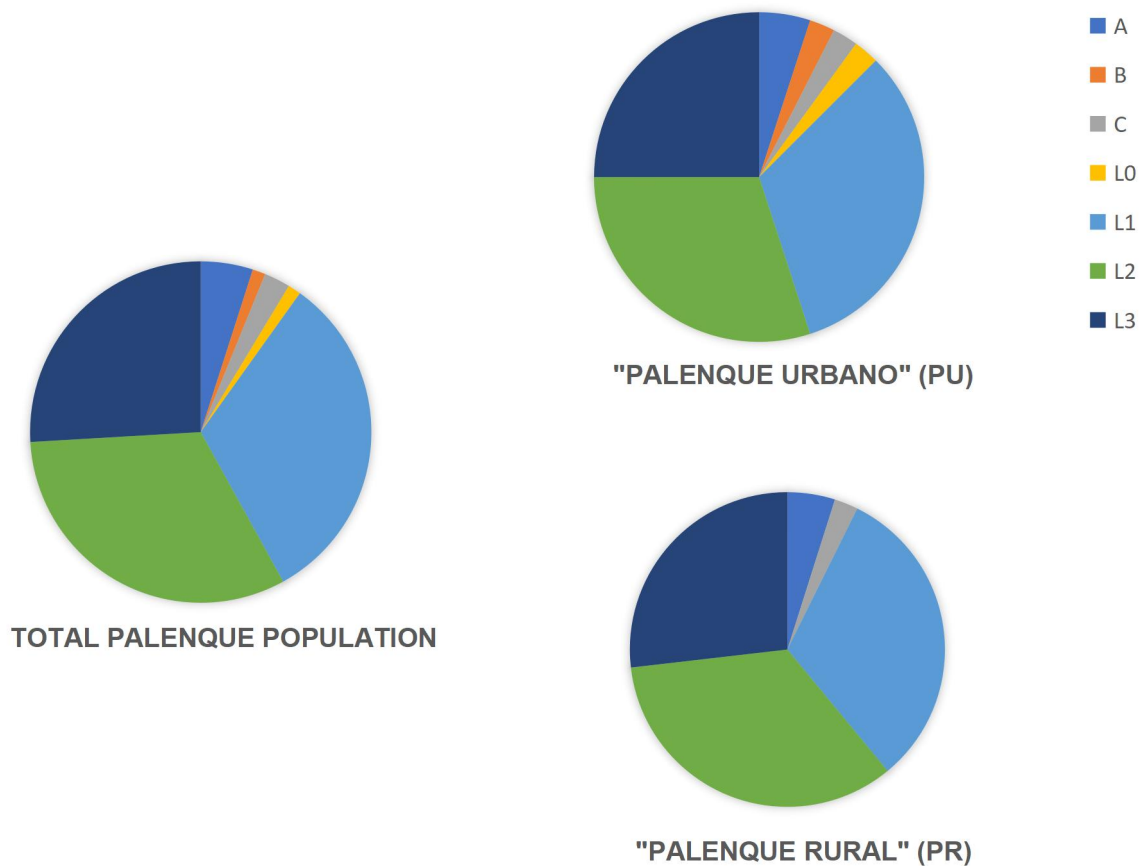


Figure 11 –Haplogroup distribution within Palenque population in comparison to the geographical subgroups in Palenque de San Basilio.

*Table 5 – Haplogroup frequencies in Palenque discriminated according to sampling region.*

| Haplogroup | Region/ Population |                |                |
|------------|--------------------|----------------|----------------|
|            | Palenque total     | Rural Palenque | Urban Palenque |
| A          | 0.049              | 0.049          | 0.050          |
| B          | 0.012              | 0.000          | 0.025          |
| C          | 0.025              | 0.024          | 0.025          |
| L0         | 0.012              | 0.000          | 0.025          |
| L1         | 0.321              | 0.317          | 0.325          |
| L2         | 0.321              | 0.341          | 0.300          |
| L3         | 0.259              | 0.268          | 0.250          |
| N          | 81                 | 41             | 40             |

There were 7 main haplogroups found in our collected samples. The majority of African maternal lineages in Palenque belonged to haplogroups L1 (32%) and L2 (32%) followed by haplogroup L3 (26%). At a lower percentage, representing approximately 1% of the total, there was 1 sample belonging to haplogroup L0. Additionally, with a smaller yet significant contribution, the Native American maternal lineages, with a total of 7 haplotypes in the Palenque population, were distributed among haplogroups A (5%), C (2,5%) and B (1%). These results are according to our expectations, since mtDNA haplotypes found in sub-Saharan Africa, as previously stated, are typically situated in the L branch of mtDNA phylogenetic tree and the studied population's History was of an afro-descendant population. Furthermore, since this population is located in the American continent, the presence of Amerind haplotypes within the sample was also expectable.

In Colombia there was abundant admixture with Europeans, Africans and Natives. Because of that fact, it would be expectable to find other contributions beyond African and Native American in this sample. The absence of European input could be explained by the fact that Palenque is located in a region that until recently was quite secluded near Cartagena, called Montes de María. For this reason, the slaves took refuge in this region. Additionally, this city was founded by then fugitive slaves that ventured into the woods of the Amazonian rainforest. Some of these former slaves were thought to have been born on American soil. This could explain findings of MSY

lineages from European descent but the same was very unlikely to happen for mtDNA lineages because of a sex-biased mating effect in which, at the time of the probable foundation of Palenque, European women would not mate with non-European men. The presence of native haplogroups in the sample, even though this was an isolated population, is explained by the fact that the location of Palenque is surrounded by native communities, even now. That could make it easier for the input of Amerind women into the community.

Even though the Palenque de San Basilio's population as a whole displays no sub-structure, there is still a fairly perceptible difference between both geographic regions sampled regarding the haplogroups found. It is also logical that Urban Palenque would have more haplogroup diversity than Rural Palenque, due to the simple fact that it is more likely to be accessible to outsiders and therefore it could be more prone to new genetic inflow, over time. Haplogroup frequencies on both geographic groups can be observed in Figure 11 and in Table 5.

According to a more specific division, a total of 22 different subclades were found in the 81 samples collectively, being that 17 are of African ancestry. Regarding the latter, within 74 genotyped samples with African ancestry, the most common subclade was L2a1+16189(+16192) representing 27% of the samples allocated (Appendix).

However, if the division of haplogroups was less profound, there was only 12 different sub-haplogroups throughout the sample. In this case, the most common sub-haplogroup of African ancestry in our sample was L2a with 32% of the samples. Also frequent in our sample are L1b (26%), L3d (15%) and L3e (12%). All these groups are quite common in admixed populations throughout the American continent (Figure 12).

The 7 Native American haplotypes found in the Palenque population are distributed among sub-haplogroups A2, B2 and C1, being A2 still divided in 3 other subclades, as perceived by the phylogenetic Network illustrated in Figure 13.

All through the American Continent, as is perceived in Figure 12, the haplogroup distribution of African ancestry (L0 to L3) is very diverse and most, if not all, main sub-haplogroups are present, unlike the distribution present in our population (Salas *et al.*, 2004). As previously mentioned, this population suffered a strong genetic drift, due to the founder effect caused by their forced departure from the main population. It is also worth noting that this community remained isolated and with a low

effective, for a long period, and the existence of a certain degree of inbreeding cannot be neglected.

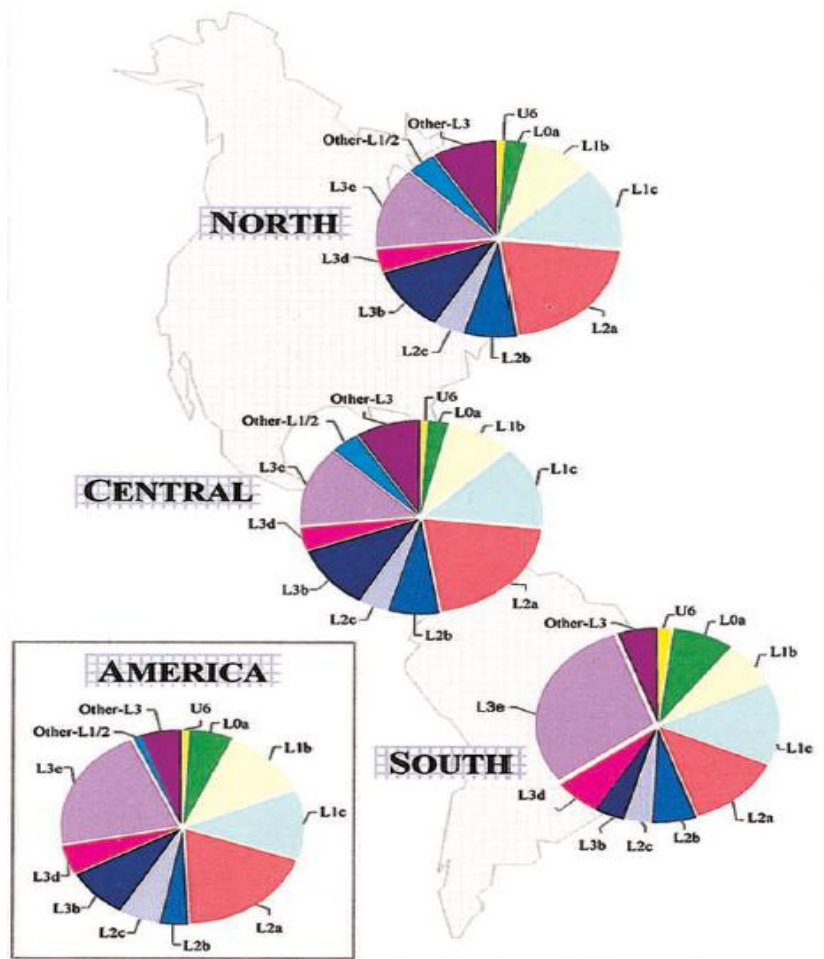
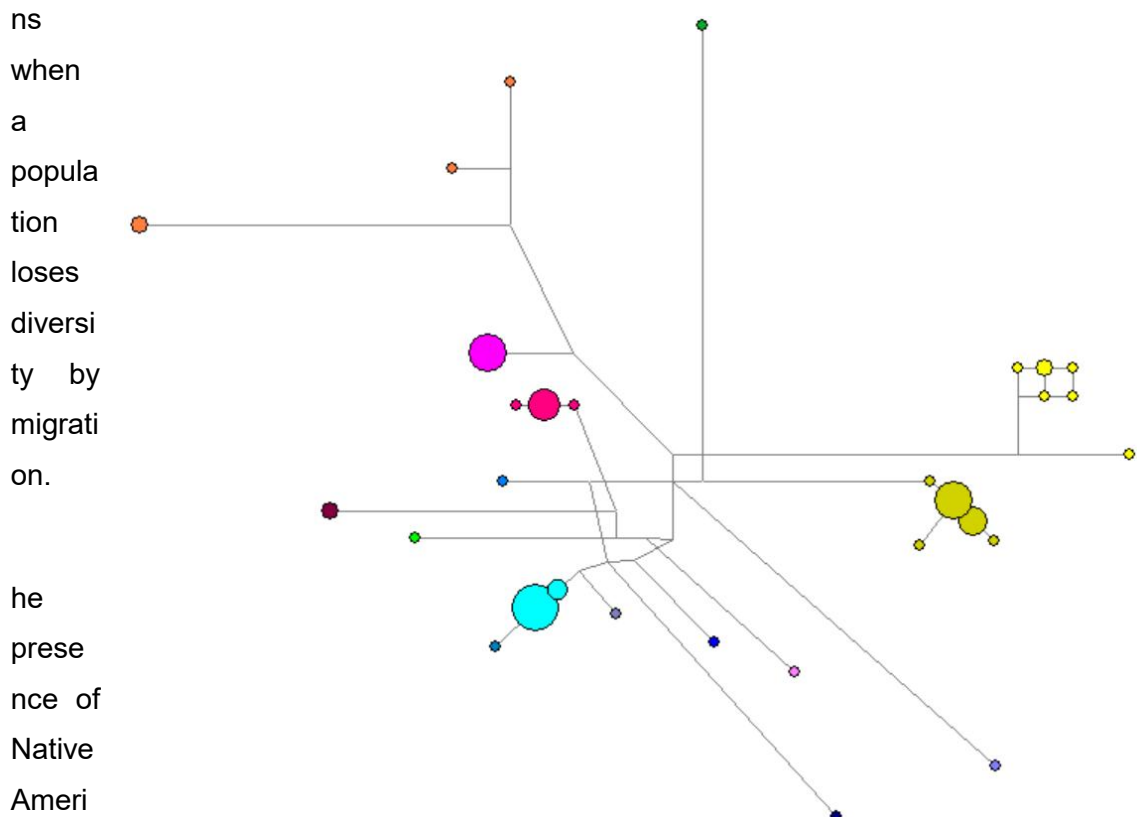


