FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

Tracing South American Colonization by mtDNA analysis in Colombian Populations

Catarina Gomes Alves Xavier

2012

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

IPATIMUP
Instituto de Patologia e Imunologia Molecular da Universidade do Porto

Tracing South American Colonization by mtDNA Analysis in Colombian Populations

Catarina Gomes Alves Xavier

2012

MSc

Tracing South American Colonization by mtDNA Analysis in Colombian Populations

Catarina Gomes Alves Xavier
Mestrado em Genética Forense
Departamento de Biologia
2012

Orientador
Ana Goios Borges de Almeida, PhD, IPATIMUP

Coorientador
Maria Leonor Rodrigues Sousa Botelho Gusmão, PhD, IPATIMUP

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

IPATIMUP
Instituto de Patologia e Imunologia Molecular da Universidade do Porto

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

O Presidente do Júri,

Porto, _____/_____/_____

AGRADECIMENTOS

À minha orientadora, Ana Goios, por toda a dedicação e amizade que me deu durante este ano. Obrigado pela ajuda, pelos conselhos e por me teres deixado levar este trabalho de uma forma independente, ainda que bem supervisionada. Muito obrigado pela ajuda nesta fase final, nunca me poderei esquecer da ajuda que me deste a ver opções para o futuro. Tive muita sorte e estive mesmo "muito bem entregue".

À minha co-orientadora, Leonor Gusmão, muito obrigado pela disponibilidade e por arranjares sempre um tempo para mim. Muito obrigado pelos incentivos e por me fazeres acreditar em continuar a fazer aquilo de que gosto.

Ao Juan José Builles, ao médico José Miguel Ospino do Hospital de San Juan de Dios do Município de Segovia (Antioquia – Colômbia) e ao Laboratorio Genes Ltda pela colheita e cedência das amostras, sem as quais não seria possível realizar este estudo.

À dupla Verónica e Rodrigo que me acompanhou no meu percurso laboratorial, obrigado pelas dicas e pelos sorrisos! Muito obrigado pela disponibilidade e pela ajuda. Espero que muitas alegrias venham a caminho!

Ao Professor António Amorim, por me ter aceitado neste Mestrado e pelas aulas *brainstorming* que me ajudaram a ver a ciência de outra perspectiva e para além do óbvio. Cresci muito nestes dois últimos anos e agradeço-lhe muito ter-me permitido esta experiência.

Obrigado ao grupo de Genética Populacional do IPATIMUP por me ter recebido em Casa e por me ter facultado todas as condições para realizar um bom trabalho num óptimo ambiente.

Agradeço muito à minha família, por me ter educado para ser uma pessoa com valores de humildade e generosidade. Obrigado pela insistência e pelos incentivos a que continue a minha educação. Sei que acreditam em mim e espero nunca vos desiludir!

Aos meus amigos de sempre, os que conto pelos dedos da mão. Obrigado pela vossa presença mesmo quando eu não dou conta pelos meses a passar. Obrigado ao meu melhor amigo, por me acompanhares sempre. Obrigado pela ajuda, pela amizade e por tudo o que fica por dizer. Obrigado aos meus amigos de Licenciatura e de Mestrado, obrigado pelas vossas opiniões sinceras e pelos momentos de

descontração, um agradecimento especial às meninas de Mestrado no grupo de Genética Populacional pelos momentos incríveis e pelos conselhos!

Esta tese é dedicada aos meus avôs, que não tive oportunidade de conhecer: António Joaquim Xavier e José Maria Gomes Alves.

Muito obrigado por todas as oportunidades e alegrias!

ABSTRACT

America was the last continent to be colonized by mankind during the late Pleistocene, and even though a consensual opinion was achieved concerning the entrance through Beringia and the main Asian origins for Native Americans, there are still doubts on the dispersion routes taken within the continent.

The colonization of South America is a widely debated issue that raises many doubts considering the number and the relevance of migrations. Despite being the main entrance point into South America, Colombia's genetic composition is still far from fully determined. Intending to provide new knowledge on the South American colonization routes, 98 mtDNA control region sequences from two Colombian regions were analysed (Antioquia representing a Chibchan related group and Cauca constituted by several ethnic and linguistic groups). Lineage markers (such as mtDNA) allow tracing back the history of populations because they are transmitted without recombination to the descendants; these lineages are grouped into haplogroups, distinguished by specific polymorphisms that tend to be geographically restricted. There have been described 4 Pan-American mtDNA haplogroups – A, B, C and D- and another lineage restricted to the northern subcontinent, X.

The vast majority of haplogroups found in both Colombian regions are typically Native American. Our results show that while in the Antioquia region, the Emberá population presents a reduced number of haplotypes, all belonging to haplogroups A, B and D, the Cauca region is more diverse and has a significant percentage of C haplogroup lineages. When dividing the Cauca group into smaller speaking groups it is visible that they are distinct and behave as small populations that have suffered evolutionary forces along time such as genetic drift and bottlenecks. When comparing with other populations from literature, there is a notable proximity between Chibchan speaking groups, whereas non-Chibchan remain differentiated. Regarding a geographic separation, there is no visible substructure. Instead, distinct patterns are visible both in northern and southern populations within Colombia which may result from distinct ancient routes.

Finally, the new data on mtDNA in Native American Colombian populations made available through this work has allowed to increase the number of sequences included in EMPOP©, a forensic database that is useful in forensic casework analyses.

RESUMO

O continente Americano foi o último a ser colonizado pelo ser humano durante o fim do Pleistoceno. Ainda que um consenso tenha sido encontrado relativamente à entrada nas Américas pelo Estreito de Bering (antiga Beringia) e quanto às origens asiáticas dos Nativo-Americanos, permanecem muitas dúvidas quanto às rotas de dispersão dentro do continente.

O processo de colonização da América do Sul ainda levanta muitas dúvidas quanto ao número e à importância das migrações. Apesar da Colômbia ser considerada o principal ponto de entrada neste subcontinente, a caracterização genética deste país ainda está por definir por completo. Com o intuito de contribuir para uma melhor compreensão das rotas de colonização da América do Sul, foram analisadas 98 sequências da Região Controlo do DNA mitocondrial de duas regiões colombianas (Antioquia composta por uma população Embéra-Chamí cuja língua está relacionada com os dialetos Chibcha e Cauca composta por indivíduos de vários grupos línguisticos e étnicos). A análise de marcadores de linhagem (como o DNA mitocondrial) permite construir um estudo histórico das populações uma vez que são transmitidos sem alterações à descendência, excetuando a ocorrência de uma mutação. As linhagens agrupam-se em haplogrupos pela partilha de certos polimorfismos e estes tendem a ser geograficamente restritos. No continente Americano existem 4 linhagens de DNA mitocondrial ubíquas – A, B, C e D – e uma linhagem que se encontra restrita na América do Norte, X.

A grande maioria dos haplótipos encontrados nas regiões analisadas foram classsificados como pertencendo a linhagens Pan-Americanas (A, B, C e D). Os nossos resultados mostram que, enquanto a população Embéra que habita a região de Antioquia apresenta um número de haplótipos reduzido pertencentes aos haplogrupos A, B e D, a região de Cauca apresenta maior diversidade e uma prevalência do haplogrupo C. No entanto, ao dividir Cauca em dois grupos com associações linguísticas, observa-se que estes são claramente distintos e se comportam como pequenas populações sujeitas a efeitos de deriva genética e bottlenecks. Ao comparar os nossos grupos com populações da literatura, verifica-se uma clara aproximação dos grupos de língua Chibcha enquanto que outros grupos linguísticos se mantêm distanciados. Quanto a um critério geográfico observa-se uma forte diferenciação entre regiões americanas e dentro da Colômbia verifica-se ainda uma forte separação entre regiões a Norte e a Sul tendo em conta as frequências de haplogrupos, que poderão ter sido uma consequência de diferentes rotas de migração.

Finalmente, os novos dados relativos ao mtDNA em populações Colombianas de Nativo-Americanos disponibilizados a partir deste trabalho permitiram aumentar o número de sequências incluídas numa base de dados forense EMPOP©, e são úteis em investigações forenses.

KEYWORDS

South America, Colombia, colonization routes, maternal lineages.

PALAVRAS-CHAVE

América do Sul, Colômbia, rotas de colonização, linhagens maternas.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVIATIONS

DNA	DeoxyriboNucleic Acid
RFLP	Restriction Fragment Length Polymorphism
PCR	Polymerase Chain Reaction
STR	Short Tandem Repeats
SNP	Single Nucleotide Polymorphism
InDel	Insertion Deletion
mtDNA	Mitochondrial DNA
bp	Base Pair
kb	Kilo Bases
CR	Control Region
D-Loop	Displacement Loop
HVR	Hyper-Variable Region
VR	Variable Regions
BP	Before Present
YBP	Years Before Present
MSY	Male Specific Region of the Y Chromosome
DEL	Deletion
H	Haplotype Diversity
Π	Nucleotide Diversity
S	Number of Segregating Sites
MDS	MultiDimensional Scaling
PS	Present Study

PREFACE

Colombia is undeniably a country of contrasting traditions and cultures. The rich ethnic and linguistic patrimony is evident when studying this country. All this diversity appeared as a consequence of several settlements of Amerinds from different ethnic groups belonging to various linguistic families during the pre-Hispanic period. But this diversity had other contributing factors such as the colonization by Europeans during the late 15th century and sequential settlements of Spaniards and other migrants with African and Jewish ancestry.

This country is also of extreme importance in the history of the colonization of the Americas by the first Americans since it is the major entrance point in the South American subcontinent. In fact, scholars suggest that the first Americans entered South America through the Isthmus of Panama and the Colombian territory, and some of them also indicate that the dispersion routes could have been subdivided within Colombia.

The aim of this study is to understand deeper the migratory movements that ended with the colonization of the entire subcontinent by characterizing the maternal lineages of two regions within Colombia. The acknowledgment of the genetic composition of the native groups is relevant not only in anthropology, allowing further insights on possible migrations throughout the continent, but also in a forensic perspective because it allows forensic databasing of some populations which are difficult to reach.

In order to achieve a better understanding of these issues an introduction was designed to encompass the major subjects. It starts by a short explanation on Population Genetics and is followed by a short description on the major advances in techniques related to the study of human diversity. Afterwards mitochondrial DNA is described as well as its major characteristics and applications to justify the choice of this marker for this study. In the second chapter of the introduction there is an explanation of the entrance in the Americas and the major models considered through the last century. Additionally, a description is made of the most accepted theories on the Entrance in South America, followed by a historical, demographical, genetic and linguistic characterization of Colombia.

1.INTRODUCTION

1.1 Population Genetics

Population genetics is a field that started to define its bases around the beginning of the 20th century under the work of several biologists and mathematicians (Fisher & Bennett, 1930; Wright, 1931; Haldane, 1932) that were able to join the Darwinian continuous evolution with the Mendelian laws of inheritance and then initiating the period of "Evolutionary Synthesis" and the dawn of population genetics (Millstein & Skipper, 2006).

The genetic composition of populations can be estimated by gathering frequencies of different genotypes. The pursuit to understand how these frequencies vary and what is the meaning for these variations occurring among populations, are some of the main objectives of population genetics. This scientific field intends to define mathematically how these variations occur in time and also to comprehend how they are shaped by evolutionary causes such as mutation, migration, genetic drift and selection (Millstein & Skipper, 2006; Griffiths *et al.*, 2008).

Population genetics has many applications in other scientific fields such as anthropology, forensic genetics or even disease studies. Concerning anthropology, population genetics allows a deeper understanding of population behaviour along time and of the demographic events and evolutionary forces that these populations endured. It is also possible to investigate phylogenies and therefore foresee the populations' ancestry and history. Applications to forensic genetics fall into the genotyping of individuals for forensic purposes such as paternity tests or individual identification in various situations such as crime scenes or accidents. A DNA profile obtained must be compared with a reference population and a probability of match for a population must be calculated. To do so, large databases are needed with the markers currently used in the forensic field and containing numerous individuals from different populations, which are normally the product of population genetics research.

1.1.1 Genetic Variation and Human Diversity

Genetic variation is found among individuals within and between populations as a consequence of various factors (Griffiths *et al.*, 2008). The only source of variation is mutation but differentiation between populations can be achieved by recombination, migration, genetic drift and selection.

Although mutation is not the faster way of increasing diversity in a population (mutation rate is relatively low), it is intrinsically the main cause of variation as it leads to the creation of new alleles. Recombination is a faster way to increase variation because of

the numerous combinations of alleles that can be formed during meiosis. Migration acts by introducing new alleles in a population and altering their genotypic frequencies and it can be faster than mutation. Selection can also alter the genotypic frequencies by increasing or decreasing the frequency of a specific genotype that has a higher or lower fitness. Genetic drift on the other hand works randomly, leading either to the fixation or the elimination of a certain genotype (Griffiths *et al.*, 2008).

Over the last century, scholars started to search for the genetic reasons behind the observed variation among human populations seen in the numerous phenotypic characteristics. The first genetic studies on human variability used classical genetic markers such as ABO blood groups or other immunological assays that revealed variation between populations and noted that it was frequent to have numerous variants of a single protein (Hirszfeld & Hirszfeld, 1919; Pauling *et al.*, 1949; Cavalli-Sforza & Feldman, 2003). The advent of new technologies that allowed a direct analysis of the DNA molecule such as Restriction Fragment Length Polymorphisms (RFLP) that used restriction enzymes to identify polymorphic patterns in individuals (Cavalli-Sforza & Feldman, 2003) led to a new scale of resolution in forensic and anthropologic studies.

In the late 80s and 90s techniques were developed that revolutionized the study of DNA such as the Polymerase Chain Reaction (PCR) and automated DNA sequencing. These techniques permitted the analysis of genome variation and were integrated in anthropologic and forensic studies (Cavalli-Sforza & Feldman, 2003). New polymorphisms have been described through the use of the latest techniques such as Short Tandem Repeats (STR), Single Nucleotide Polymorphisms (SNP), Insertions and Deletions (InDels) among others (Cavalli-Sforza & Feldman, 2003). These techniques also allowed exploring different types of variation, particularly in lineage markers such as the Mitochondrial DNA (mtDNA) and the Y-chromosome (Saiki *et al.*, 1985; Amorim, 2007).

Over the last years a wide range of studies were performed with the aim of unveiling the demographic processes that led to the present human genetic patterns worldwide, under evolutionary interpretations. Non-recombining markers like mtDNA and Y-chromosome have been extensively used to unveil the history of populations because of their ability to identify lineages and therefore perceive the major demographic events that the populations underwent such as migrations and bottlenecks.

1.1.2 Mitochondrial DNA

Mitochondria are energy producing organelles located in the cytoplasm of Eukaryotic cells, which present a double membrane structure that is a reminiscence of their origin. These organelles are thought to have originated from prokaryotic cells that were introduced in larger and anaerobic cells, cohabiting in an endossymbiotic manner: anaerobic cells beneficiated from the aerobic or photosynthetic abilities of the prokaryote and prokaryotes found resources easier inside these larger cells. This theory first initiated by Ivan Wallin in 1920 and later formalized by Lynn Margulis was named Endossymbiotic Theory and has extensive supporting evidence in the structure of the mitochondrion itself (double membrane like in prokaryotic cells), the circular genome found within mitochondria and the fission replication processes (Margulis, 1981).

Even though mitochondria are essentially involved in the production of energy, they play a role in other functions such as apoptosis and synthesis of some compounds as steroids and heme among others reviewed in (Butler, 2005).

As a reminiscence of their prokaryotic origin, mitochondria have an independent genome that is small and circular. In fact, the mammalian mitochondrial genome is about 16,569 base pairs long, much smaller than the nuclear genome that presents about 3.3 x 10^9 base pairs (bp). Although small, the mitochondrial genome is very economical as the distribution of genes is very dense (1 gene per 0,45kb instead of 1 gene per 100kb present in nuclear genome) and some genes overlap. In addition it lacks introns and presents little intergenic DNA. It is noteworthy to say that during the evolution of these organelles several genes were lost and others were transposed to the nuclear genome reviewed in (Burger *et al.*, 2003).

MtDNA is composed of two strands, one named H (Heavy) and richer in Guanine and the other called L (Light) and carrying more Cytosine. The genome is divided in two regions, the Coding and the Non-Coding Region. The Coding Region is composed of 37 genes: 13 polypeptides, 2 ribosomal RNAs and 22 transfer RNAs. The Non-Coding region is commonly called Control Region (CR) or D-Loop (Displacement Loop) because during the synthesis of a small fragment of the H Chain (7S DNA) it forms a triple chain structure. The CR represents about 7,2% (around 1,122bp) of the complete genome and plays a role on the replication and transcription of mitochondrial DNA. The CR presents Hyper-Variable Regions (HVR) with a high level of variation among individuals, namely the HVRI that incorporates the region from base 16,024 until 16,365 in a total of 342bp, the HVRII that extends from position 73 to 340 (267bp) and

the HVRIII that ranges from position 438 until 576 (121bp). The CR also encompasses Variable Regions (VR) that reveal less variation among individuals than the previous regions and therefore are more conserved segments, respectively VR1 and VR2 that are located in between the HVRs (Butler, 2005).

Mitochondrial DNA has some characteristics that make this marker valuable to anthropological and forensic applications as will be discussed below.

1.1.2.1. Pattern of Transmission

MtDNA is maternally inherited, meaning that mothers transmit mtDNA to their children and only the daughters will further transmit it, as observable in Figure 1.A. The reason behind this is that during the fertilization the father's mitochondria are located in the sperm tail leading to a nearly null contribution. If by chance a paternal mitochondrion enters the oocyte, besides being largely diluted in the maternal contribution (100,000 maternal mitochondria present in the oocyte), it also is marked with ubiquitin and follow an elimination path, remaining doubts on whether this elimination procedure acts at the fertilization or soon after it occurs (Manfredi *et al.*, 1997). In 2002 Schwartz and co-workers reported a case where the paternal transmission of mitochondria occurred, however this was considered a rare event and the exclusively maternal inheritance is presently accepted as being the rule (Schwartz & Vissing, 2002).

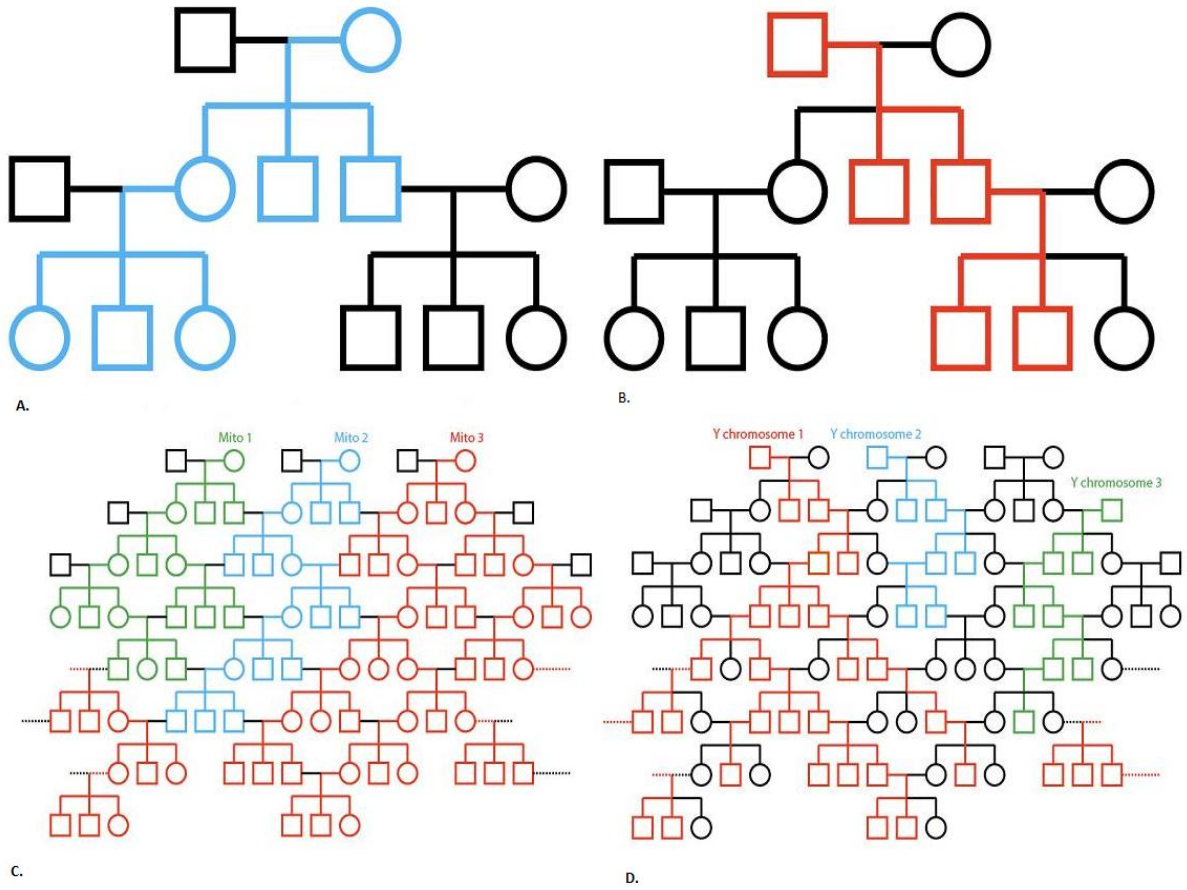


Figure 1 - Shcematic illustration of the patterns of inheritance of the uniparental markers mtDNA (A) and Y-Chromosome (B). Figures C and D represent respectively the transmission pattern of mtDNA and Y-Chromosome across generations. Squares represent males and circles represent females. Adapted from www.biologos.org.

Due to its mode of inheritance, it is possible to say that, in a similar way as for the Y-Chromosome (which is paternally transmitted, Fig 1.B), the effective population size of the mtDNA is ¼ relative to autosomes. This small effective population size makes these markers more sensitive to detect demographic events like bottlenecks and population expansions. Lineage markers also lack recombination, thus enabling the perception of lineages (Figure 1C and 1D) that tend to be geographically restricted. These features make both mtDNA and Y Chromosome useful in the study of populations' ancestries.

1.1.2.2. Copy Number

In the majority of somatic cells it is possible to find about 1,000-10,000 mitochondria. Additionally, each mitochondrion carries approximately 4 to 5 molecules of mtDNA, however the variation can be from 1 molecule to 15. The high copy number allows the mtDNA to be easily retrieved and isolated from the cells' tissues, simplifying the DNA extraction and amplification and therefore making mtDNA an excellent marker to study in cases where nuclear DNA is degraded (Butler, 2005).

The high number of mtDNA copies per cell may increase the complexity of the mtDNA analysis, because sometimes it is possible to have different copies within the same individual or tissue, a condition called heteroplasmy. The proportions of the variants may differ between tissues and change in different life stages of the individual.

1.1.2.3. Mutation Rate

MtDNA is also characterized by a high mutation rate. This is a consequence of a series of combined factors as the lower efficacy of the DNA repairing processes in mtDNA, the high number of replicating cycles, the high levels of oxygen radicals present inside mitochondria that damage the DNA and other structural deficits as absence of histones (Butler, 2005).

The CR reveals higher mutation rates than the other parts of this genome due to the formation of a temporary single stranded structure during the replication process. The single stranded DNA presents a depurination rate around 4 times higher than double stranded DNA and therefore accumulates more mutations (Butler, 2005). Moreover, mutation rate is not uniform along the CR as some positions, called hotspots, are more prone to mutate and others are more conserved.

1.1.2.4. Lack of Recombination

Contrarily to other markers (Autosomes or X-Chromosome), uniparental markers (mtDNA and MSY- Male Specific Region of the Y chromosome) do not suffer recombination and are transmitted as haplotypes. Therefore, apart from mutations, mtDNA is transmitted intact from mothers to offspring, meaning that all maternally related individuals will carry the same mtDNA sequence (Budowle *et al.*, 2003). Consequently, mtDNA markers are not suitable to identify individuals but female lineages.

Despite numerous publications on the possibility of recombination in mtDNA, there was no direct support evidence on this matter and so the subject was put aside by the majority of researchers (Ingman *et al.*, 2000; Elson *et al.*, 2001; Wiuf, 2001; Herrnstadt *et al.*, 2002).

1.1.2.5. Applications in Population Genetics

Lineage markers such as mtDNA or Y-Chromosome display some characteristics that confer them unique abilities in the field of population genetics. The uniparental transmission and the lack of recombination allow the perception of phylogenies which if associated with geographic patterns allow phylogeographic inferences such as the enlightening of populations' main migratory movements. The low effective population size makes these markers more sensitive to genetic drift, increasing the levels of population differentiation and substructure.

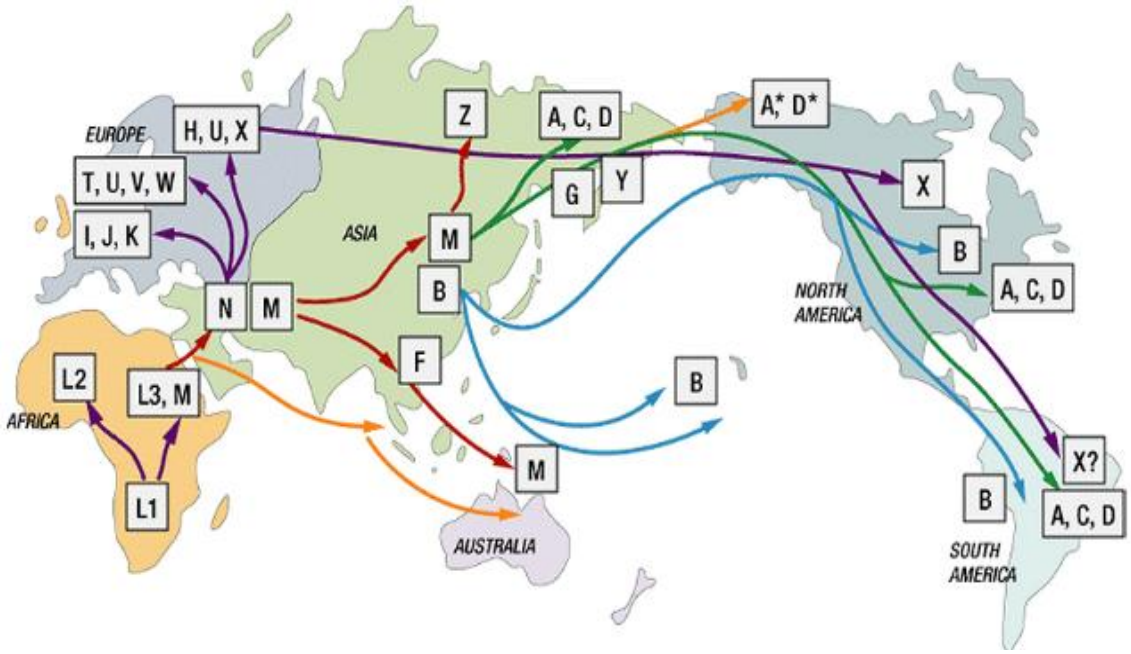


Figure 2 - Schematic representation of the human main migrations by the geographic distribution of the major mtDNA haplogroups. Haplogroups are designated by their letter. Adapted from familytree.com.