Figure 12 – Haplogroup frequencies of recent African ancestry present in the Americas (adapted from Salas et al., 2004).

The haplotype variation of these haplogroups is represented in a median-joining network (Figure 13), showing a distinct separation of the haplotypes and a clear founder effect. This is confirmed by the fact that the number of Haplogroups found is similar to the number of different haplotypes, indicating a repetition, caused by strong founder effects.

Regarding Native American haplotypes, it seems to have been single entries, as an explanation for the low number of haplotypes but with significant difference between them, with at least 5 different haplotypes.

Within African ancestral haplotypes, there are 5 that have the highest frequency of repetition within the sample. This is a result of an event of founding effect, which happens when a population loses diversity by migration.



The presence of Native American haplotypes proves this predominantly African community did not remain completely isolated from neighbouring population. Otherwise, admixture could not have been present.

- A
- B
- C
- L0
- L1b1a
- L1c3a
- L2a1
- L2a1+16189+(16192)
- L2a1+143+@16309
- L2a1c3b2
- L2a1f3
- L2b1a
- L2d+16129
- L3d1a1a
- L3e1d
- L3f1b+16292

Figure 13 – Network of Palenque's population. It shows a clear separation between its major haplogroups found. It is also shown that there is low haplotype diversity

The network in figure 13 clearly shows that this population has a high haplogroup diversity but low haplotype diversity, which once again was a result of strong founder effects.

As shown in figure 14, the most commonly found haplogroup in the northern region of Colombia, in Amerindian populations, is haplogroup A, which could be the reason for this haplogroup to be the most frequent in our studied population. However, we cannot be completely certain due to a small number of haplotypes allocated in Native American haplogroups, present in the studied population. Additionally, in Colombia's total Native American population, haplogroups A and C both form the majority of haplogroups found in that region, which is also consistent with our distribution.

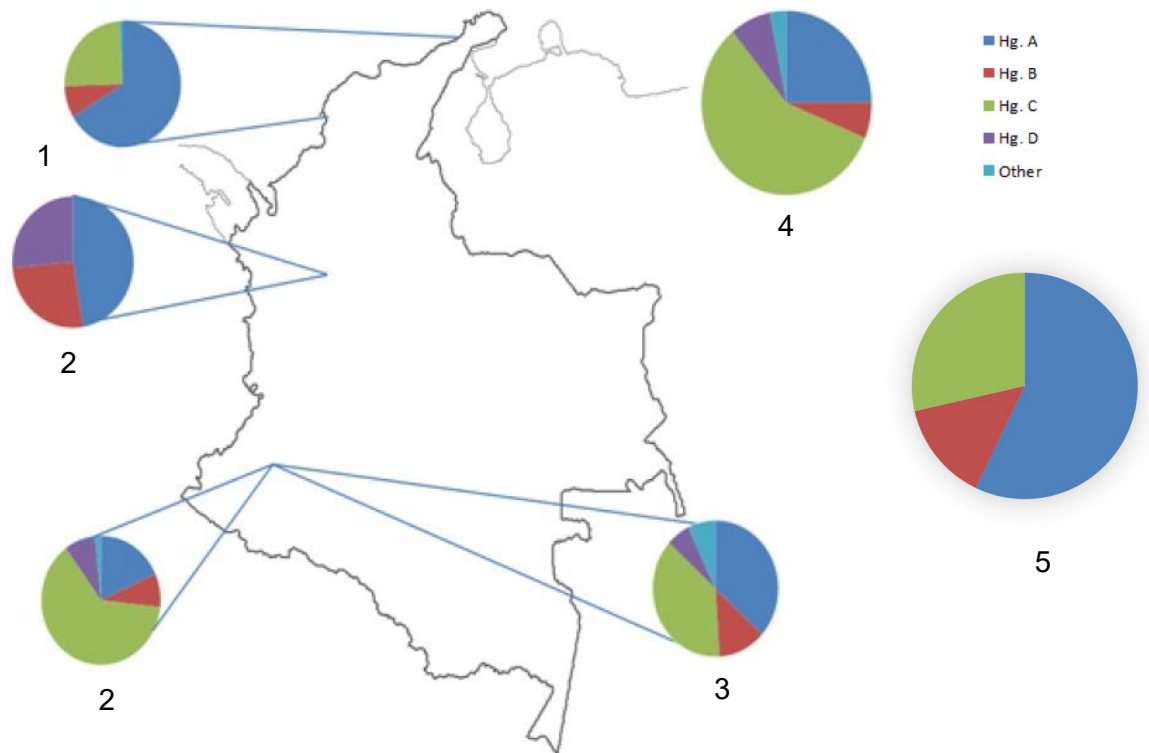


Figure 14 – Amerindian Haplogroups found in Colombia and Palenque (Adapted from Xavier *et al.*, 2015). 1 (Melton *et al.*, 2007); 2 (Xavier *et al.*, 2015); 3 (Salas *et al.*, 2008); 4 (Torres *et al.*, 2006); 5 - Present Study.

## 4.2 Genetic Diversity Comparisons

The result of Mean Number of Pairwise Differences (MNPD) for the population of Palenque was rather surprising. There was a higher number of differences than expected in most pairwise comparisons (MNPD of 16.907) (Table 6).

The Mismatch Distribution (MMD) of haplotypes in the Palenque sample shows variability as to pairwise differences within the population (Figure 15).

Point heteroplasmy was observed in 6 samples altogether, and all heteroplasmies occurred at positions considered major hotspots in the mtDNA genome (16189Y, 16192Y, 16355Y and 263R).

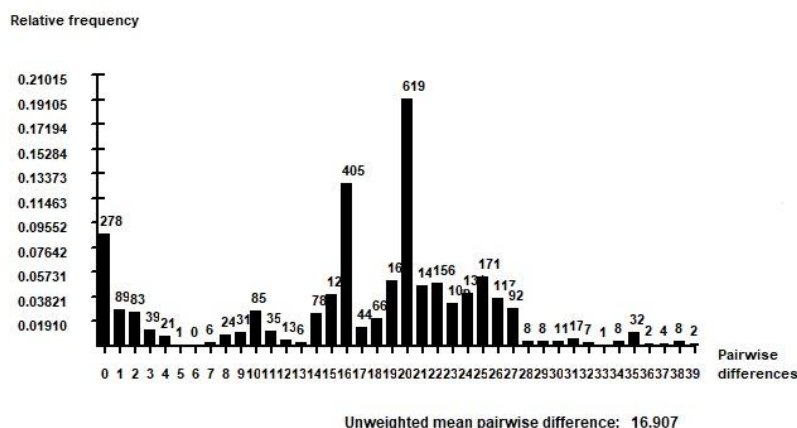


Figure 15 –MMD of Palenque population based on MNPD values.

The calculated haplotype diversity ( $H$ ) found in Palenque was of  $0.9225 \pm 0.0162$  (Table 6). When comparing to other populations, it is perceptible the lower value of  $H$  found in Palenque. This most likely means that this population suffered strong founder effects and/or genetic drift. 30 unique haplotypes were found in this sample.

Population comparisons were made according to the geographic regions because the interest of this study was to investigate African ancestry. As such, despite a miscegenation of two ancestral populations is present in Palenque, only samples with African ancestry and samples from that same geographic region were accounted for comparison purposes. Results can be observed in Tables 6-8.

Most of the data collected were typed for shorter HVR segments than the present study, with the maximum resolution of range possible for all mtDNA sequences

being from 16,030 to 16,391 for the HVRI and a range from 73 to 322 for the HVRII, which was used for population comparisons in the present study.

As expected, values of haplotype diversity (H) obtained differed when using the complete CR or when based only on shorter versions. However, as previously mentioned, some works used for comparison in this study did not sequence all the CR and therefore only a lower resolution could be obtained. This also draws the attention to the fact that future works with high resolution methods are needed for forensic analysis and will also enable a better understanding of the populations' structures and histories.