In anthropological terms, mtDNA has been used to understand not only the origin of modern Human populations by tracing the mitochondrial eve, but also recent Human migrations and demographic events that populations underwent. The differences found between individuals are called polymorphisms and the sharing of certain polymorphisms defines Haplogroups. The high mutation rate of this marker combined with the lack of recombination allows some polymorphisms to be geographically or population restricted.

Under the theory of human evolution "Out of Africa", it has been shown that all non-African mitochondrial variation is originated from Macro-haplogroup L3. It is estimated that the first migration out of Africa occurred between 50,000-70,000YBP (Soares *et al.*, 2009) and led to the formation of M and N which originated the R haplogroup. These haplogroups colonized the Middle East and migrated to Europe and Asia (Macaulay *et al.*, 2005; Roostalu *et al.*, 2007). Some Asian haplogroups reached Beringia (21,000-18,000YBP) and originated haplogroups A, B, C and D that continued to the Americas (Figure 2 and 3) (More details on section 1.2.2.2.1. Mitochondrial DNA Evidence).

The fact that mtDNA is easily extracted and amplified (see section 1.1.2.2. High Copy Number) makes it an excellent marker to use in ancient DNA studies (when the DNA is degraded), leading to the elucidation of numerous historical or even pre-historical cases.

1.1.2.6. Applications in Forensic Genetics

The advantages of the use of mtDNA in forensic cases are the high number of copies per cell and the fact that mtDNA is less prone to degradation, leading to an easier recovery of DNA. There are many applications in forensic casework, and most of the times the mtDNA is used in cases where samples are too degraded or too ancient to retrieve autosomal DNA. In this situation, mtDNA can be useful in parentage testing when the only relative available is related by the mother's side; or even in cases of missing persons and mass disasters when all the closer relatives are not available for testing (Butler, 2005).

The pattern of inheritance and the lack of recombination are advantageous, because they allow linking an individual to a lineage, however these characteristics are also disadvantageous because they only identify the lineage and not the individual itself.

The uniparental markers, namely mtDNA and Y-Chromosome, have a smaller effective population size and therefore are more sensitive to genetic drift and enable a better

insight into the structure of the population. The analysis of the population structure is also important under a forensic and anthropological perspective. Consequently, it is important that databases include different populations, cosmopolite and indigenous, however the latter are of difficult access. Furthermore, in order to calculate the haplotypic frequency by direct counting it is important that the databases are large to prevent biased results. Gathering and depositing of data in forensic and population genetics databases (such as EMPOP©) can improve the accuracy of the reports (Salas *et al.*, 2007). EMPOP© was created to answer the need for a high quality forensic and population genetic mtDNA database. MtDNA haplotypes deposited in EMPOP© are subjected to a series of high quality control tests regarding the quality of the typing and the phylogenetic haplotype calling and haplogroup classification (Parson *et al.*, 2004).

1.2. Colonization of the Americas

1.2.1. Entrance in the Americas

Contrarily to other continents where visibly archaic forms of *Homo sapiens* were discovered, such as *Homo sapiens neanderthalensis* in Europe or *Homo erectus* in Asia, the American continent reveals no such presence.

The colonization of the American continent has been under scientific scrutiny for a long time. Since the first European settlers arrived in America, theories arose about how the Americas were first inhabited. In fact, the first inductive theory was described in *Natural and Moral History of the Indies* by a Spanish Jesuit Priest, José de Acosta in 1590 (Acosta *et al.*, 2002). This theory suggests that Native Americans must derive from Asian populations and predicts an entrance by an overland via, foreseeing the Bering Strait (Mazières, 2011; O'Rourke & Raff, 2010).

The first model for the American Colonization was suggested by Ales Hrdliča in 1937 and named The Clovis First, The Single Origin model or the Blitzkrieg model (Rothhammer & Dillehay, 2009). The Clovis complex is an archaeological tradition commonly found in North America, dated between 11,500 and 10,900 radiocarbon years (Waters & Stafford, 2007) and associated with large mammalian remains, that indicates the Clovis people sustainability came from hunting. The model states that the Paleoindians entered the Americas through Beringia around 11,500YBP during the Clovis time and were the ancestors of the Amerindians (Dillehay, 1999; Dixon, 2001; Adovasio & Page, 2003; Haynes, 2002; Meltzer, 2004; Goebel *et al.*, 2008; Rothhammer & Dillehay, 2009). More recently, archaeological data gathered in South America predated the Clovis and concluded that these were not the first inhabitants of the continent (Bryan, 1986; Dillehay, 1997; Dillehay, 2000; McAvoy *et al.*, 1997; Adovasio *et al.*, 1999; Stanford, 2002; Meltzer, 2004; Meltzer, 2006; Waters & Stafford, 2007; Goebel *et al.*, 2008).

Greenberg's Tripartite Model arose as the first interdisciplinary perspective of the colonization of the Americas, gathering information from various fields of study, such as biology, archaeology and linguistics. Greenberg *et al*., (1986) and Turner (1987) works postulated that the Americas were colonized by three migrations: the first conducted by the Amerinds that reached South America; the second performed by Na-Dene speakers that colonized the North-West Pacific coast and the third by the Eskimo-Aleut people that stayed in the Artic area (Greenberg *et al.*, 1986; Turner, 1987; O'Rourke & Raff, 2010; Rothhammer & Dillehay, 2009). This model was supported by the early

classical genetic markers' studies, since the ABO blood markers differentiated the three linguistic groups (Estrada-Mena *et al.*, 2010; Mourant, 1985). Nevertheless, this model was questioned by numerous researchers both on linguistic and biological basis (Morell, 1990; Neves & Pucciarelli, 1991; Szathmary & E., 1993; Szathmary, 1993; Lahr, 1995; Merriwether *et al.*, 1995).

In the early 90s another model arose due to the divergences found on craniometric records from American skulls that indicated that distinct populations entered in the Americas at different times. Two morphological patterns were found, one more similar with the Australians and sub-Saharan Africans and the other more similar with northern Asians (Neves & Pucciarelli, 1990; Neves & Pucciarelli, 1991; Neves *et al.*, 2003; Neves & Hubbe, 2005; Rothhammer & Dillehay, 2009). Despite being supported by some archaeological findings that propose the presence of two distinct technologies (Dixon, 2001), further studies on morphology stated that these variations were part of a gradient (Mazières, 2011).

Recent Studies

With new genetic studies, new hypotheses appeared regarding the colonization of the Americas. The first genetic studies were based on classical genetic markers and supported the Tripartite model. However, studies based on uniparental markers denied any major contribution to the Americas colonization besides the Asiatic, as almost all lineages found in the uniparental markers were found in North Asian indigenous groups, even though there could have been smaller alternative contributions.

Regarding mtDNA, 4 lineages were first discovered in Native Americans during the first studies - A, B, C and D - and later another one called X (Torroni *et al.*, 1993; Forster *et al.*, 1996; Brown *et al.*, 1998). While the first four lineages are found throughout the continent regardless of ethnic and linguistic groups, the X haplogroup is mainly found in North and Central America and nearly absent in South America (Dornelles *et al.*, 2005). This absence can be explained by a founder effect during the migration from North to Central and South America. Another explanation is that haplogroup X reached South America but in small frequencies and eventually became extinct during time (Dornelles *et al.*, 2005).

Despite several interpretations based on these findings, the latest and most accepted thesis is that all lineages were carried into America through a migration from a single population of origin. However, researchers remain doubtful in deciding if there was only one migratory movement or if there were more migrations from the same ancestral

population (Mazières, 2011; Perego *et al.*, 2009; Schurr & Sherry, 2004; Schurr, 2004). Y-Chromosome studies revealed that only haplogroup Q was characteristic of Native American populations (O'Rourke & Raff, 2010). Coalescent timing of all mtDNA haplogroups pointed to a common ancestor within 17,200-10,100 YBP, indicating that their separation into haplogroups must have occurred before the entrance in the Americas (O'Rourke & Raff, 2010). However different calibrations result in different coalescent time periods and Achilli *et al.*, (2008) considering only the Native American branches of each haplogroup obtained an average time for the split of different haplogroups of 20,200YBP.

The availability of the new data led Maziéres (2011) to formulate a consensus model based on several fields of study. This model postulates that Human populations migrated from Asia to the north-eastern part of Siberia during the late Pleistocene (26,000-18,000YBP) carrying a non-derived cranial morphology and a genetic background free from specific mutations. By this period the sea level decreased due to the glaciation and a landmass emerged between Siberia and Alaska, called Beringia. Archaeological and paleoecological data suggest that populations settled in Beringia and did not go further because North America was still buried in ice-sheets (Mazières, 2011; Tamm *et al.*, 2007). This natural barrier allowed for a population settlement and growth and also for the appearance of specific mutations, while still conserving an Asian genetic background and morphological traits. Since 18,000YBP until the end of the Pleistocene (around 10,000YBP) the mean temperature of the planet arose and the deglaciation took place leading to a rise in sea level, to the opening of some coastal routes and also of the continental Ice-Free Corridor. This promoted a reduction of the Beringian plain and compelled the populations to move southwards colonizing the Americas (Mazières, 2011).

Even though this model gives a consensus thesis on the entrance in the Americas, it does not clarify the issue of the dispersion routes within the continent. In fact, the time of the entrance in America is commonly settled around at least 15,000 years ago. By that time the northern subcontinent would be under the Wisconsin glaciation (25,000-10,000YBP) that covered the land with two extensive glaciers: Laurentide and Cordilleran, causing serious difficulties to the dispersion of the earlier populations southwards (Dillehay, 2009). Johnson in 1933 considered the existence of an ice-free corridor between the two ice-sheets Laurentide and Cordilleran (Rothhammer & Dillehay, 2009). Despite being considered, until recently, the most accepted dispersion route, geological and archaeological records argue against the existence of this corridor (Jackson *et al.*, 1997; Mandryk *et al.*, 2001; Clague *et al.*, 2004). In

consequence, newer approaches on dispersion routes taken by the first Americans were considered, in particular coastal routes.

Nowadays the coastal routes are gaining relevance, being considered the most congruent theses of genetic, archaeological and environmental data. Currently, it was proposed the occurrence of coastal migrations via the southern coast of Beringia through the use of watercraft followed by southwards and inland dispersion and settlement. This model explains the fast colonization of the continent and is supported by early archaeological findings in South America and corroborated by some genetic studies on mtDNA of Native Americans (Keefer *et al.*, 1998; Sandweiss *et al.*, 1998; Fix, 2005; Dixon, 2006; Wang *et al.*, 2007; Dillehay, 2009).

Genetic studies also have not yet reached a consensus: while some studies claim that the coastal routes of dispersion combined with riverine routes would have allowed a faster colonization process, therefore justifying the early archaeological evidence in South America (Fix, 2005; Wang *et al.*, 2007), other reports based on frequencies of rare mtDNA variants indicate that both inland Ice-Free Corridor and the Coastal Routes of dispersion were used (Schurr & Sherry, 2004; Kitchen *et al.*, 2008; Perego *et al.*, 2009).

Additional Contributions

During the 20th century, several entrance points were considered as candidates for American colonization, such as maritime and coastal or more inland routes. Other contributions to the American colonization were considered as well, with alternative entrances in the main continent such as via the Pacific Ocean by crossing the Melanesian and Polynesian Islands and through the Atlantic Ocean.

The Pacific Ocean thesis is based on archaeological findings in northwest South America that resemble the cultures found in some Melanesian Islands as well as in South Asian populations. This theory has recently gained some refreshment with data from the Y-Chromosome and mtDNA (O'Rourke & Raff, 2010; Estrada *et al.*, 1962) and will be discussed below (1.2.2.1. Entrance Points and Routes of Dispersion).

With regard to the entrance via the Atlantic Ocean hypothesis, it relies on archaeological data and states that the Clovis complex (archaeological tradition found in North America) resembles in manufacturing procedures the European Solutrean tradition found in the Iberian Peninsula and southern France (Bradley & Stanford, 2004).