*Table 6 - Haplotype Diversity (H) and Mean Number of Pairwise Differences of Palenque vs other populations (positions between 16030-16391 and 73-322 were used for minimal resolution).<sup>i</sup>*

| Population <sup>ii</sup> | HVRI+HVRII (610 bp) |         |                 |          | CR (1124 bp) |                 |       |
|--------------------------|---------------------|---------|-----------------|----------|--------------|-----------------|-------|
|                          | N                   | K/S     | H + sd.         | MNPD     | K/S          | H + sd.         | MNPD  |
| Palenque <sup>PS</sup>   | 81                  | 29/78   | 0.9099 ± 0.0180 | 13.32377 | 30/79        | 0.9225 ± 0.0162 | 16.97 |
| Tunisia <sup>1</sup>     | 64                  | 54/105  | 0.9931 ± 0.0047 | 11.74901 | -            | -               | -     |
| Morocco <sup>1</sup>     | 56                  | 46/82   | 0.9838 ± 0.0104 | 9.59481  | -            | -               | -     |
| Morocco <sup>2</sup>     | 508                 | 362/187 | 0.9942 ± 0.0013 | 9.76066  | -            | -               | -     |
| Kenya <sup>3</sup>       | 100                 | 90/124  | 0.9972 ± 0.0021 | 15.35010 | 93/130       | -               | 15.31 |
| Uganda <sup>4</sup>      | 261                 | 184/160 | 0.9951 ± 0.0013 | 15.98344 | -            | -               | -     |
| Guinea <sup>5</sup>      | 79                  | 66/91   | 0.9942 ± 0.0034 | 13.84745 | -            | -               | -     |
| Ghana <sup>6</sup>       | 192                 | 117/111 | 0.9905 ± 0.0020 | 12.88416 | 127/-        | 0.987           | 15.55 |
| Angola <sup>7</sup>      | 365                 | 172/132 | 0.9914 ± 0.0010 | 16.50048 | -            | -               | -     |
| Angola <sup>8</sup>      | 42                  | 36/79   | 0.9919 ± 0.0074 | 16.13124 | -            | -               | -     |

<sup>i</sup> Note: N is the number of individuals, K is the number of different haplotypes, S is the number of segregating sites, H is the haplotypic diversity and MNPD stands for Mean Number of Pairwise Differences. Sd is the standard deviation associated with the haplotypic diversity.

<sup>ii</sup> PS stands for present study; 1 (Turchi *et al.*, 2009) 2 (Aboukhalid *et al.*, 2013); 3 (Brandstätter *et al.*, 2004); 4 (Gomes *et al.*, 2015); 5 (Carvalho *et al.*, 2011); 6 (Fendt, Röck, *et al.*, 2012); 7 (Coelho *et al.*, 2009); 8 (Plaza *et al.*, 2004); 9 (Cerezo *et al.*, 2016); 10 (Fendt, Huber, *et al.*, 2012).

|                         |     |         |                 |          |   |   |   |
|-------------------------|-----|---------|-----------------|----------|---|---|---|
| Angola <sup>9</sup>     | 105 | 85/110  | 0.9943 ± 0.0026 | 16.93553 | - | - | - |
| Angola <sup>10</sup>    | 129 | 36/81   | 0.8769 ± 0.0226 | 10.27859 | - | - | - |
| Mozambique <sup>9</sup> | 146 | 103/109 | 0.9891 ± 0.0030 | 16.45508 | - | - | - |

The fact that this population has been separated from its main origin population for a few centuries means that diversity changes between them, accumulated over time, are guaranteed to have occurred. In addition, as mentioned before, the present studied population suffered events of founder effect and genetic drift, reason why the percentages of haplotypes found could differ from the original population(s).

To evaluate if there were significant differences in haplotype frequencies, pairwise  $F_{ST}$  genetic distances were calculated between African lineages in Palenque and samples from African populations, using available data based on HVRI and HVRII sequences (Table 7). For a more thorough understanding of the African maternal origin, as the main objective was to investigate the African origin of these lineages, Amerindian haplotypes were excluded from the remainder of the analysis.

All calculated genetic distances with Palenque were significant ( $P$ -value<0.05), meaning these populations are indeed genetically distant from our subject population, which is expected due to the amount of time passed since the formation of this new population. Nevertheless, the closest values of the comparisons made were from Ghana ( $F_{ST}$  =0.03074), Guinea-Bissau ( $F_{ST}$  =0.03877) and Angola ( $F_{ST}$  =0.03695), as shown in Table 7. This is consistent with the hypothesis of more than one origin, but it does not rule out assumptions made by Ansari-pour *et al.* (2016), given that the Congo region is geographically close to all of these locations. Therefore, it is difficult to ascertain which theory is closer to the truth, being most likely that both conjugated are the most likely scenario.

As expected, highly overall significant differences were observed between Palenque and North African populations (Tunisia and Morocco) as well as a Khoisan group from South-west Africa (Angola4). This population seems to diverge from the rest of African populations as well.

*Table 7* – Matrix of pairwise  $F_{ST}$  genetic distances (below the diagonal) and corresponding P-values (above diagonal) obtained for 10100 permutations, for mtDNA control region, between 74 Palenque samples (with African ancestry) and the 12 other African populations found in the literature.

|            | Tunisia | Morocco            | Morocco            | Kenya              | Uganda             | Guinea             | Ghana              | Angola1            | Angola2            | Angola3            | Angola4            | Mozambique         | Palenque           |
|------------|---------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Tunisia    | —       | 0.17701+<br>0.0036 | 0.05851+<br>0.0024 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 |
| Morocco    | 0.00453 | —                  | 0.29314+<br>0.0045 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 |
| Morocco    | 0.00489 | 0.00111            | —                  | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 |
| Kenya      | 0.05601 | 0.08290            | 0.10584            | —                  | 0.01426+<br>0.0012 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00050+<br>0.0002 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 |
| Uganda     | 0.05663 | 0.08154            | 0.10538            | 0.00800            | —                  | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00129+<br>0.0003 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 |
| Guinea     | 0.04518 | 0.08691            | 0.10098            | 0.05302            | 0.03971            | —                  | 0.01673+<br>0.0011 | 0.00000+<br>0.0000 | 0.03326+<br>0.0017 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00030+<br>0.0002 |
| Ghana      | 0.06236 | 0.10703            | 0.12279            | 0.07541            | 0.06336            | 0.01047            | —                  | 0.00000+<br>0.0000 | 0.02406+<br>0.0013 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00010+<br>0.0001 |
| Angola1    | 0.08008 | 0.11132            | 0.14283            | 0.03802            | 0.03011            | 0.03753            | 0.05056            | —                  | 0.08544+<br>0.0027 | 0.04802+<br>0.0017 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 |
| Angola2    | 0.07033 | 0.11845            | 0.14900            | 0.04047            | 0.02831            | 0.01579            | 0.01567            | 0.00860            | —                  | 0.11494+<br>0.0030 | 0.00000+<br>0.0000 | 0.00495+<br>0.0008 | 0.00396+<br>0.0006 |
| Angola3    | 0.08801 | 0.12431            | 0.16529            | 0.04843            | 0.03892            | 0.04336            | 0.05726            | 0.00527            | 0.00961            | —                  | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 |
| Angola4    | 0.37253 | 0.40387            | 0.42297            | 0.27822            | 0.25405            | 0.30886            | 0.31313            | 0.21963            | 0.28326            | 0.27585            | —                  | 0.00000+<br>0.0000 | 0.00000+<br>0.0000 |
| Mozambique | 0.11286 | 0.14642            | 0.19256            | 0.04306            | 0.03737            | 0.07548            | 0.09547            | 0.02146            | 0.03338            | 0.04156            | 0.23532            | —                  | 0.00000+<br>0.0000 |
| Palenque   | 0.10564 | 0.15839            | 0.17957            | 0.10751            | 0.08646            | 0.03877            | 0.03074            | 0.06651            | 0.03695            | 0.07506            | 0.33723            | 0.10566            | —                  |

Note: Non-significant values coloured in red. Lowest yet significant values (against Palenque) coloured in orange

The MultiDimensional Scaling (MDS) plot of pairwise  $F_{ST}$  matrix (depicted in Figure 16) shows a clear genetic drift from African populations Palenque population seems to show a clear separation from its ancestral African populations. More data on ancestral and present populations is most likely required to completely remove doubts about Palenque' population origins.

It is known that the origin of the slaves brought to America were Bantu, as well as the cultural and linguistic influences present in Palenque, reason why the Khoisan population from Angola was excluded from further MDS calculations (Figure S1). Also, due to the high  $F_{ST}$  distance values between this population and the remaining populations.

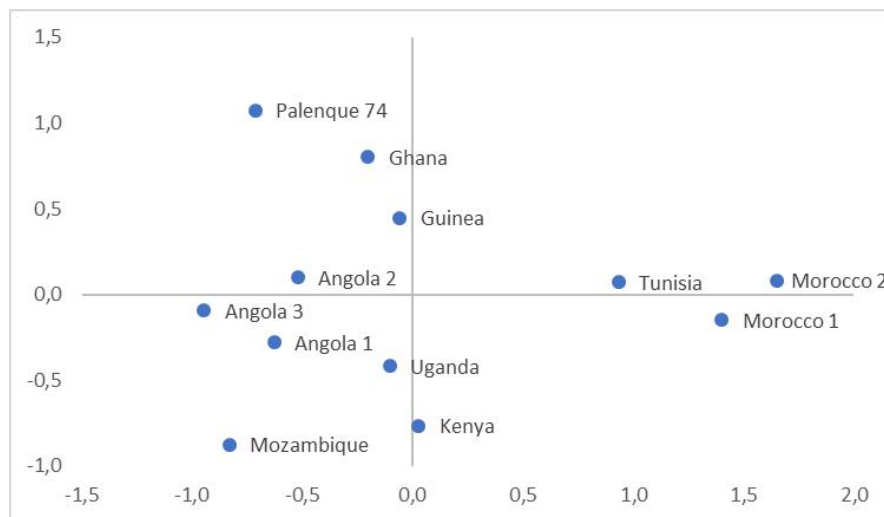


Figure 166 – MDS plot based on the  $F_{ST}$  genetic distances, between the 13 populations analysed, using the 74 samples from Palenque with African ancestry. Angola Khoi-san population was excluded from this MDS.

Ansari-Pour *et al.* (2016) were not able to determine the exact location of origin for the maternal lineages, which could be from the sole use of genetic distance as mtDNA analysis, and in cases of a high genetic drift such as this population suffered is comprehensible the lack of a palpable conclusion. Still, he suspected that it would be the same as the Y chromosome, which he determined to be from the Congo region, as a confirmation to the linguistic studies previously performed, which puts the language of Palenque closer to Bantu language family (Ansari-pour *et al.*, 2016; Schwegler, 2016).

However, in contrary to what is described by Ansari *et al.*, the data shown by the MDS (Figure 16) seems to give evidence of a multi-origin for the population of Palenque.

### 4.3 Phylogeography of lineages from Palenque

To further confirm the previous conclusions, we conducted a more detailed analysis resorting to phylogenetic networks. A table with haplogroup frequencies (Table 8) was also constructed in order to once again verify the presence of a genetic drift in Palenque's population.

For clarity purposes, networks for each haplogroup were calculated separately (Figures 17-21). In order to avoid reticulation, only lineages, from the literature, belonging to the same haplogroup as the samples studied were considered.