Recently, some considerations have been taken on an alternative entrance via Beringia's northern coast, which followed the Atlantic coastlines southwards. The northern coast of Beringia has been inhabited for about 30,000 years. Therefore, if these populations had a coastal economy and used watercraft systems, it would be possible that these populations have reached Alaska before the Last Glacial Maximum (LGM) (Brigham-Grette *et al.*, 2004; O'Rourke & Raff, 2010; Ebenesersdóttir *et al.*, 2011).

1.2.2. Reaching the South

Even though more recent, the colonization of South America raises even more controversial issues than North America since the number of migratory events and dispersion routes taken by the first Native American populations within the subcontinent remain yet to be unveiled.

The entrance and colonization of South America was markedly different from the process that occurred in the northern part of the continent. In terms of archaeological records, it is noteworthy that in South America there was not the prevalence of a certain culture, in the way the Clovis did in the northern subcontinent. This was probably a consequence of the high variability of environments of which some were seriously adverse leading to the appearance of local and regional traditions. Moreover in South America the ice glaciers were restricted to high altitude zones (Clapperton, 1993), contrarily to North America where the glaciers covered an extensive area of landmass limiting the population movements (Rothhammer & Dillehay, 2009).

During the late Pleistocene (between 11,000-10,000YBP), South America endured various environmental and climatic changes causing the alteration of landscapes and of the distribution of fauna and flora (Rothhammer & Dillehay, 2009). Following these massive modifications, the conditions led to a demographic increase, technological innovations, the advent of agricultural practices and cultural rituals solidification. These processes were not only rapid but also promoted the development of regional traditions (Rothhammer & Dillehay, 2009; Dillehay, 2000; Lavallee, 2000).

1.2.2.1. Entrance Points and Routes of Dispersion

Biogeographically, South America can be divided in four main regions: the Andean chains; the humid and fertile plains of Colombia, Venezuela and Brazil; the eastern Brazilian highlands and finally the southern part of the subcontinent formed by the Guyanas, Patagonia and southern pampa. Some of these regions acted as natural barriers for the demographic movements causing the isolation of some populations and

contributing to the appearance of distinct local cultural and technological practices (Rothhammer & Dillehay, 2009).

Concerning the entrance points into this subcontinent, the most accepted hypothesis is that South America was colonized via the Isthmus of Panama and reached Colombia (more detail in section 1.2.3. Colombia). Nevertheless, there is an alternative hypothesis which states that there was another contribution to the South American gene pool, originated from populations from South-eastern Asia that colonized the Australo-melanesian and Polynesian Islands between 5,000-3,000YBP (Arnaiz-Villena *et al.*, 2010). These latter settlers would have reached the Pacific Coast of South America using watercraft. Recently, genetic evidence was found that can confirm this thesis but it is restricted to some Y-Chromosome lineages found in the Pacific coast of South America (Roewer *et al.*, 2012) and also to the high frequencies of mtDNA haplogroup B in the Andean region (more details on section 1.2.2.2.1 Mitochondrial DNA Evidence). Archaeologists also discovered remains that culturally resemble those found in some populations on the other side of the Pacific Ocean (O'Rourke & Raff, 2010). Recently, other genetic evidence was found on admixture between Amerindians and southeast Asians and other Pacific inhabitants regarding HLA genes (Arnaiz-Villena *et al.*, 2010).

The number of migrations into South America is still widely discussed; some researchers argue that there was only one migration (Moraga *et al.*, 2000), others suggest that there were two migratory movements (Greenberg *et al.*, 1986; Wallace & Torroni, 1992; Fox, 1996; Lalueza *et al.*, 1997; Keyeux *et al.*, 2002; Achilli *et al.*, 2008); Recently Fuselli *et al.* (2003) proposed a two fase migration that initiated descending the Andes and then came upwards through the Amazon basin (Fuselli *et al.*, 2003; Tamm *et al.*, 2007).

The dispersions routes within the southern subcontinent also have been extensively debated, although some theses are more accepted. Since the first hunter gatherers entered America through the Isthmus of Panama and the Colombian territory they could have easily reached the Andes via Cauca and other river valleys within Colombia. After having adapted to this environment and developed agricultural practices and culture rituals, these populations might have settled in the highlands of the Andean chain and moved southwards. It is important to mention that some of the agriculture practices used in the highlands are difficult, if not impossible, to establish in lower altitude areas. In agreement, the populations that developed such cultures probably remained in the highlands, migrating towards South through the Andean

cordilleras, and populations that settled in the eastern lowlands of the Amazonian plains pursued their migration nearby the river basins reaching the inner regions of Brazil (Rothhammer & Dillehay, 2009). Although these are the most cited, there have also been described migrations following the Pacific coastline associated with the development of fishing tools and watercraft (Rothhammer & Silva, 1992; Cavalli-Sforza *et al.*, 1994; Keefer *et al.*, 1998; Sandweiss *et al.*, 1998; Stothert, 1998; Wiesner, 1999; Wang *et al.*, 2007; Rothhammer & Dillehay, 2009) and throughout the Caribbean Coast into Venezuela nearby river basins and into the Amazonian basin (Rothhammer & Dillehay, 2009).

1.2.2.2. Genetic Evidence

Generally, a decreasing gradient of genetic variability from North to South of South American continent can be observed (Rothhammer & Silva, 1992; Rothhammer *et al.*, 1997; Rothhammer & Dillehay, 2009). The patterns of genetic diversity also vary from West to East in South America. Indeed, populations in eastern South America reveal lower values of genetic diversity when compared to their western counterparts. Fuselli and co-workers (2003) proposed that in the Andean region there is a uniformity caused by high long-term effective population sizes and gene flow, also seen in the uniformity of languages (only Andean languages) and cultural traditions, which were probably a consequence of the prevalence of the Inca Empire. The Inca Empire also promoted the gene flow between groups and the maintenance of higher population sizes leading to a general level of diversity that was kept higher than in other regions. On the contrary, in the eastern part of this subcontinent a heterogeneous pattern is found not only on the numerous languages that are spoken but also on the ethnicities and on the local genetic divergences. The absence of gene flow between populations and the small population sizes caused strong genetic drift effects and led to an increased divergence between populations.

These results are acknowledgeable both in mtDNA and Y-chromosome data (Tarazona-Santos *et al.*, 2001; Fuselli *et al.*, 2003) and suggest that the colonization of the western part of South America happened first and led to a homogenization of the gene pool and it was followed by the peopling of the eastern region probably by western groups. In the eastern regions these groups settled in smaller communities leading to stronger genetic drift that has caused local differentiations and the heterogeneity found today (Callegari-Jacques *et al.*, 1994; Fuselli *et al.*, 2003). This thesis was recently supported by a mtDNA study (specifically of D haplogroup) that proposes a rapid Pacific coast colonization, explaining the early settlement in Monte

Verde (Chile), and followed by several *trans-Andean* migrations which were more intense in the South of the subcontinent (Bodner *et al.*, 2012).

There were also found similarities between Mesoamerican and Andean populations interpreted as an early coastal colonization (Wang *et al.*, 2007), although Rothhammer & Dillehay (2009) explain these findings as a consequence of the contacts between the two most developed cultures in the Americas that arose in Peru (Inca) and Mexico (Maya). The latter thesis finds some support in Spanish Chronicles describing transpacific journeys that connected both civilizations, during the Inca Empire (Rivet, 1943; Valboa, 1951; Pietschmann & de Gamboa, 1906).

1.2.2.2.1. Mitochondrial DNA Evidence

Throughout the American continent five major mtDNA haplogroups can be found, A, B, C, D and X whose phylogeny can be observed in Figure 3. However, recent literature has adopted only the names of American sublineages: A2, B2, C1, D1 and X2a (Figure 3). While haplogroups A, B, C and D are considered to be Pan-American since they are ubiquitous in the continent, haplogroup X appears to be absent in South America (Dornelles *et al.*, 2005; Achilli *et al.*, 2008).

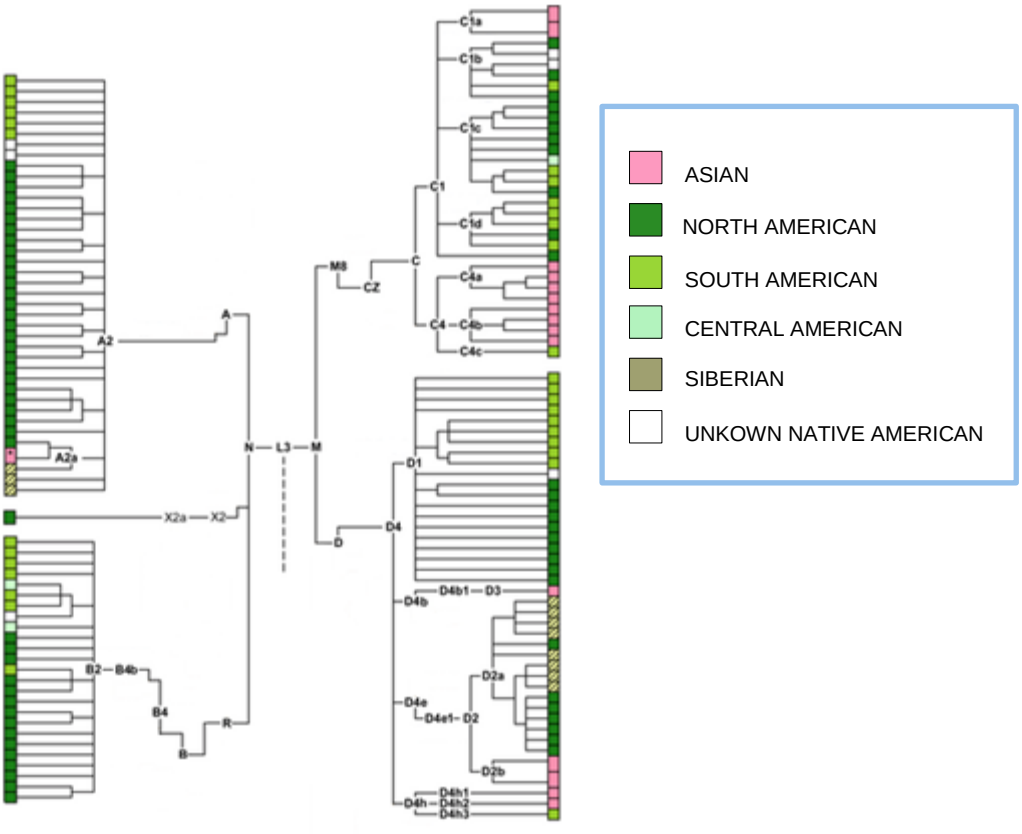


Figure 3 - The phylogenetic tree of Native American mtDNA haplogroups [adapted from reference (Tamm *et al.*, 2007)].

Figure 4, adapted from Salas *et al.* (2009), shows a representation of the distribution of the major mtDNA lineages in South and Central America. Haplogroup A is highly frequent in Mesoamerica and northwest South America, followed by haplogroups B and C. Accompanying the decrease in frequency of haplogroup A towards South, there is an increase in the frequencies of haplogroups C and D. Concerning haplogroup B, it has higher frequencies in Andean populations as referred by Fuselli *et al.* (2003). Some researchers consider the high frequency of B haplogroup in the Andean region to be a consequence of a coastal entrance, implying an alternative entrance point into the Americas, explaining also the small frequencies of this haplogroup in North America and in the southern cone of South America (O'Rourke & Raff, 2010).

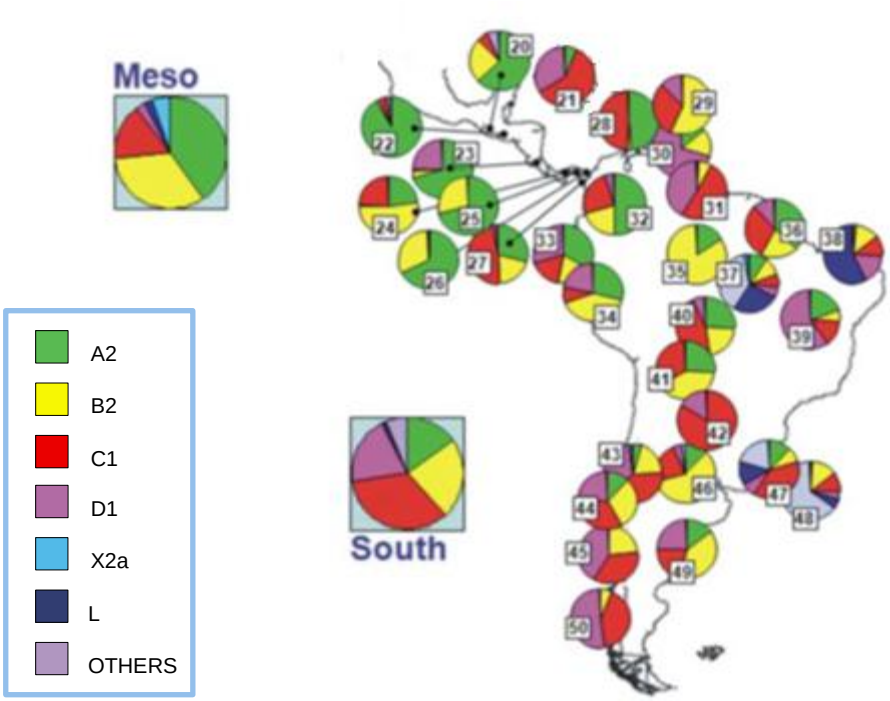


Figure 4 - Haplogroup distribution in Central and South America. Meso-America: 18 = Pima; 19 =Mexico; 20 =Quiche; 21= Cuba; 22= El Salvador; 23=Huetar; 24 =Embéra ;25= Kuna; 26 =Ngöbe; 27 =Wounan; South America: 28=Guahibo; 29 =Yanomamo from Venezuela; 30=Gaviao; 31= Yanomamo from Venezuela and Brazil; 32= Colombia; 33 = Ecuador; 34 = Cayapa; 35 = Xavante; 36 =North Brazil; 37 =Brazil; 38 = Curiau; 39 = Zoro; 40 =Ignaciano, 41 =Yuracare; 42= Ayoreo; 43 = Araucarians; 44=Pehuenche, 45=Mapuche from Chile; 46= Coyas; 47 = Tacuarembo; 48 =Uruguay; 49 =Mapuches from Argentina; 50= Yaghan. Illustration adapted from reference (Salas *et al.*, 2009).