Table 8 – African Haplogroup Frequencies found in 5 different regions of Africa and in the Palenque population.

| Haplogroup      | African Regions <sup>iii</sup> |      |              |            |            | Palenque   |
|-----------------|--------------------------------|------|--------------|------------|------------|------------|
|                 | North                          | East | Central West | South West | South East |            |
| <i>Non-L</i>    | 83%                            | 6%   | 2%           | 0%         | 0%         | 0%         |
| <i>L0a</i>      | 1%                             | 19%  | 7%           | 16%        | 25%        | 1%         |
| <i>L0d</i>      | 0%                             | 0%   | 0%           | 4%         | 6%         | 0%         |
| <i>L0</i>       | 1%                             | 3%   | 0%           | 0%         | 0%         | 0%         |
| <i>L1-other</i> | 0%                             | 0%   | 0%           | 1%         | 1%         | 0%         |
| <i>L1b</i>      | 7%                             | 0%   | 8%           | 4%         | 1%         | <b>26%</b> |
| <i>L1c</i>      | 0%                             | 3%   | 4%           | 20%        | 9%         | <b>10%</b> |
| <i>L2a</i>      | 2%                             | 14%  | 17%          | 11%        | 32%        | <b>32%</b> |
| <i>L2b</i>      | 0%                             | 1%   | 4%           | 3%         | 1%         | 1%         |
| <i>L2c</i>      | 0%                             | 0%   | 2%           | 3%         | 1%         | 0%         |

<sup>iii</sup> Note: North region corresponds to Tunisian and Moroccan populations; East region refers to Kenya and Ugandan populations; Central West corresponds to Guinea-Bissau and Ghana; South West refers to Angolan populations; and the South East region concerns Mozambique.

|                 |            |            |            |            |            |            |
|-----------------|------------|------------|------------|------------|------------|------------|
| <i>L2-other</i> | 1%         | 1%         | 2%         | 1%         | 2%         | 1%         |
| <i>L3b</i>      | 1%         | 8%         | 10%        | 3%         | 2%         | 0%         |
| <i>L3d</i>      | 1%         | 3%         | 8%         | 5%         | 5%         | <b>15%</b> |
| <i>L3e</i>      | 3%         | 4%         | 20%        | 20%        | 13%        | <b>12%</b> |
| <i>L3f</i>      | 1%         | 3%         | 13%        | 7%         | 1%         | 1%         |
| <i>L3-other</i> | 0%         | 16%        | 1%         | 1%         | 1%         | 0%         |
| <i>L4-6</i>     | 0%         | 20%        | 1%         | 0%         | 1%         | 0%         |
| <b>N</b>        | <b>628</b> | <b>361</b> | <b>263</b> | <b>471</b> | <b>147</b> | <b>74</b>  |

## Haplogroup L1 - L1b

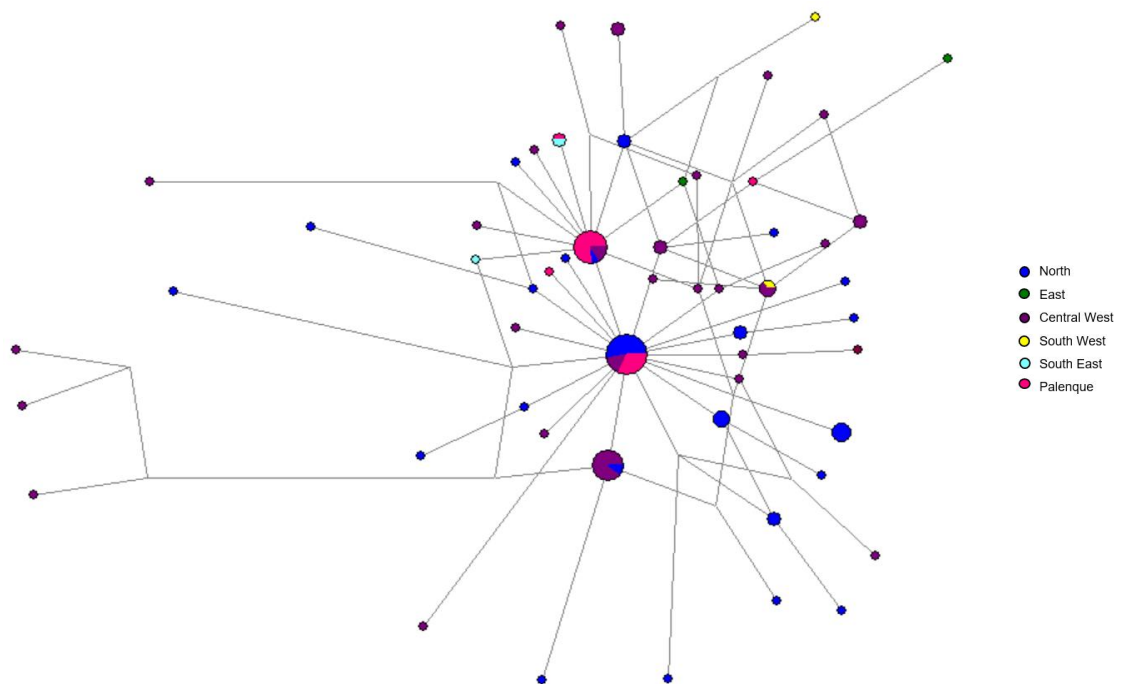


Figure 17 – Network of sub-Haplogroup L1b.

Southeast and Southwest African regions have a low frequency of L1b haplotypes (Table 8). The Network constructed for this haplogroup confirms that haplotypes belonging to L1b found in Palenque most likely came from Central West Africa.

It is also known that no slaves came from the north of the Saharan desert, so the most likely scenario, for this particular haplogroup, to have a match of haplotypes with North Africa, is that the lineages found, in both cases, originally came from the Central West Africa. Given the history of these populations, the lineages found in the North most likely came from below the Saharan desert since North and Western Africa are known to be interconnected through trading networks conducted by desert, nomadic, people (Moors in the extreme west and Tuareg elsewhere), which is more plausible because there were slaves natives to that area.

## Haplogroup L1 - L1c



Figure 18 – Network of sub-Haplogroup L1c.

L1c has the highest frequency in the South West, followed by South East Africa (Table 8).

The network in Figure 18, clearly shows our samples are much closer related to those from the West Africa (Central and South). In fact, Palenque shares haplotypes with samples from Ghana, which was a major slavery traffic hotspot in the 16<sup>th</sup> century. Furthermore, it can be seen the total absence of connection with the East Africa, which

is consistent with previous expectations. It is believed that Palenque was formed around the end of the 16<sup>th</sup> century and that slaves from the East were not trafficked to the Americas until the beginning of the 18<sup>th</sup> century.

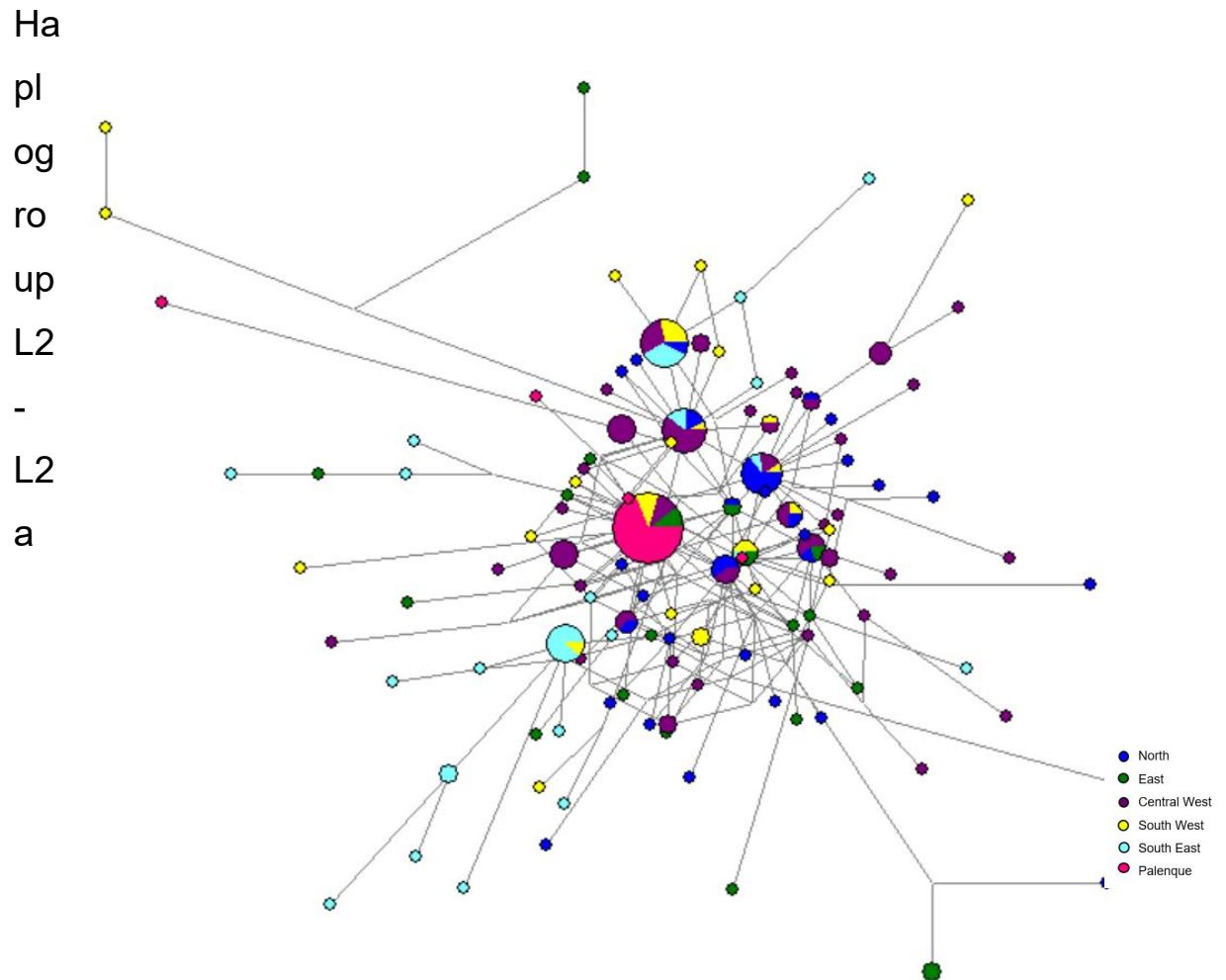


Figure 19 - Network of sub-Haplogroup L2a1.

The most common subclade of African ancestry in our sample was L2a with 32% of the samples. This subclade is widely spread in sub-Saharan Africa and has the highest frequency in the South East region (Table 8). However, as the Network in

figure 19 demonstrates, the lineages from Palenque correlate more with the West African region.

Due to high levels of reticulation in the network, it is difficult to ascertain origin of some lineages. Nevertheless, approximately 20 haplotypes appear together, sharing the circle with other populations as well, in the centre of a star-like reticulation, indicating this could have been an ancestral haplotype from which other haplotypes in this network, that are present in Africa but not in Palenque, evolved. Also worth noting is the other haplotypes present in Palenque are not shared with other African populations, indicating that they may have emerged after the formation of Palenque.

Once again, in Table 8, it is shown that L2a has a high frequency in the South East, but Palenque Lineages seem to be clearly genetically distant from lineages found in this region, again confirming a mostly Western origin of the founding slaves of Palenque, which is also consistent with data available about the slave trade traffic.

### Haplogroup L3 - L3d

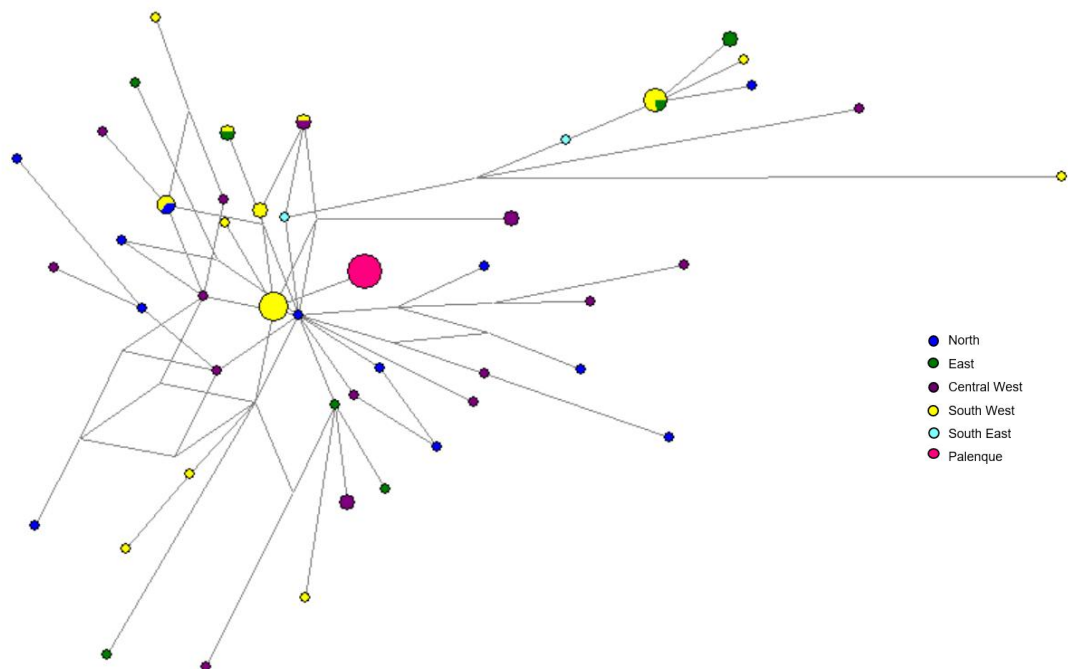


Figure 20 - Network of sub-Haplogroup L3d.

In the network constructed with data relating to sub-haplogroup L3d, shown in figure 20, the visualization of a founder effect in Palenque is clear.

All Palenque haplotypes are constricted to a single sphere, meaning the haplotypes are all equal. Furthermore, these shared haplotypes are all genetically closer related to those from the South West, suggesting that their ancestral haplotype most likely originated there.

Despite L3d being more frequent in Central West Africa than other African regions (Table 8), the network clearly shows that our samples allocated to that haplogroup came from South-western populations. This also indicates genetic drift, which caused the frequency elevation in the population.

## Haplogroup L3 - L3e

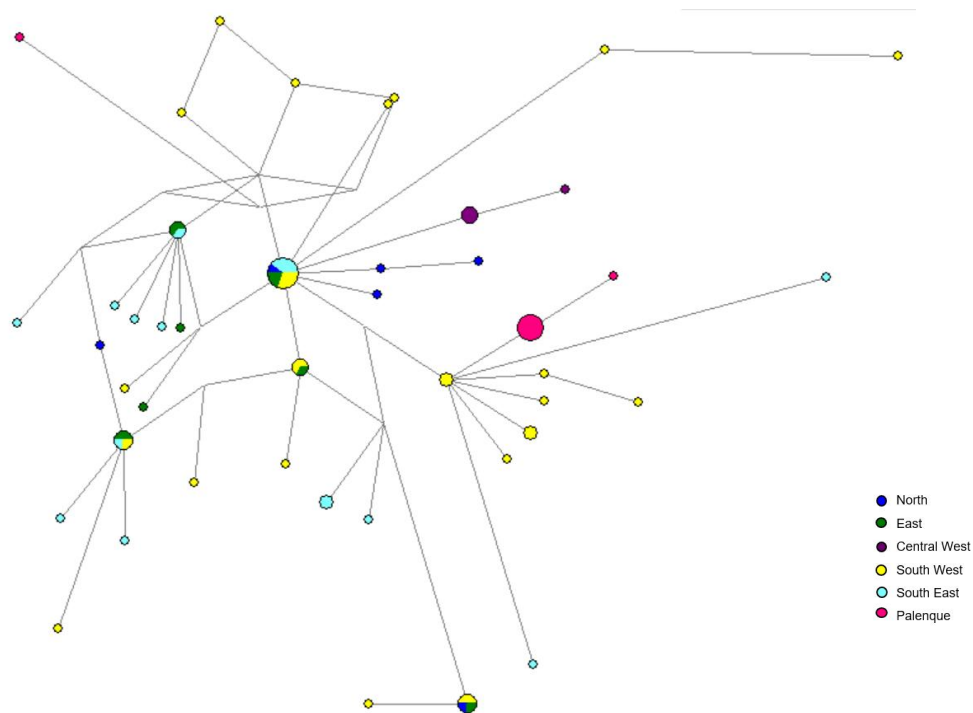


Figure 21 - Network of sub-Haplogroup L3e.

As in previous networks, the network constructed regarding haplogroup L3e (Figure 21) shows a clear founder effect and consequently genetic drift.

But unlike the previous network, the Palenque individuals belonging to haplogroup L3e do not all share the same haplotype. One of the individuals, displaying a different haplotype, offshoots from the rest of the Palenque individuals, suggesting that the mutation originating this different haplotype might have happened after the formation of the Palenque population. A second individual with a different haplotype can be found on the other end of the network. This haplotype seems to have diverged from a different subset of individuals than the rest of the Palenque haplotypes.

Even so, the fraction of our population belonging to the L3 haplogroup most likely originated in a subset of South West African populations.

Frequencies of L3e, shown in Table 8, demonstrate that this haplogroup is more frequent in the west, which is consistent with findings of the network in Figure 21. It is also frequent in the East but as previously mentioned, Palenque was founded before Eastern slaves were brought to the Americas. Although more recent migrations to Palenque would have been possible, all haplotypes from Palenque are distant from those in South East Africa.

## 5. Conclusions

This study provided a deeper and more detailed understanding of the genetic composition of Palenque de San Basilio. The mtDNA haplotypes present in Palenque, in the available sampling, were characterized, thus furthering the investigation upon the maternal ancestry of Palenque de San Basilio's population. As a result, 81 mtDNA haplotypes were deposited in EMPOP mtDNA Forensic Database and their haplogroup classification were subsequently confirmed.

The vast majority of this population's maternal founding haplogroups were determined to be of African ancestry. Additionally, a pattern of admixture is present within the population, as expected. However, this admixture is solely from ancestral Native-American and African populations, with no input from European females. This result, which is also found in indigenous populations, was also expected and it is explained here by not only the secluded location of this population as well as this city's foundation by runaway slaves that ventured into the woods of the Montes de Maria mountains. This probably meant there would not be much miscegenation with Natives nor Europeans.

Unlike the hypothesis previously postulated by Ansari-Pour and collaborators, it is perceived in this study a possibility of more than one regional origin for this population's maternal lineages. Nevertheless, both central west and south west regions of Africa were found to be somewhat closely related to this study's population.

The two Bolivar Department sub-populations studied are, to some extent, differentiated, dividing into subgroups, probably as a consequence of strong genetic drift effects and small population sizes, a pattern common to isolated populations such as this one. This, however, is uncertain, due to a small number of unique haplotypes and significant haplotype sharing frequency.

Further sampling may be needed to confirm these conclusions given the low number of individuals sampled for this study, and maybe even broaden the analyzed mtDNA region, as well as Y-chromosome data should be analysed and combined with this data.

The differences found between African-Ancestral and Palenque populations and the pattern of restricted haplotypes makes the genetic characterization of these groups of extreme relevance for forensic genetic studies, due to their uniqueness.

## References

- Aboukhalid, R. *et al.* (2013) 'Mitochondrial DNA control region variation from samples of the Moroccan population', *International Journal of Legal Medicine*, 127(4), pp. 757–759. doi: 10.1007/s00414-012-0813-4.
- de Acosta, J. *et al.* (2002) *Natural and Moral History of the Indies*. Duke University Press. doi: 10.1192/bjp.111.479.1009-a.
- Alberts, B. *et al.* (2014) *Molecular Biology of the Cell*. Sixth edit. W. W. Norton & Company. doi: 10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.
- Almécija, S. *et al.* (2013) 'The femur of *Orrorin tugenensis* exhibits morphometric affinities with both Miocene apes and later hominins', *Nature Communications*, 4(1), p. 2888. doi: 10.1038/ncomms3888.
- Almeida Mendes, A. de (2008). The foundations of the system: A reassessment of the slave trade to the Spanish Americas in the sixteenth and seventeenth centuries. In D. Eltis & D. Richardson (Eds.), *Extending the frontiers. Essays on the new transatlantic slave trade database* (pp. 63-94). Yale University Press.
- Anastasiou, E. and Mitchell, P. D. (2013) 'Evolutionary anthropology and genes: Investigating the genetics of human evolution from excavated skeletal remains', *Gene*. Elsevier B.V., 528(1), pp. 27–32. doi: 10.1016/j.gene.2013.06.011.
- Anderson, S. *et al.* (1981) 'Sequencing and organization of the human mitochondrial genome', *Nature Publishing Group*.
- Andrews, R. M. *et al.* (1999) 'Reanalysis and revision of the cambridge reference sequence for human mitochondrial DNA [5]', *Nature Genetics*, 23(2), p. 147. doi: 10.1038/13779.
- Ansari-pour, N. *et al.* (2016) 'Palenque de San Basilio in Colombia : genetic data support an oral history of a paternal ancestry in Congo', *Royal Society Publishing*. doi: <http://dx.doi.org/10.1098/rspb.2015.2980>.
- Auton, A. *et al.* (2015) 'A global reference for human genetic variation', *Nature*, 526(7571), pp. 68–74. doi: 10.1038/nature15393.
- Avise, J. (1991) '10 Unorthodox Perspectives On Evolution Prompted By Comparative Population Genetic Findings On Mitochondrial Dna', *Annual Review of Genetics*, 25(1),

pp. 45–69. doi: 10.1146/annurev.genet.25.1.45.