1.2.3. Colombia

The rich diversity in ethnicities and linguistics combined with its relevant role as an entrance point in South America makes Colombia an interesting object of study in population genetics and for forensic purposes. In order to promote a better

understanding of the population dynamics of this country, some points on history, demography, linguistics and genetics will be referred.

1.2.3.1. History

Colombia pre-colonial history is based on archaeological findings discovered throughout the country. Such discoveries seem to suggest the existence of three distinct periods in the pre-colonial history. The first one is characterized by the presence of *Paleoindians* which were hunter-gatherers that entered in South America. Afterwards the *Herrera* culture prevailed with a tradition known to be associated with the advent of agriculture and ceramics. Finally the *Agroceramic* tradition established the agricultural practices and manufactured more sophisticated pottery (Casas-Vargas *et al.*, 2011).

Christopher Columbus arrived in the Americas for the first time in the year 1492, followed by the arrival of Pedro Álvares de Cabral in Brazil in 1500. During the following centuries both Spanish and Portuguese and afterwards other Europeans colonized the Americas and imposed their cultures and governments.

By the time the Spanish navigator, Alonso de Ojeda, arrived in Colombian territory in 1499 the distribution of the Native populations differed from the one visible now. Some examples are the Chibcha-speaking groups that occupied Sierra Nevada and the prevalence of Caribe speaking populations in the Atlantic Coast. Moreover, various Chocó groups inhabited the Pacific coast, whereas in the southern part of the Country lived numerous ethnic groups as Pasto and Nasa (Arango & Sánchez, 2004; Arango & Sanchéz, 1998; DANE, 2007). More detailed observations on ethnic groups will be discussed below.

Only in the year of 1503 did the Spanish Royalty allow the slavery of Native people if they did not accept both the Spanish Realm and Christianity or if proven they were cannibals or idolaters. The several martial conflicts that were conducted between Indigenous and Europeans and the spread of European diseases led to a dramatic reduction of the Native's numbers. The consequence of this reduction was a lack of workers that were needed at that time to create an empire, and so the Spanish crown conceded *Licencias* that allowed the legal introduction of African slaves into the New World. In 1538, the *Encomienda* was created in order to construct villages where the dominance of the Crown was more prevalent, in which the *Encomendero* kept the lands and the slaves (both Africans and Native) (Arango & Sánchez, 2004; Arango & Sanchéz, 1998; DANE, 2007) and these were distributed through the villages by

separating the ethnic groups (Klein, 1999; Friedemann, 1993). In 1536-1561, the Spanish King instituted the first *Resguardos* for indigenous people, located in lands previously occupied by Native populations and internally organized based on their traditional hierarchies (Arango & Sánchez, 2004; Arango & Sanchéz, 1998; DANE, 2007).

In the 19th century, the country of Colombia gained independency and constituted a Republic. During this period several arrangements were made with the purpose of removing the lands of the *Resguardos* from the Native populations to become part of the Governmental estates. Even though Simón Bolívar had declared the return of the *Resguardos* territories to the Indigenous populations, by 1821 this decree had been ignored and all the lands had been divided or invaded. Furthermore, the Republic elaborated a plan for converting the Indigenous populations to the civilization standards and to Christianity. However such measures did not have the expected results and caused various rebellions and resistance to the normative and religion imposed. Additionally these impositions led to the movement of some populations into the forests to conceal themselves. In 1890, a law was created to minimize the damages and the agricultural crisis caused by the war confronts promoting the reinstate of the *Resguardos* lands to the Natives and also allowing the populations to establish their own governments (Arango & Sánchez, 2004; Arango & Sanchéz, 1998; DANE, 2007).

During the 20th century and until today several enforcements have been put into practice to grant the Indigenous communities various social rights, such as the retrieval of usurped lands, the defence of the ethnic groups and their languages under a bicultural and bilingual education and the control and improvement of their products and economies. Other languages were also made official in the areas of the country where they are spoken (Arango & Sánchez, 2004; Arango & Sanchéz, 1998; DANE, 2007).

1.2.3.2. Demography

Scholars do not share a consensus opinion about the demography of indigenous populations in Colombia before the arrival of the Europeans in the 15th century. A unanimous certainty is that Colombia was inhabited by several ethnic groups which carried a wide number of languages.

Although the European Conquest in the New World led to a significant reduction of the number of Native Americans (about 90%), in Colombia the remains of such diversity are still observable as it contains about 87 recognized ethnic groups. Even though the

Castellan is the present official language in Colombia, there are still 64 indigenous languages spoken throughout the country (DANE, 2007). The last census demonstrated the existence of 1,392,623 (3.43%) indigenous people in a total of 40,607,408 people that inhabit Colombia.

Their distribution is not even throughout the country and more information on which ethnic groups inhabit each Department and on the percentage of indigenous individuals per Department is observable on Table 1. The Departments of Vaupés, Guanía, La Guajira, Vichada and Amazonas reveal the highest percentages of Amerindians in Colombian territory. In fact, the majority of Native people inhabit rural zones of the country or Indigenous reserves; anyhow there is a small minority that lives in the cities, normally due to the lack of lands in the reserves and to changes in their cultures (DANE, 2007).

Table 1 - Distribution of indigenous individuals per Colombian Department. Antioquia and Cauca Departments are in bold because of their relevance in this work. Table adapted from (DANE, 2007).

DEPARTMENT	% OF ETHNIC POPULATION
Amazonas	43.43
Antioquia	**0.53**
Arauca	2.24
Atlántico	1.33
Bogotá	0.23
Bolívar	0.11
Boyacá	0.49
Caldas	4.3
Caquetá	1.61
Casanare	1.48
Cauca	**21.55**
Cesar	5.15
Córdoba	10.39
Cundinamarca	0.34
Chocó	12.67
Guainía	64.9
Guaviare	4.3
Huila	1.05

La Guajira	44.94
Magdalena	0.81
Meta	1.28
Nariño	10.79
Norte de Santander	0.61
Putumayo	20.94
Quindío	0.41
Risaralda	2.9
San Andrés	0.1
Santander	0.13
Sucre	10.96
Tolima	4.32
Valle	0.56
Vaupés	66.65
Vichada	44.35

1.2.3.3. Ethnic Groups

In this section a set of ethnic groups are described, both in terms of the main historical occurrences and their present demographic distribution (Figure 5). There is also a small description of the main subsistence ways of the populations and their languages. In order to simplify, as Colombia presents 87 ethnic groups, the descriptions are restricted to those groups that constitute the sampling of the present study.

Embéra

The Embéra ethnic group used to include several groups that shared the same pre-Hispanic histories such as the Katíos, the Chamí and the Eperara-siapidara. However, in the last census these ethnic groups were considered separately (Embéra; Embéra-Chamí; Embéra-Katío; Eperara-siapidara) and these categories are the ones here adopted. Their language belongs to the Chocoan family that was recently inserted in the Paezan group (Figure 6).

In the pre-colonial period, the Embéras inhabited the higher channels of the rivers Atrato and San Juan in the Department of Antioquia in the northeast part of the country. However, because of the Hispanic-Indigenous conflicts and the impositions of the Spanish Crown, the Embéra populations fled in small groups to the coastal plains and to the Andes cordilleras. Nowadays they are about 49,686 people and occupy the

Departments of Antioquia, Bolívar, Caldas, Caquetá, Cauca, Chocó, Córdoba, Nariño, Putumayo, Risaralda and Valle de Cauca (Figure 5).

The Embéra-Chamí ethnic group shares the same pre-hispanic history with the Embéras but this group resisted the Hispanic invasions until the 17th century when they escaped to the riverine areas within the wet forests. The Embéra-Chamí populations are dispersed close to river basins as they are culturally adapted to humid and tropical areas. At the present-day they are about 5,511 people and inhabit the Departments of Risaralda, Antioquia and Valle de Cauca primordially, but it is possible to find other settlements in Quindio and Caldas.

The Embéra-Katío group seems to have favoured an escape towards the mountain chain of the Andes, instead of settling close to the river basins.

All these ethnic groups share similar social and economic traditions. The populations are governed by the *Jaibaná* who has the responsibility to defend, cure and educate. They are structured into nuclear families and usually all members of the same family live in the same house. The economy is based on fishing and hunting but also the farming of sugar cane and maize among other cultures (Arango & Sánchez, 2004).

Coconuco

In the last census, 2005, the Coconuco amounted to 6,767 people and occupied the right shore of the Cauca River in the *"Resguardos"* of Coconuco, Puracé and Paletará (Figure 5). Nowadays the Coconuco language is extinct but the researches consider it as belonging to the Paezan group. This culture used to live in the Central cordillera of the Andes, however with the Hispanic invasions they were forced to change their location. Since the Republic period Coconucos are socially organised to claim for the restoration of indigenous lands (Arango & Sánchez, 2004).

Guambiano

There is no consensual opinion on the history of the Guambiano people and while some propose they were introduced in Colombia from the Ecuador with the Spanish conquerors, other researchers suggest that there was a vast ethnic group named *Pubenses* that was constituted by all the groups that inhabited the area and were under the government of two leaders. The Guambia (or Guambiano) people were distributed into the *Encomiendas* as workers and their lands have been restored recently (Arango & Sánchez, 2004). Archaeological findings support that the Guambianos have

inhabited the Guambia *Resguardo* in organized communities since before the arrival of the Europeans (Llanos, 1981; Romoli, 1974).

At the present-day, the Guambiano are about 23,462 people, of which 77% live in the *Resguardo* of Guambia in Silvia, Department of Cauca (Figure 5). The Guambiano are known to live in small villages of 20 individuals each, where they are socially organized in nuclear families and their subsistence is provided by agriculture. Their language is also a polemic issue, as investigators are doubtful if it belongs to Paez or to Barbacoan families, both inserted in the Paezan group (Arango & Sánchez, 2004).

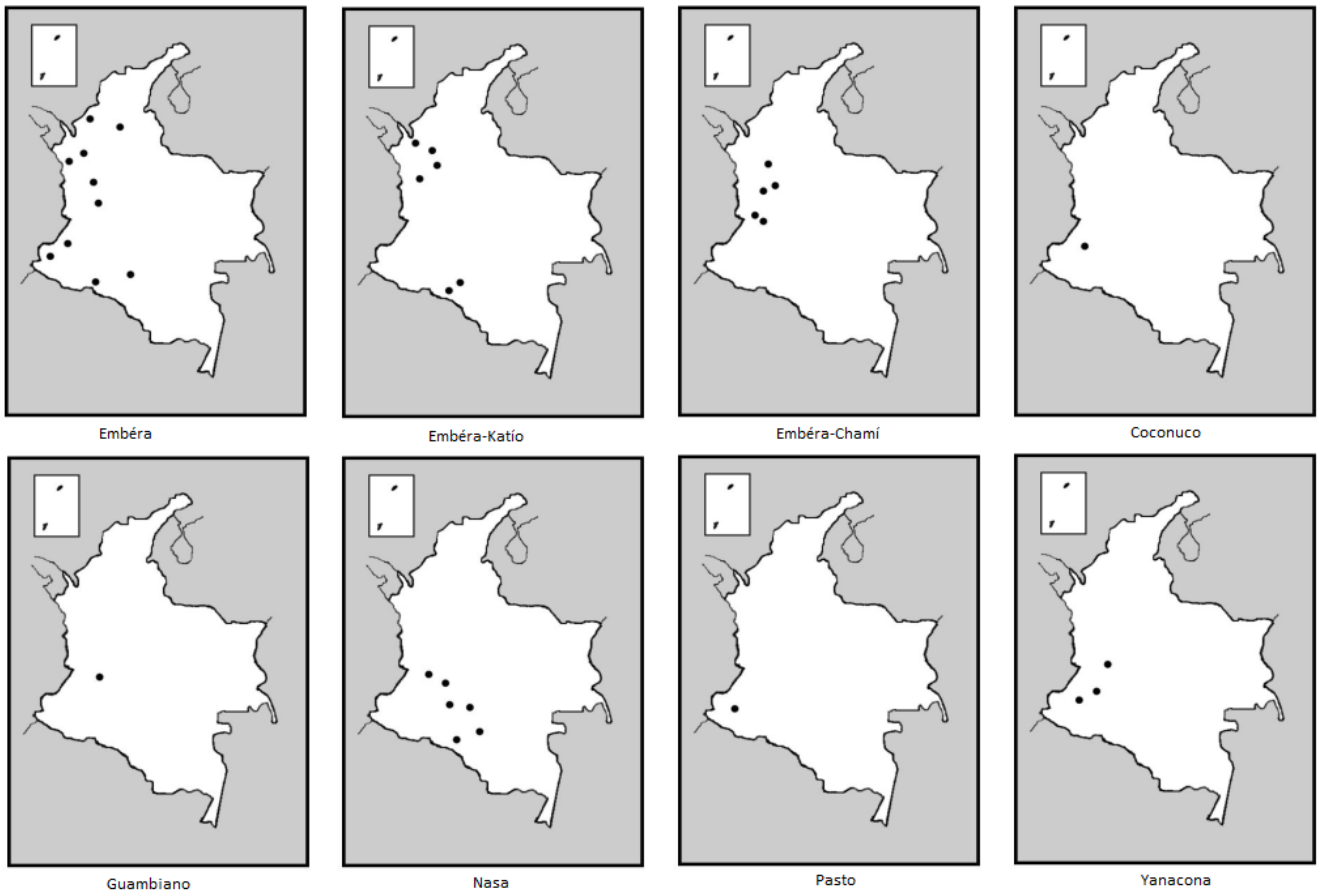


Figure 5 - Present distribution of the ethnic groups discussed in this work in the Country of Colombia. Adapted from Arango & Sánchez (2004).

Nasa

The Nasa people speak nasa yuwe that falls within the Paezan group and normally occupy areas in the Andes, in the southern part of the country (Figure 5). Nasa people are thought to be originate from the tropical forests, where they lived in dispersed small groups under the command of one leader. They were part of the resistance forces and

after the *Encomiendas* and missions that caused the loss of their lands, the Nasa started to organize with the intention of retrieving those lands. Contrarily to other indigenous groups, the Nasa subsistence lays not only in agriculture but also in pastoral practices and commerce of handcrafted items (Arango & Sánchez, 2004)

Pasto

The Pasto inhabit mainly *Resguardos* and are mostly localized in the southern part of Colombia (Figure 5). The estimates of the last census point to 69,789 people and their original language is extinct. In ancient times, before the arrival of the Europeans, the Pasto lived in the Andean regions following the Guáitara River southwards until the Ecuador. Their social organization is divided in nuclear families and it is common that women move to the village of their husbands once they are married. They live mainly from agricultural products but also have some milk production and wool. They have commercial practices where they sell handicraft and dairy products (Arango & Sánchez, 2004).

Yanacona

Yanaconas live essentially in the southeast of the Department of Cauca (Figure 5), and even though their original dialect is no longer spoken, researchers think it belonged to the Quechua family (Figure 6). In the 2005 census, Yanacona were estimated to be around 21,457 people. Their territory was declared as property of the Spanish Crown in the 16th century and the populations were sent to work as slaves in the gold mines. Their religion is slightly different from what is common in indigenous populations because there is a visible influence of Catholicism; they are very devoted to virgins who are the founders of the different villages and the ones that end the conflicts (Arango & Sánchez, 2004).

1.2.3.4. Linguistic Families

As far as Amerindian languages are concerned, they are still deeply discussed between linguistic researchers. South American Amerindian languages are even more controversial as this subcontinent reveals higher linguistic diversity than North and Central America combined (Campbell, 2000). A review was performed to the classification and structure of Amerindian languages in 2007 and the results were as represented in Figure 6.

However there are still many doubts not only on the position of the Andean, Chibchan and Paezan groups but also on the relationship between them. Greenberg *et al*.,

(2007) considers the Chocoan and the Barbacoan families to be included in the Paezan group and while agreement has been achieved in what concerns the Barbacoan enclosure, the Chocoan family raises some questions since some consider it to be closer to the Chibchan group (Campbell, 2000; Constenla & Margery, 1991). The Paez family, also included in the Paezan group, contains the Guambiano, Paez and Coconuco languages which were mentioned above.

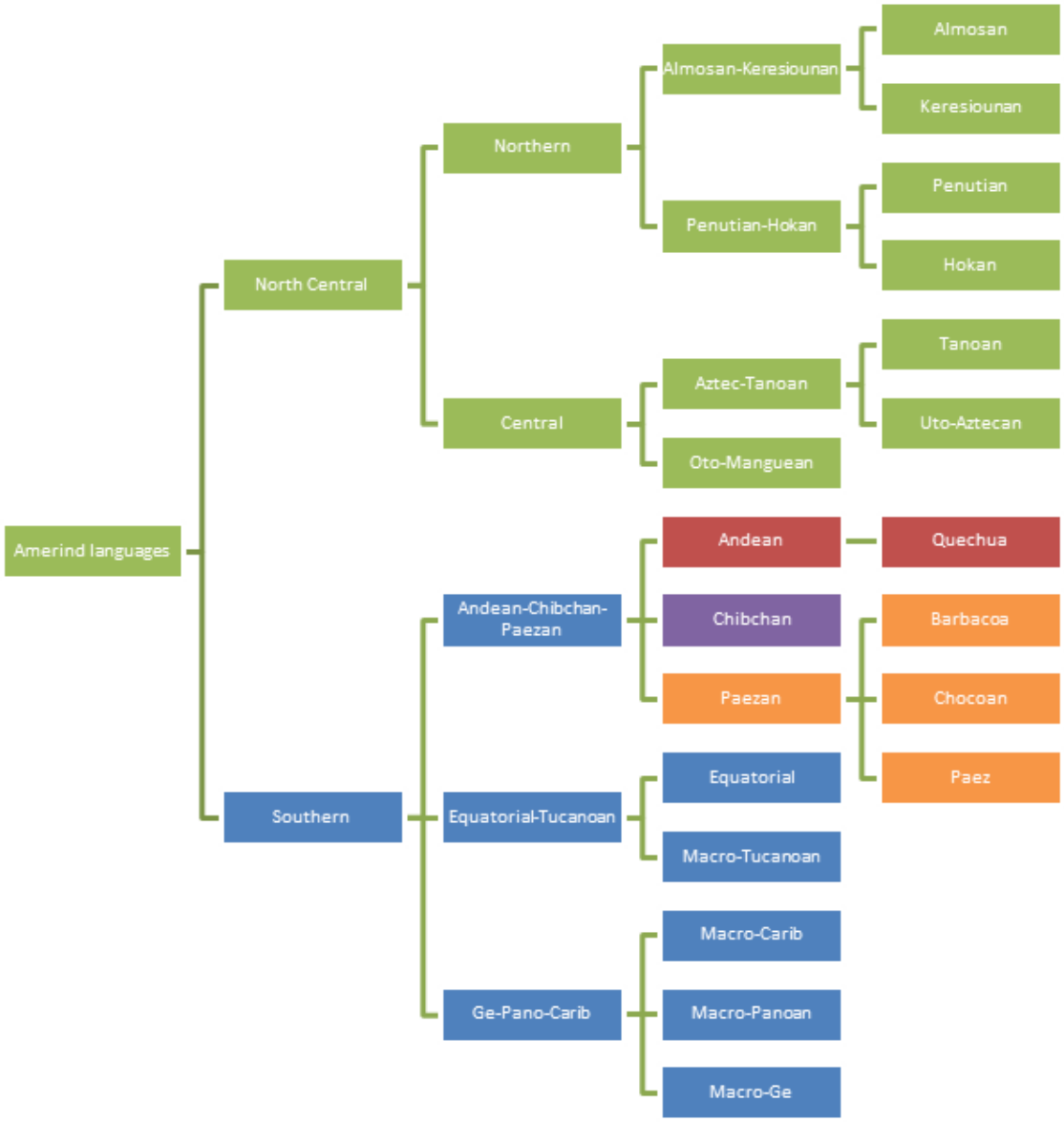


Figure 6 - Amerindian major linguistic groups spoken throughout America. In green are represented the languages spoken in North and Central America, in blue the major linguistic groups spoken in South America. Red boxes represent the Andean group (showing one linguistic family – Quechua - as example), the Chibcha group is coloured in purple and the Paezan group is represented in orange (showing some linguistic families as examples) (Greenberg & Ruhlen, 2007).

South America is divided in several linguistic areas; the following are of relevance to this study: the Colombian-Central American Area that is constituted of several

Chibchan, Chocoan and Barbacoan languages and the Ecuadoran-Colombian Subarea that is located in the Colombia-Ecuador border, which is characteristically an Andean zone and where Paez, Guambiano, other Barbacoan and several Quechuan (Andean) languages are spoken (Campbell, 2000). The areas described can be observed in detail in Figure 7.

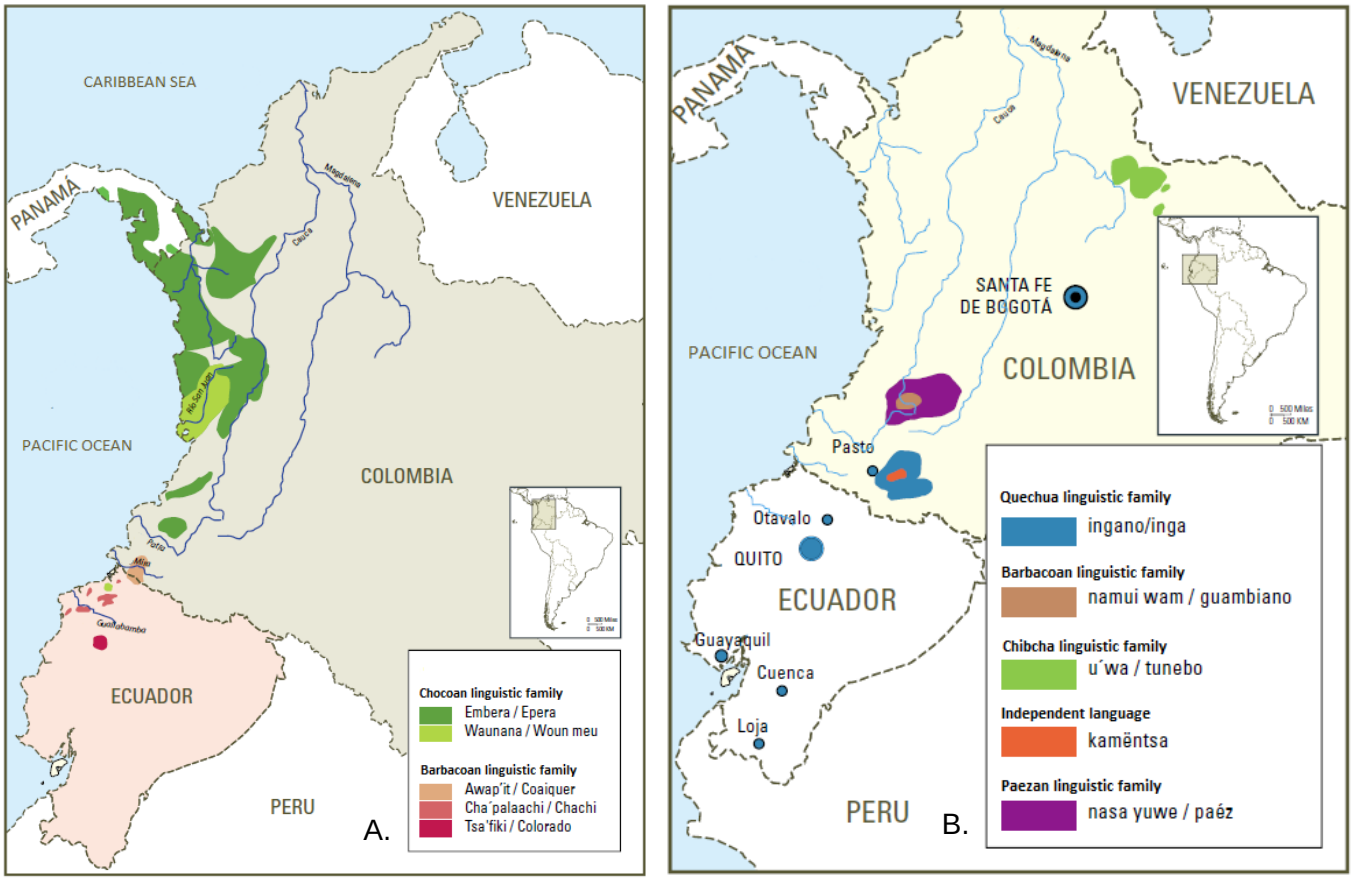


Figure 7 - Linguistic areas with relevance for this work are illustrated here. In figure 7.A. there is a description of the distribution of the Chocoan and Barbacoan linguistic families (classified as part of the Paezan major group) along the Pacific Coastline of Colombia and Ecuador. Figure 7.B. shows a display of several linguistic families, the majority belonging to Andean and Paezan macro-families in the Andean area of Colombia and Ecuador. (Sichra *et al.*, 2009; Curieux *et al.*, 2009).