Bandelt, H. J., Forster, P., Sykes, B. C., & Richards, M. B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics*, 141(2), 743–753.

Bandelt, H.J., Forster, P. & Rohl, A. (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37-48

Bandelt, H.J. & Parson, W. (2008) Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *International Journal of Legal Medicine*, 122, 11-21.

Bandelt, H. J. *et al.* (2014) 'The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies', *Journal of Human Genetics*, 59(2), pp. 66–77. doi: 10.1038/jhg.2013.120.

Bateson, W., *et al.* (1905). Experimental studies in the physiology of heredity. Reports to the Evolution Committee of the Royal Society 2, 1–55, 80–99

Battaglia, V. *et al.* (2013) 'The First Peopling of South America: New Evidence from Y-Chromosome Haplogroup Q', *PLoS ONE*, 8(8). doi: 10.1371/journal.pone.0071390.

Behar, D. M. *et al.* (2012) 'A “copernican” reassessment of the human mitochondrial DNA tree from its root', *American Journal of Human Genetics*, 90(4), pp. 675–684. doi: 10.1016/j.ajhg.2012.03.002.

Brandstätter, A. *et al.* (2004) 'Mitochondrial DNA control region sequences from Nairobi (Kenya): Inferring phylogenetic parameters for the establishment of a forensic database', *International Journal of Legal Medicine*, 118(5), pp. 294–306. doi: 10.1007/s00414-004-0466-z.

Budowle, B. *et al.* (2003) 'Forensics and Mitochondrial DNA: Applications, Debates, and Foundations', *Annual Review of Genomics and Human Genetics*, 4(1), pp. 119–141. doi: 10.1146/annurev.genom.4.070802.110352.

Burr, S. P., Pezet, M. and Chinnery, P. F. (2018) 'Mitochondrial DNA Heteroplasmy and Purifying Selection in the Mammalian Female Germ Line', *Development Growth and Differentiation*, 60(1), pp. 21–32. doi: 10.1111/dgd.12420.

Carracedo, A., Bär, W., Lincoln, P., Mayr, W., Morling, N., Olaisen, B., Schneider, P., Budowle, B., Brinkmann, B., Gill, P., Holland, M., Tully, G. & Wilson, M. (2000) DNA Commission of the International Society for Forensic Genetics: guidelines for

mitochondrial DNA typing. *Forensic Science International*, 110, 79-85.

Carvalho, M. *et al.* (2011) 'Paternal and maternal lineages in Guinea-Bissau population', *Forensic Science International: Genetics*, 5(2), pp. 114–116. doi: 10.1016/j.fsigen.2010.10.007.

Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. (1994) *The History and Geography of Human Genes*. Princeton University Press.

Cavalli-Sforza, L. L. and Feldman, M. W. (2003) 'The application of molecular genetic approaches to the study of human evolution', *Nature Genetics*, 33(3S), pp. 266–275. doi: 10.1038/ng1113.

Cerezo, M. *et al.* (2016) 'Comprehensive Analysis of Pan-African Mitochondrial DNA Variation Provides New Insights into Continental Variation and Demography', *Journal of Genetics and Genomics*. Elsevier Ltd, 43(3), pp. 133–143. doi: 10.1016/j.jgg.2015.09.005.

Chakravarti, A. 1999. Population genetics—making sense out of sequence. *Nature genetics*, 21, 56-60.

Charlesworth, B. (2009) 'Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation', *Nature Reviews Genetics*, 10(3), pp. 195–205. doi: 10.1038/nrg2526.

Coelho, M. *et al.* (2009) 'On the edge of Bantu expansions: MtDNA, y chromosome and lactase persistence genetic variation in southwestern Angola', *BMC Evolutionary Biology*, 9(1), pp. 1–18. doi: 10.1186/1471-2148-9-80.

Correns CG., (1900). Regel über das Verhalten der Nachkommenschaft der Rassenbastarde. *Berichte der Deutschen Botanischen Gesellschaft*. 18:158–68.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 1st edn. John Murray, London.

Diamond, J. (1991) *The rise and fall of the third chimpanzee*.

Dobzhansky, Th. (1937) *Genetics and the Origin of Species*. Columbia Univ. Press, New York); 2nd Ed., 1941; 3rd Ed., 1951.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Therer, T. & Wilson, A. (2011) *Geneious*. V5.4 (ed.).

- Excoffier, L. and Lischer, H. E. L. (2010) 'Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows', *Molecular Ecology Resources*, 10(3), pp. 564–567. doi: 10.1111/j.1755-0998.2010.02847.x.
- Fendt, L., Huber, G., *et al.* (2012) 'Mitochondrial DNA control region data from indigenous Angolan Khoe-San lineages', *Forensic Science International: Genetics*, 6(5), pp. 662–663. doi: 10.1016/j.fsigen.2012.02.010.
- Fendt, L., Röck, A., *et al.* (2012) 'MtDNA diversity of Ghana: A forensic and phylogeographic view', *Forensic Science International: Genetics*, 6(2), pp. 244–249. doi: 10.1016/j.fsigen.2011.05.011.
- Fisher, R.A. & Bennett, J.H. (1930) *The Genetical Theory of Natural Selection: A Complete Variorum Edition*. Oxford, University Press.
- Flegontov, P. *et al.* (2019) 'Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America', *Nature*. Springer US, 570(7760), pp. 236–240. doi: 10.1038/s41586-019-1251-y.
- Fregel, R. and Delgado, S. (2011) 'HaploSearch: A tool for haplotype-sequence two-way transformation', *Mitochondrion*. Elsevier B.V. and Mitochondria Research Society, 11(2), pp. 366–367. doi: 10.1016/j.mito.2010.11.001.
- Gomes, V. *et al.* (2010) 'Digging deeper into East African human y chromosome lineages', *Human Genetics*, 127(5), pp. 603–613. doi: 10.1007/s00439-010-0808-5.
- Gomes, V. *et al.* (2015) 'Mosaic maternal ancestry in the Great Lakes region of East Africa', *Human Genetics*. Springer Berlin Heidelberg. doi: 10.1007/s00439-015-1583-0.
- Goodwin, W., Linacre, A. and Hadi, S. (2007) *An introduction to forensic genetics*, Wiley. John Wiley & Sons Ltd.
- Green, R. E. *et al.* (2010) 'A draft sequence of the neandertal genome', *Science*, 328(5979), pp. 710–722. doi: 10.1126/science.1188021.
- Griffiths, A. J. F. *et al.* (1987) 'An Introduction to Genetic Analysis', *JAMA: The Journal of the American Medical Association*, 257(16), p. 2223. doi: 10.1001/jama.1987.03390160109041.
- Hagelberg, E. *et al.* (2008) 'A genetic perspective on the origins and dispersal of the Austronesians: Mitochondrial DNA variation from Madagascar to Easter Island', *Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics*, pp.

356–375. doi: 10.4324/9780203926789.

Haldane, J.B.S. (1932) *The causes of evolution*, by J.B.S. Haldane. London, New York [etc.]: Longmans, Green and co.

Hartl, D. and Clark, A. (1997) *Principles of Population Genetics*. 3<sup>rd</sup> ed. Sinauer Associates, inc. Publishers.

Hedrick, P. 2009. *Population Genetics and Ecology*. The Princeton guide to ecology. Princeton Univ Pr.

Jobling, M. (2013) *Human Evolutionary Genetics*. 2<sup>nd</sup> Ed, *Garland Science*. doi: 10.1016/b0-08-043076-7/03088-6.

Karafet, T. M. *et al.* (1999) 'Ancestral Asian source(s) of new world Y-chromosome founder haplotypes', *American Journal of Human Genetics*, 64(3), pp. 817–831. doi: 10.1086/302282.

Klinge, C. M. (2008) 'Estrogenic control of mitochondrial function and biogenesis', *Journal of Cellular Biochemistry*, 105(6), pp. 1342–1351. doi: 10.1002/jcb.21936.

Kloss-Brandstätter, A., *et al.* (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Human Mutation*, 32, 25-32

Lareu, M. *et al.* (1994) Investigation of the STR locus HUMTH01 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. *Forensic science international*, 66, 41-52.

Li, X. *et al.* (2013) 'Targeting mitochondrial reactive oxygen species as novel therapy for inflammatory diseases and cancers', *Journal of Hematology and Oncology*, 6(1), pp. 1–19. doi: 10.1186/1756-8722-6-19.

Lodish, H. *et al.* (2003) 'Molecular Cell Biology', *Freeman, W. H & Company*, (5th ed), p. 973.

Margulis, L. (1981) *Symbiosis in Cell Evolution: Life and Its Environment on the Early Earth*. New York: W. H. Freeman.

Martínez, B. *et al.* (2017) 'Ancestry estimates in afrodescendant population from San Basilio de Palenque, Colombia', *Forensic Science International: Genetics Supplement Series*. Elsevier, 6(August), pp. e224–e225. doi: 10.1016/j.fsigss.2017.09.105.

Melton, P. E. *et al.* (2007) 'Biological relationship between central and South American Chibchan speaking populations: Evidence from mtDNA', *American Journal of Physical*

*Anthropology*, 133(1), pp. 753–770. doi: 10.1002/ajpa.20581.

Mendel G. (1866). Versuche über Pflanzen-Hybriden, Verhandlungen des naturforschenden Vereines, Abhandlungen. Brünn, 4:3–47, English translation at <http://www.mendelweb.org/Mendel.html>

Meyer, S., Weiss, G. and Haeseler, A. Von (1999) 'Pattern of Nucleotide Substitution and Rate Heterogeneity in the hyper variable regions I and II of human mtDNA', *Genetics*, 152, pp. 1103–1110.

Miller, S. A., Dykes, D. D. and Polesky, H. F. (1988) 'A simple salting out procedure for extracting DNA from human nucleated cells', *Nucleic Acids Research*, 16(3), p. 1215. doi: 10.1093/nar/16.3.1215.

Millstein, R.L. & Skipper, R.A. (2006) 'Population Genetics', *The Cambridge Companion to the Philosophy of Biology*, Cambridge University Press, p. 35. doi: 10.1017/CCOL9780521851282.

Mishmar, D. *et al.* (2003) 'Natural selection shaped regional mtDNA variation in humans', *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), pp. 171–176. doi: 10.1073/pnas.0136972100.

Morgan, T. H. (1910) Sex-limited inheritance in *Drosophila*. *Science* 132, 120–122

Morgan, T. H (1911) Random segregation versus coupling in Mendelian inheritance. *Science* 34, 384

Navarrete, M. C. (2012) Palenques: cimarrones y castas en el Caribe colombiano — Sus relaciones sociales (siglo XVII). In Maglia & Schwegler (Eds.), pp. 257-284.