1.2.3.5. Genetics of Colombian Populations

Colombia presents all Pan-American mtDNA haplogroups – A, B, C and D – with rare or absent haplogroup X as it is expected in South American countries (Dornelles *et al.*, 2005). Studies have shown that the genetic structure in the Colombian populations was consistent with other populations in South America. Indeed, a high level of differentiation among populations was found concerning allelic frequencies probably as a consequence of the strong drift effects in small population sizes and founder effects (Mesa *et al.*, 2000).

Concerning the distribution of mtDNA haplogroups in Colombia, two different distributions of mtDNA major haplogroup frequencies were discovered within the country and were interpreted as the consequence of two distinct migratory waves of dispersion into South America. Indeed, one distribution is observed in the northwest Colombia with higher frequencies of haplogroup A and lower frequencies of D and that showed more resemblances to the Central Amerindian populations with whom they maintained contact (Keyeux *et al.*, 2002). The high frequency of haplogroup A in these regions was also interpreted as a consequence of several sequential Chibchan migrations from the Central America into the Pacific coasts of Colombia and backwards (Melton *et al.*, 2007). More recently another study proposes the same Chibchan migrations to be responsible by the introduction of A haplogroup into an ancient Guane population that would have had a predominance of haplogroup B followed by D haplogroup and lacking any C lineage (Casas-Vargas *et al.*, 2011). They concluded that Guane people were originated by two waves: the first would have introduced the B haplogroup and was followed by the introduction of haplogroup A due to the expansion of Chibchan speakers from Central America. These conclusions were very recently corroborated by a whole genome study in Native American populations (Reich *et al.*, 2012).

The other mtDNA haplogroup frequencies distribution is visible in populations from the southeast of the country which showed lower frequency of haplogroup A but higher of haplogroups D and C and was more similar to the pattern seen in other Amerindian populations from South America. Presence of haplogroups C and B was reported in several ancient mtDNA studies concerning remains of different periods: Paleoindians, Herrera and Agroceramic, which indicates the continuity of these haplogroups in Colombian populations (Monsalve *et al.*, 1996; Fernández, 1999; Sánchez, 2007; Silva *et al.*, 2008). Furthermore, a “null haplotype”, later discovered to be assigned to haplogroup C, was found in diverse ethnic and linguistic groups, which suggests that this mutation arose before the separation into tribes and speaking groups (Torres *et al.*, 2006) and indicates the presence of this haplogroup before the formation of ethnic groups.

In regard to the extent of European and African genetic input in Colombian populations, studies clearly demonstrate that the main non-Native American presence in the studied populations was via Y-Chromosome that presented 90% of the chromosomes with European ancestry. These results contrast shockingly with the maternal input, where 90% of the individuals studied presented Native American mtDNA haplogroups. This reflects the fact that, during and after the colonial period, European men migrated into

the New World and crossed preferentially with Native Women. Additionally this miscegenation flow was maintained due to social preference, while the opposite flow was socially refused (Carvajal-Carmona *et al.*, 2000).

2. OBJECTIVES

The study of Colombian populations has recently gained more and more interest from scholars, not only due to the high diversity of ethnic and linguistic groups found in Colombian territory, but also because of its location, proposed as the most likely entrance in South America.

There are still many doubts on the routes taken by the first Native Americans towards South. Even if several reports address mtDNA diversity in Colombian populations, only few analyse the complete CR of the sequences. Therefore, a more polished approach is needed to complement these studies and provide a better insight into the colonization processes and also the demographic events that came to pass in Colombian populations.

In this context, the aims of this work were:

i. To characterize the female lineages from two Colombian regions;
    a. Perceive the proportions of the Native American and the non-Native American (European and African) maternal contributions;
    b. Try to enlighten the history of the groups sampled in the two regions under study;

ii. By comparing with data from literature try to apprehend if:
    a. There are differences among Colombian populations that could indicate different migratory routes taken southwards;
    b. The studied groups reveal geographic or linguistic associations;

iii. To contribute to the enlargement of the mtDNA Forensic Database, EMPOP©.

3. MATERIALS AND METHODS

3.1. Sampling

A total of 98 samples were collected from two different locations within Colombia. In the Department of Antioquia (Segovia) blood samples were taken from 38 Embéra-Chamí Native Americans during regular medical visits. From the DNA bank for parentage tests, 60 Amerindian samples were selected, belonging to various ethnic and linguistic groups from the Department of Cauca. In Figure 8 it is possible to see the exact sampling locations; the A on the map indicates the Department of Antioquia and the B corresponds to the Department of Cauca.

Both samples do not represent the mtDNA lineages of the Departments of Antioquia and Cauca, but represent Amerindian communities that are a minority in their Departments. The Embéra-Chamí population, located in Segovia (Department of Antioquia), more precisely in the Tagua del Pó Indian Reservation, belongs to the Chocó language family. Considering the sample from Cauca, it is composed by several ethnic groups as Yanacona, the Emberá, the Coconucos, the Nasa and the Guambianos, with different languages as Quechua, Chocó and Chibcha and also other languages from Paezan group (Arango & Sánchez, 2004; DANE, 2007; Greenberg & Ruhlen, 2007). Detailed information on each sampled individual can be obtained on Table 2.

Figure 8 - Location of both populations sampled in this study. The letter A describes the population Embéra-Chamí sampled in the Department of Antioquia (Segovia) and the letter B relates to the group sampled in the South of the country, Department of Cauca that is constituted of several ethnic and linguistic groups. Adapted from Google Maps.

Table 2 - Description on the sample and ethnic characteristics of each individual.

SAMPLE	DEPARTMENT	LOCATION	ETHNICITY	LANGUAGE
L01	Antioquia	Segovia	Embéra-Chamí	Embéra
L08	Antioquia	Segovia	Embéra-Chamí	Embéra
L09	Antioquia	Segovia	Embéra-Chamí	Embéra
L11	Antioquia	Segovia	Embéra-Chamí	Embéra
L18	Antioquia	Segovia	Embéra-Chamí	Embéra
L21	Antioquia	Segovia	Embéra-Chamí	Embéra
L24	Antioquia	Segovia	Embéra-Chamí	Embéra
L25	Antioquia	Segovia	Embéra-Chamí	Embéra
L30	Antioquia	Segovia	Embéra-Chamí	Embéra
L31	Antioquia	Segovia	Embéra-Chamí	Embéra
L34	Antioquia	Segovia	Embéra-Chamí	Embéra
L35	Antioquia	Segovia	Embéra-Chamí	Embéra
L36	Antioquia	Segovia	Embéra-Chamí	Embéra
L42	Antioquia	Segovia	Embéra-Chamí	Embéra
L43	Antioquia	Segovia	Embéra-Chamí	Embéra
L48	Antioquia	Segovia	Embéra-Chamí	Embéra
L49	Antioquia	Segovia	Embéra-Chamí	Embéra
L90	Antioquia	Segovia	Embéra-Chamí	Embéra
L51	Antioquia	Segovia	Embéra-Chamí	Embéra
L52	Antioquia	Segovia	Embéra-Chamí	Embéra
L53	Antioquia	Segovia	Embéra-Chamí	Embéra
L55	Antioquia	Segovia	Embéra-Chamí	Embéra
L57	Antioquia	Segovia	Embéra-Chamí	Embéra
L60	Antioquia	Segovia	Embéra-Chamí	Embéra
L63	Antioquia	Segovia	Embéra-Chamí	Embéra
L64	Antioquia	Segovia	Embéra-Chamí	Embéra
L65	Antioquia	Segovia	Embéra-Chamí	Embéra
L70	Antioquia	Segovia	Embéra-Chamí	Embéra
L74	Antioquia	Segovia	Embéra-Chamí	Embéra
L76	Antioquia	Segovia	Embéra-Chamí	Embéra
L77	Antioquia	Segovia	Embéra-Chamí	Embéra
L79	Antioquia	Segovia	Embéra-Chamí	Embéra
L81	Antioquia	Segovia	Embéra-Chamí	Embéra
L82	Antioquia	Segovia	Embéra-Chamí	Embéra
L85	Antioquia	Segovia	Embéra-Chamí	Embéra
L88	Antioquia	Segovia	Embéra-Chamí	Embéra
L89	Antioquia	Segovia	Embéra-Chamí	Embéra
L91	Antioquia	Segovia	Embéra-Chamí	Embéra
P1531	Cauca	Catrú	Embera	Embera
P6208	Cauca	Bolivar	Embera-Chami	Embera
P2941	Cauca	Popayan	Nasa	Chibcha
H2941	Cauca	Popayan	Nasa	Chibcha

P3634	Cauca	Popayan	Nasa	Chibcha
H3634	Cauca	Popayan	Nasa	Chibcha
P3791	Cauca	El Tambo	Nasa	Chibcha
P3654	Cauca	Silvia	Guambiano	Guambiano
H3654	Cauca	Silvia	Guambiano	Guambiano
P3666	Cauca	Puracé	Coconuco	Guambiano
H3666	Cauca	Puracé	Coconuco	Guambiano
P3699	Cauca	-	-	-
H3699	Cauca	-	-	-
P3728	Cauca	-	-	-
H3728	Cauca	-	-	-
H3791	Cauca	El Tambo	Nasa	Chibcha
P3839	Cauca	Jambalo	Nasa	Chibcha
P3792	Cauca	Puracé	Coconuco	Guambiano
H3792	Cauca	Puracé	Coconuco	Guambiano
P3807	Cauca	Puracé	Coconuco	Guambiano
H3807	Cauca	Puracé	Coconuco	Guambiano
P3826	Cauca	Silvia	Guambiano	Guambiano
P5279	Cauca	Paez	Nasa	Chibcha
P4213	Cauca	Silvia	Guambiano	Guambiano
H4213	Cauca	Silvia	Guambiano	Guambiano
P5324	Cauca	Paez	Nasa	Chibcha
P5679	Cauca	Paez	Nasa	Chibcha
P4288	Cauca	Silvia	Guambiano	Guambiano
H4288	Cauca	Silvia	Guambiano	Guambiano
H4453	Cauca	Puracé	Coconuco	Guambiano
P4557	Cauca	Puracé	Coconuco	Guambiano
PP4645	Cauca	Puracé	Coconuco	Guambiano
H5679	Cauca	Paez	Nasa	Chibcha
P5942	Cauca	El Tambo	Nasa	Chibcha
P4803	Cauca	Silvia	Guambiano	Guambiano
P4985	Cauca	Silvia	Guambiano	Guambiano
P5076	Cauca	Puracé	Coconuco	Guambiano
H5076	Cauca	Puracé	Coconuco	Guambiano
H6564	Cauca	Popayan	Nasa	Chibcha
P6750	Cauca	El Tambo	Nasa	Chibcha
P5325	Cauca	Silvia	Guambiano	Guambiano
P2 4668	Cauca	Ipiales	Pastos	Chibcha
P3633	Cauca	La Vega	Yanacona	Quechua
H3633	Cauca	La Vega	Yanacona	Quechua
P6013	Cauca	Silvia	Guambiano	Guambiano
P6056	Cauca	Silvia	Guambiano	Guambiano
P4214	Cauca	Sotará	Yanacona	Quechua
P6336	Cauca	Silvia	Guambiano	Guambiano
H6336	Cauca	Silvia	Guambiano	Guambiano
P6337	Cauca	Silvia	Guambiano	Guambiano

P6427	Cauca	Silvia	Guambiano	Guambiano
H6427	Cauca	Silvia	Guambiano	Guambiano
P6428	Cauca	Silvia	Guambiano	Guambiano
H6428	Cauca	Silvia	Guambiano	Guambiano
P6429	Cauca	Piendamo	Guambiano	Guambiano
P4286	Cauca	La Vega	Yanacona	Quechua
P1 4668	Cauca	La Vega	Yanacona	Quechua
P6751	Cauca	Silvia	Guambiano	Guambiano
H6767	Cauca	San Sebastian	Yanacona	Quechua
P6793	Cauca	Silvia	Guambiano	Guambiano

3.2. DNA Analysis

3.2.1. DNA Amplification and Sequencing

J. J. Builles and J. M. Ospino from the University of Antioquia, Colombia, collected blood samples in FTA cards from 98 individuals, who inhabit the 2 different regions in Colombia described above.

DNA was extracted from all samples following the Chelex (Bio-Rad Laboratories, Inc) protocol, and the entire mtDNA Control Region (16024-16569 e 1-576) was amplified using the primers L15978 (5'-CACCATTAGCACCCAAAGCT-3') and H639 (5'-GGGTGATGTGAGCCCGTCTA-3') under the following PCR conditions: 15' at 95°C for initial denaturation, after that 35 cycles of 30'' at 94°C for denaturation, 90'' under 58ºC for annealing and the elongation for 90'' under 72°C, followed by a final extension for 10' under the temperature of 72°C.