Nei, M. (1987) *Molecular Evolutionary Genetics*. New York: Columbia University Press. doi: 10.1126/science.235.4788.599.

Nei, M. (1995) 'Genetic support for the out-of-Africa theory of human evolution', *Proceedings of the National Academy of Sciences of the United States of America*, 92(15), pp. 6720–6722. doi: 10.1073/pnas.92.15.6720.

Newson, L. A. & Minchin, S. (2004). Slave mortality and African origins: A view from Cartagena, Colombia in the early seventeenth century. *Slavery and Abolition*, 25, 18-43.

Newson, L. A. & Minchin, S. (2007) From capture to sale: The Portuguese slave trade to Spanish South America in the early seventeenth century. Leiden: Brill.

Newson, L. A. (2012) Africans and Luso-Africans in the Portuguese slave trade on the

Upper Guinea coast in the early seventeenth century. *Journal of African History*, 53(1), 1-24

Noguera, M. C. *et al.* (2013) 'Colombia's racial crucible: Y chromosome evidence from six admixed communities in the department of bolivar', *Annals of Human Biology*, 41(5), pp. 453–459. doi: 10.3109/03014460.2013.852244.

O'Rourke, D. H. and Raff, J. A. (2010) 'Review: The Human Genetic History of the Americas: The Final Frontier', *Current Biology*. Elsevier Ltd, 20(4), pp. R202–R207. doi: 10.1016/j.cub.2009.11.051.

Orita, M. *et al.* (1989) 'Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms', *Proceedings of the National Academy of Sciences of the United States of America*, 86(8), pp. 2766–2770. doi: 10.1073/pnas.86.8.2766.

van Oven, M. and Kayser, M. (2009) 'Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation.', *Human mutation*, 30(2), pp. 386–394. doi: 10.1002/humu.20921.

Pakendorf, B. and Stoneking, M. (2005) 'Mitochondrial Dna and Human Evolution', *Annual Review of Genomics & Human Genetics*, 6, pp. 161–165. doi: 10.1146/annurev.genom.6.080604.162249.

Parson, W. *et al.* (2004) EMPOP—the EDNAP mtDNA population database concept for a new generation, high-quality mtDNA database. International Congress Series, 1261, 106-108.

Parson, W. *et al.* (2014) 'DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing', *Forensic Science International: Genetics*, 13, pp. 134–142. doi: 10.1016/j.fsigen.2014.07.010.

Parsons, T. J. *et al.* (1997) 'A high observed substitution rate in the human mitochondrial DNA control region', *Nature Genetics*, 15, pp. 57–61.

Plaza, S. *et al.* (2004) 'Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola', *Human Genetics*, 115(5), pp. 439–447. doi: 10.1007/s00439-004-1164-0.

Posth, C. *et al.* (2018) 'Reconstructing the Deep Population History of Central and South America', *Cell*, 175(5), pp. 1185-1197.e22. doi: 10.1016/j.cell.2018.10.027.

Potter, B. A. *et al.* (2018) A N T H R O P O L O G Y Current evidence allows multiple

models for the peopling of the Americas. *Sci. Adv*, 4(August), pp 1–9.  
<https://doi.org/10.1126/sciadv.aat5473>

R Development Core Team (2011) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Relethford, J. H. and Harding, R. M. (2001) 'Population Genetics of Modern Human Evolution', *Encyclopedia of Life Sciences*, pp. 1–6. doi: 10.1038/npg.els.0001470.

Roberts, A. (2011) *Evolution - The Human Story*. Dorling Kindersley Publishing.

Robin, E. D. and Wong, R. (1988) 'Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells', *Journal of Cellular Physiology*, 136(3), pp. 507–513. doi: 10.1002/jcp.1041360316.

Rojas, W. *et al.* (2010) Genetic Make Up and Structure of Colombian Populations by Means of Uniparental and Biparental DNA Markers. *American Journal of Physical Anthropology* 143: 13–20.

Rozas, J. (2009) 'DNA Sequence Polymorphism Analysis Using DnaSP', *Bioinformatics for DNA Sequence Analysis, Methods in Molecular Biology*, 537, pp. 233–242. doi: 10.1007/978-1-59745-251-9.

Rozas, J. *et al.* (2017) 'DnaSP 6: DNA sequence polymorphism analysis of large data sets', *Molecular Biology and Evolution*, 34(12), pp. 3299–3302. doi: 10.1093/molbev/msx248.

Salas, A. *et al.* (2004) 'The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade', *American Journal of Human Genetics*, 74(3), pp. 454–465. doi: 10.1086/382194.

Salas, A. *et al.* (2008) 'The mtDNA ancestry of admixed Colombian populations', *American Journal of Human Biology*, 20(5), pp. 584–591. doi: 10.1002/ajhb.20783.

Santibanez-Koref, M. *et al.* (2019) 'Assessing mitochondrial heteroplasmy using next generation sequencing: A note of caution', *Mitochondrion*, 46(April 2018), pp. 302–306. doi: 10.1016/j.mito.2018.08.003.

Saraste, M. (1999) 'Oxidative phosphorylation at the fin de siecle', *Science*, 283(5407), pp. 1488–1493. doi: 10.1126/science.283.5407.1488.

Schwegler, A. *et al.* (2011) 'On the Extraordinary Revival of a Creole : Appeared in On the Extraordinary Revival of a Creole : Palenquero ( Colombia )', pp. 153–161.

- Schwegler, A. (2016) 'Combining population genetics (DNA) with historical linguistics: On the African origins of Latin America's black and mulatto populations', *Spanish language and sociolinguistic analysis*, pp. 33–88. doi: 10.1075/ihll.8.02sch.
- Sikora, M. *et al.* (2019) 'The population history of northeastern Siberia since the Pleistocene', *Nature*. Springer US. doi: 10.1038/s41586-019-1279-z.
- Simão, F. *et al.* (2018) 'Defining mtDNA origins and population stratification in Rio de Janeiro', *Forensic Science International: Genetics*, 34(December 2017), pp. 97–104. doi: 10.1016/j.fsigen.2018.02.003.
- Stellwagen, N. C. (2009) 'Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution', *Electrophoresis*, 30(SUPPL. 1), pp. 188–195. doi: 10.1002/elps.200900052.
- Stewart, J. B. and Chinnery, P. F. (2015) 'The dynamics of mitochondrial DNA heteroplasmy: Implications for human health and disease', *Nature Reviews Genetics*. Nature Publishing Group, 16(9), pp. 530–542. doi: 10.1038/nrg3966.
- Stone, A. C. (2019) 'Human lineages in the far north', *Nature*, 570, pp. 170–172.
- Stoneking, M. (2017) *An Introduction to Molecular Anthropology*. Wiley-Blackwell.
- Stoneking, M. and Krause, J. (2011) 'Learning about human population history from ancient and modern genomes', *Nature Reviews Genetics*. Nature Publishing Group, 12(9), pp. 603–614. doi: 10.1038/nrg3029.
- Stringer, C. (2003) 'Out of Ethiopia', *Nature*, 423(June), pp. 692–695.
- Thomas, H (1997). *The slave trade: the story of the Atlantic slave trade, 1440-1870*. New York: Simon & Schuster.
- Tishkoff, S. A. and Verrelli, B. C. (2003) 'PATTERNS OF HUMAN GENETIC DIVERSITY: Implications for Human Evolutionary History and Disease', *Annual Review of Genomics and Human Genetics*, 4(1), pp. 293–340. doi: 10.1146/annurev.genom.4.070802.110226.
- Torres, M. M. *et al.* (2006) 'A revertant of the major founder Native American haplogroup C common in populations from Northern South America', *American Journal of Human Biology*, 18(1), pp. 59–65. doi: 10.1002/ajhb.20461.
- Torrioni, A. *et al.* (1993) Asian affinities and continental radiation of the 4 founding Native-American mtDNAs. *American Journal of Human Genetics*, 53, pp. 563-590.

- Torroni, A. *et al.* (2006) 'Harvesting the fruit of the human mtDNA tree', *Trends in Genetics*, 22(6), pp. 339–345. doi: 10.1016/j.tig.2006.04.001.
- Turchi, C. *et al.* (2009) 'Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: An enrichment of forensic mtDNA databases with Northern Africa data', *Forensic Science International: Genetics*, 3(3), pp. 166–172. doi: 10.1016/j.fsigen.2009.01.014.
- Turner, C.G. (1987) Late Pleistocene and Holocene population history of east Asia based on dental variation. *American Journal of Physical Anthropology*, 73, pp. 305-321.
- Voet, D; Voet J. G. and Pratt C. W. (2013). *Fundamentals of Biochemistry*. John Wiley and Sons, 5th Edition.
- de Vries, H. (1900) *Rev. Gen. Bot.* 12, pp. 257–271.
- Walberg, M. W. and Clayton, D. A. (1981) 'Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA', *Nucleic Acids Research*, 9(20), pp. 5411–5421. doi: 10.1093/nar/9.20.5411.
- Wallace, D. C. (1994) 'Review: Mitochondrial DNA sequence variation in human evolution and disease (mutations/human origins/Leber hereditary optic neuropathy/Alzheimer and Parkinson diseases/aging)', *Proceedings of the National Academy of Sciences*, 91(September), pp. 8739–8746.
- Wang, S. *et al.* (2007) 'Genetic variation and population structure in Native Americans', *PLoS Genetics*, 3(11), pp. 2049–2067. doi: 10.1371/journal.pgen.0030185.
- Wheat, D (2009). *The Afro-Portuguese maritime world and the foundations of Spanish Caribbean society, 1570-1640*. Doctoral Dissertation, Vanderbilt University, Nashville, Tennessee.
- Wheat, D (2011). *The first great waves: African provenance zones for the transatlantic slave trade to Cartagena de Indias, 1570-1640*. *Journal of African History*, 52, 1-22.
- Wright, S. (1931) *Evolution in Mendelian Populations*. *Genetics*, 16, 97-159.
- Xavier, C. *et al.* (2015) 'Admixture and genetic diversity distribution patterns of non-recombining lineages of native american ancestry in colombian populations', *PLoS ONE*, 10(3), pp. 1–13. doi: 10.1371/journal.pone.0120155.

## Web Resources

<http://www.phylotree.org>

<http://www.discoveringbristol.org.uk/slavery/people-involved/enslaved-people/enslaved-africans/transatlantic-slave-trade/>

<https://ich.unesco.org/en/RL/cultural-space-of-palenque-de-san-basilio-00102>

[http://www.understandingslavery.com/index.php-option=com\\_content&view=article&id=369&Itemid=145.html](http://www.understandingslavery.com/index.php-option=com_content&view=article&id=369&Itemid=145.html)

<https://haplosearch.com>

## Annexes

Table S1 – Table of haplotypes found in the population of Palenque

| <i>Sample ID</i> | <i>CR Mutated Positions</i>                                                                                                                                                  | <i>Haplogroup</i>  |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| <i>PR004</i>     | 16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del                              | L1b1a1'4           |
| <i>PR005</i>     | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                              | L2a1+16189+(16192) |
| <i>PR007</i>     | 16189C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G 146C 152C 195C 263G 315.1C                                                                                | L2a1c3b2           |
| <i>PR009</i>     | 16129A 16183C 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16360T 16390A 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del | L1c3a              |
| <i>PR010</i>     | 16111T 16126C 16223T 16290T 16319A 16362C 64T 73G 146C 153G 235G 263G 309.1C 309.2C 315.1C 523del 524del                                                                     | A2aI               |
| <i>PR011</i>     | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                | L1b1a+189          |
| <i>PR013</i>     | 16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del                              | L1b1a1'4           |
| <i>PR018</i>     | 16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del                              | L1b1a1'4           |
| <i>PR024</i>     | 16189C 16192Y 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| <i>PR026</i>     | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                              | L2a1+16189+(16192) |
| <i>PR027</i>     | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                | L1b1a+189          |
| <i>PR033</i>     | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                              | L2a1+16189+(16192) |
| <i>PR038</i>     | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C 73G 152C 182T 185T 195C 228A 247A 263G 309.1C 315.1C 357G 523del 524del                                       | L1b1a7a            |
| <i>PR039</i>     | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                          | L3d1a1a            |
| <i>PR040</i>     | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                | L1b1a+189          |
| <i>PR041</i>     | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                          | L3d1a1a            |
| <i>PR051</i>     | 16223T 16265G 16298C 16319A 16325C 16327T 16519C 73G 249del 263G 290del 291del 309.1C 315.1C 489C                                                                            | C1c3               |
| <i>PR052</i>     | 16111T 16176T 16189Y 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C                                                                                         | L3e1d              |
| <i>PR058</i>     | 16093C 16129A 16189C 16278T 16300G 16311C 16354T 16390A 16399G 16519C 73G 146C 150T 152C 195C 263G 315.1C 456T 524.1A 524.2C                                                 | L2d+16129          |
| <i>PR061</i>     | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                          | L3d1a1a            |
| <i>PR062</i>     | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                          | L3d1a1a            |
| <i>PR068</i>     | 16223T 16278T 16294T 16390A 73G 143A 146C 152C 195C 263G 309.1C 315.1C 523del 524del                                                                                         | L2a1+143+@16309    |
| <i>PR071</i>     | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                | L1b1a+189          |
| <i>PR072</i>     | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                              | L2a1+16189+(16192) |
| <i>PR073</i>     | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                          | L3d1a1a            |
| <i>PR077</i>     | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                              | L2a1+16189+(16192) |
| <i>PR081</i>     | 16189C 16191.1C 16192Y 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                              | L2a1+16189+(16192) |
| <i>PR082</i>     | 16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C                                                                                                | L3e1d              |

|       |                                                                                                                                                                                       |                    |
|-------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| PR085 | 16038.1A 16129A 16183C 16189C 16193.1C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del | L1c3a1b            |
| PR086 | 16189C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                                | L2a1+16189+(16192) |
| PR087 | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                         | L1b1a+189          |
| PR088 | 16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C                                                                                                         | L3e1d              |
| PR090 | 16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                       | L1b1a1'4           |
| PR091 | 16189C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                                | L2a1+16189+(16192) |
| PR093 | 16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C                                                                                                         | L3e1d              |
| PR095 | 16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 315.1C                                                                                                                | L3e1d              |
| PR097 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PR099 | 16092C 16111T 16223T 16290T 16319A 16362C 64T 73G 146C 153G 235G 263G 315.1C 482C 523del 524del                                                                                       | A2+(64)            |
| PR100 | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                                   | L3d1a1a            |
| PR102 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PR103 | 16038.1A 16129A 16183C 16189C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del          | L1c3a1b            |
| PU004 | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                                   | L3d1a1a            |
| PU005 | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                         | L1b1a+189          |
| PU006 | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                                   | L3d1a1a            |
| PU009 | 16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                       | L1b1a1'4           |
| PU011 | 16183C 16189C 16217C 16519C 73G 263G 309.1C 309.2C 315.1C 498del 499A                                                                                                                 | B2d                |
| PU015 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PU016 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PU019 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PU027 | 16129A 16183C 16189C 16193.1C 16215G 16223T 16278T 16294T 16311C 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del                 | L1c3a              |
| PU029 | 16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16319A 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                | L1b1a1'4           |
| PU030 | 16111T 16223T 16290T 16319A 16360T 16362C 89C 97A 106del 107del 108del 109del 110del 111del 146C 198T 235G 263G 309.1C 315.1C                                                         | A2af1a1            |
| PU037 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PU040 | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                                   | L3d1a1a            |
| PU041 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PU042 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PU043 | 16111T 16223T 16290T 16319A 16360T 16362C 89C 97A 106del 107del 108del 109del 110del 111del 146C 198T 235G 263G 309.1C 315.1C                                                         | A2af1a1            |
| PU044 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PU045 | 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 198T 263G 315.1C                                                                                                         | L2a1f3             |
| PU047 | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                                       | L2a1+16189+(16192) |
| PU050 | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                         | L1b1a+189          |

|              |                                                                                                                                                                              |                    |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|
| <i>PU053</i> | 16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C                                                                                                | L3e1d              |
| <i>PU055</i> | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263R 309.1C 315.1C 357G 523del 524del                                | L1b1a+189          |
| <i>PU057</i> | 16223T 16265G 16298C 16319A 16325C 16327T 16519C 73G 249del 263G 290del 291del 309.1C 315.1C 489C                                                                            | C1c3               |
| <i>PU062</i> | 16114A 16129A 16174T 16213A 16223T 16278T 16355T 16362C 16390A 73G 146C 150T 152C 182T 195C 198T 204C 263G 309.1C 315.1C 418T 523del 524del                                  | L2b1a              |
| <i>PU063</i> | 16114A 16191.1C 16192Y 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                              | L2a1               |
| <i>PU065</i> | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                          | L3d1a1a            |
| <i>PU067</i> | 16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C                                                                                                | L3e1d              |
| <i>PU069</i> | 16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C                                                                                                | L3e1d              |
| <i>PU073</i> | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                | L1b1a+189          |
| <i>PU076</i> | 16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del                                                                                                          | L3d1a1a            |
| <i>PU079</i> | 16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del                              | L1b1a1'4           |
| <i>PU080</i> | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                | L1b1a+189          |
| <i>PU081</i> | 16209C 16223T 16292T 16311C 16519C 73G 189G 195C 200G 263G 315.1C                                                                                                            | L3f1b+16292        |
| <i>PU092</i> | 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C                                                                              | L2a1+16189+(16192) |
| <i>PU094</i> | 16129A 16183C 16189C 16193.1C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del | L1c3a1b            |
| <i>PU095</i> | 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16311C 16320T 64T 93G 152C 189G 200G 236C 247A 263G 309.1C 315.1C 523del 524del                        | L0a1a+200          |
| <i>PU097</i> | 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del                                     | L1b1a18            |
| <i>PU098</i> | 16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C                                                                                                | L3e1d              |
| <i>PU100</i> | 16129A 16183C 16189C 16193.1C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del | L1c3a1b            |
| <i>PU103</i> | 16129A 16183C 16189C 16215G 16223T 16278T 16294T 16311C 16355Y 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del          | L1c3a              |

Table S2 – Relative frequency of haplogroups found in separate regions of Palenque.

| HG         | %    | N(PR) | HG         | %    | N(PU) |
|------------|------|-------|------------|------|-------|
| <b>L0A</b> | 0%   | 0     | L0a        | 3%   | 1     |
| <b>L1B</b> | 26%  | 10    | L1b        | 25%  | 9     |
| <b>L1C</b> | 8%   | 3     | L1c        | 11%  | 4     |
| <b>L2A</b> | 34%  | 13    | L2a        | 31%  | 11    |
| <b>L2B</b> | 0%   | 0     | L2b        | 3%   | 1     |
| <b>L2D</b> | 3%   | 1     | L2d        | 0%   | 0     |
| <b>L3D</b> | 16%  | 6     | L3d        | 14%  | 5     |
| <b>L3E</b> | 13%  | 5     | L3e        | 11%  | 4     |
| <b>L3F</b> | 0%   | 0     | L3f        | 3%   | 1     |
| <b>TOT</b> | 100% | 38    | <b>TOT</b> | 100% | 36    |

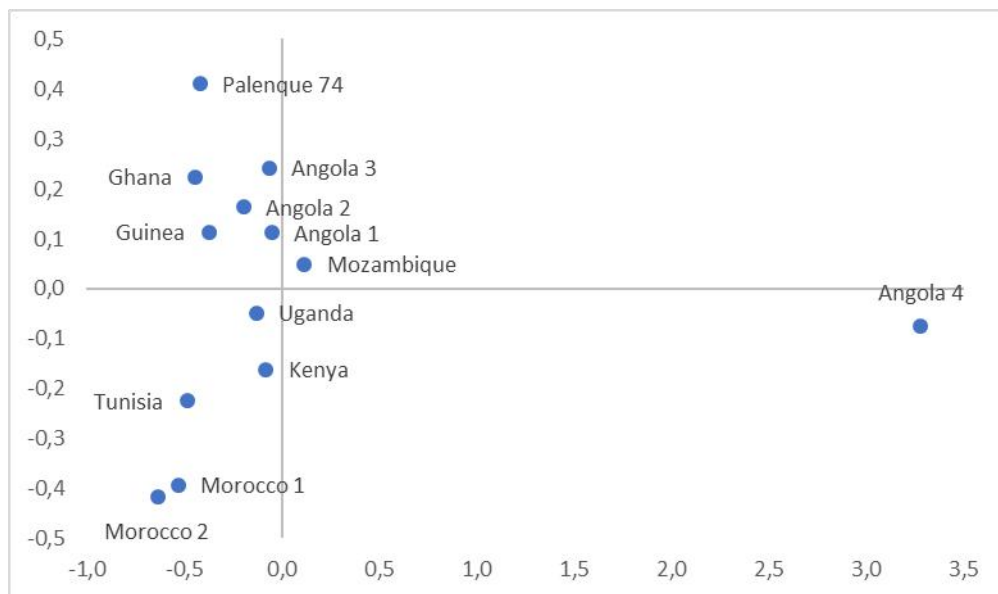


*Figure S1* – MDS plot based on the  $F_{ST}$  genetic distances, between the 13 populations analysed, utilizing the 74 samples from Palenque with African ancestry. Angolan Khoisan population, represented here by Angola 4, is not remotely close, genetically, to any other African population.