PCR mix was composed by 1µL of primer mix (2µL of each primer forward and reverse from the initial stocks at 100µM of concentration, and 96µL of water, in a final concentration of 2µM), 5µL of Quiagen PCR Kit (2x), 3µL of H_2O and 1µL of DNA. Amplification was verified by Polyacrilamide gel (T9%, C5%) electrophoresis and silver nitrate stain, following standard protocol.

The sequencing procedures started with a previous purification of the PCR product using ExoSAP-IT (GE Healthcare) in the thermocycler for 15' at the temperature of 37°C followed by 15' at 80°C. To the 1.5µL of purified PCR product, we added 0.5µL of primer (from an initial concentration of 2.5µM) corresponding to the desired fragment (Table 3), 1µL of the Big Dye Terminator v3.1 cycle Sequencing Kit (Applied Biosystems®), 1µL of buffer (Applied Biosystems®) and 1µL of H_2O. The mtDNA was

sequenced for L-strand (forward) in all samples, and in cases where a length heteroplasmy or poli-C tracts were found the H-strand (reverse) was also sequenced.

Afterwards, the sequencing reaction was conducted under the following PCR conditions: initial denaturation of 2‘ at 96°C, followed by 35 cycles of 15‘‘at 96°C and 60°C for 2‘, and a final extension for 10‘at 60°C.

After the sequencing reaction, the samples were purified with Illustra Sephadex™ DNA Grade (GE Healthcare) following the regular protocol, before being run on the automatic sequencer ABI 3130XL (Genetic Analyser 3000®, Applied Biosystems®).

Table 3 - Primers used for the regions analysed in each sequencing reaction. Note that L-strand fragments (reverse sequences) were only analysed when a heteroplasmy or slippage due to a poli-C tract occurred.

REGION TO ANALYSE	PRIMER
HVI Forward	L15978 (5’-CAC CAT TAG CAC CCA AAG CT-3’)
HVII Forward	L16536 (5’-CCC ACA CGT TCC CCT TAA AT-3’)
HVI Reverse	H036 (5’-CCC GTG AGT GGT TAA TAG GGT-3’)
HVII Reverse	H639 (5’-GGG TGA TGT GAG CCC GTC TA-3’)

Primer designations are labeled according to the 5’ end position of each primer in the mtDNA rCRS sequence.

3.2.2. Haplotypes and Haplogroups discrimination

The sequences were analysed using the software *Geneious Pro 5.5.3*® (Drummond *et al.*, 2011) by assembling the sequences against the revised Cambridge Reference Sequence – rCRS (Andrews *et al.*, 1999) - leading to the identification of the polymorphisms and consequently the description of the haplotypes. This analysis was performed using the phylogenetic approach suggested by Bandelt & Parson (2008) and the haplotypes followed the nomenclature described in Carracedo *et al.* (2000), adopted by the International Society for Forensic Genetics (ISFG).

Haplogrep software (Kloss-Brandstätter *et al.*, 2011) was used to assign the sequences into haplogroups according to the most recently updated version of the mtDNA phylogenetic tree, *Phylotree 2012* (van Oven & Kayser, 2009).

The haplotypes were submitted to the EMPOP© Forensic and Population Genetics that analysed and checked them as well as confirmed the haplogroup classifications. EMPOP© mtDNA database (Parson *et al.*, 2004) follows the criteria established by the International Society for Forensic Genetics, namely for the nomenclature rules and phylogenetic approach (Bandelt & Parson, 2008; Carracedo *et al.*, 2000). EMPOP© was created to satisfy the urgent need of a high quality mtDNA database for forensic

purposes but allows an analysis of the quality of the haplotypes and of the haplogroup classification before publication (Parson *et al.*, 2004).

3.3. Data Analysis

3.3.1. Comparative data

In order to do a comparative analysis with our study groups we collected sequences from publicly available literature on mtDNA (HVRI, HVRII and complete Control Region) from Colombian populations and South, Central and North Amerindians (Table 4).

Table 4 - Description of the literature data collected for comparison purposes. The ethnic group, country and language groups are described as well as the number of individuals and the region of mtDNA analysed. Language groups are according to those used in Yang *et al.* (2010).

PAPER	REGION/POPULATION	COUNTRY	LANGUAGE (MACRO-FAMILY)	N	ANALYSED REGION
Melton *et al.* (2007)	Kogui	Colombia	Chibchan-Paezan	48	HVRI
	Ijka	Colombia	Chibchan-Paezan	40	HVRI
	Arsario	Colombia	Chibchan-Paezan	50	HVRI
	Wayúu	Colombia	Arawak	50	HVRI
Torres *et al.* (2006)	Cubeo	Colombia	Equatorial Tucanoan	4	HVRI + HVRII
	Curripaco	Colombia	Equatorial Tucanoan	2	HVRI + HVRII
	Desano	Colombia	Equatorial Tucanoan	1	HVRI + HVRII
	Embéra	Colombia	Chibchan-Paezan	4	HVRI + HVRII
	Guahibo	Colombia	Equatorial Tucanoan	4	HVRI + HVRII
	Huitoto	Colombia	Ge-Pano-Carib	7	HVRI + HVRII
	Ingano	Colombia	Andean	1	HVRI + HVRII
	Jebero	Colombia	Andean	1	HVRI + HVRII
	Ocaina	Colombia	Ge-Pano-Carib	4	HVRI + HVRII
	Paez	Colombia	Chibchan-Paezan	2	HVRI + HVRII
	Piapoco	Colombia	Equatorial Tucanoan	11	HVRI + HVRII
	Puinave	Colombia	Equatorial Tucanoan	5	HVRI + HVRII
	Saliva	Colombia	Equatorial Tucanoan	3	HVRI + HVRII
	Ticuna	Colombia	Equatorial Tucanoan	4	HVRI + HVRII
	Wayúu	Colombia	Equatorial Tucanoan	3	HVRI + HVRII
	Yagua	Colombia	Ge-Pano-Carib	3	HVRI + HVRII
	Zenu	Colombia	?	5	HVRI + HVRII
Salas *et al.* (2009)	Departament of Cauca	Colombia	Several	98	HVRI
Yang *et al.* (2010)	Ache	Brazil	Equatorial Tucanoan	11	CR
	Arhuaco	Colombia	Chibchan-Paezan	16	CR
	Aymara	Chile	Andean	17	CR
	Cabecar	Costa Rica	Chibchan-Paezan	15	CR
	Cree	Canada	Northern Amerind	11	CR
	Embéra	Colombia	Chibchan-Paezan	9	CR

Guarani	Brazil	Equatorial Tucanoan	8	CR
Guaymi	Costa Rica	Chibchan-Paezan	13	CR
Huilliche	Chile	Andean	20	CR
Inga	Colombia	Andean	16	CR
Kogi	Colombia	Chibchan-Paezan	16	CR
Mixe	Mexico	Central Amerind	20	CR
Mixtec	Mexico	Central Amerind	17	CR
Ojibwa	Canada	Northern Amerind	16	CR
Quechua	Peru	Andean	18	CR
Kaqchikel	Guatemala	Central Amerind	16	CR
Ticuna	Colombia	Equatorial Tucanoan	12	CR
Waunana	Colombia	Chibchan-Paezan	18	CR
Wayúu	Colombia	Chibchan-Paezan	18	CR
Zapotec	Mexico	Central Amerind	19	CR
Zenu	Colombia	Chibchan-Paezan	15	CR
Kaingang	Brazil	Ge-Pano-Carib	2	CR

A total of 272 mtDNA HVRI and HVRII sequences from Colombian populations were compiled, suggesting a rather wide sampling throughout the country (Melton *et al.*, 2007; Torres *et al.*, 2006; Salas *et al.*, 2008). Furthermore, we were also able to collect data from several Amerind ethnic and linguistic groups from South, Central and North America. A paper from Yang *et al.* (2010) contained the sequences from the complete CR of 323 Amerindian individuals widely sampled throughout America (Yang *et al.*, 2010). Details on sampling can be observed on Table 4.

All sequences were re-classified into haplogroups by *Haplogrep* software (Kloss-Brandstätter *et al.*, 2011) and the updated version of *Phylotree 2012 (van Oven & Kayser, 2009)*. InDels from literature data (Yang *et al.*, 2010) that appeared to be result from sequencing errors and incomplete sequences were ignored, except for 290DEL, 291Del and 498DEL that are important for haplogroup definition.

3.3.2. Intra and Inter-population analysis

In order to promote a coherent study, only the individuals carrying a Native-American haplogroup were considered for subsequent analyses. The haplogroup frequencies were calculated by direct counting. Data collected from the literature were organized under two different criteria, one concerning the geographic region and the other one related to the language, as in Yang *et al.* (2010). All analyses that included our data and data from literature considered only the mtDNA region shared by all sequences.

Diversity indices were calculated not only for our study populations, but also for comparison populations using *Dnasp v.5* software (Librado & Rozas, 2009).

Accordingly, we calculated the nucleotide diversity (π) (Tajima, 1983; Nei, 1987), haplotype diversity (H) and the number of segregating sites (S).

Network analyses were performed using *NETWORK v.4.6.1.0* software (Fluxus Technology, 2004-2007) and applying the reduced median and the median joining methods (Bandelt *et al.*, 1995; Bandelt *et al.*, 1999) sequentially in order to minimize reticulation. These phylogenetic analyses were performed to check for haplotype sharing and to confirm the haplogroup classification within our groups and also with comparative data. In all networks the positions 16182C, 16183C, 16519C, 309.1C, 309.2C and 315.1C were excluded since they are not considered for haplogroup discrimination.

For comparative purposes, population pairwise F_{ST} genetic distances were calculated by *Arlequin v.3.5* software (Excoffier & Lischer, 2010) with haplogroup frequencies data. Bonferroni correction was applied by dividing α (0.05) by the number of pairwise population comparisons. In order to have a better perception of the genetic distances based on F_{ST} values, a MDS (MultiDimensional Scaling) analysis, representing the two dimensional spatial plans, was accomplished with the *SPSS v.20* software (SPSS, 2001).

4. RESULTS

To enable an easier understanding and discussion, the results presented in this section are organized according to the analyses performed. The analyses contemplate all the comparative stages, starting with a description of our results followed by a comparison with data from other groups within Colombia and finally the comparison between our data and other data from the Americas.

Ninety eight mtDNA control region haplotypes were identified and deposited in a Forensic and Population Genetic mtDNA database, EMPOP©. Details on the polymorphisms and haplogroups attributed to each sample can be obtained in appendix.

4.1. Haplogroup Frequencies

The majority of haplogroups found in our samples are typically Native American (A, B, C and D). Frequencies on both geographic groups can be observed in Figure 9 and in Table 5 (see Antioquia and Cauca).

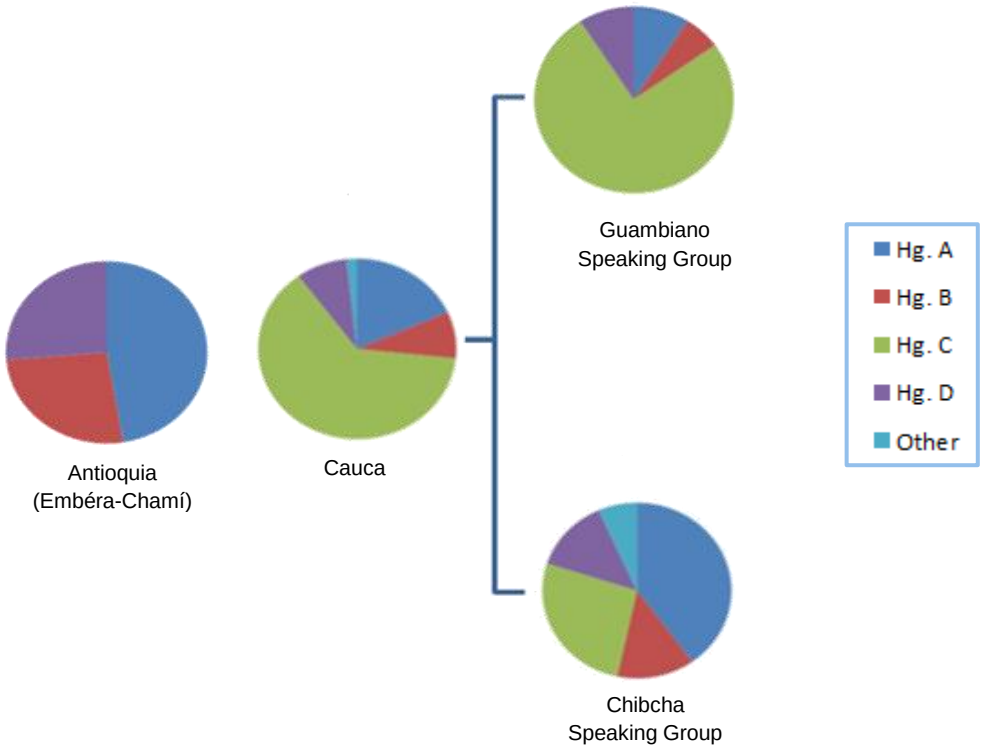


Figure 9 - Frequencies of the major mtDNA haplogroups (Hg) of both geographic regions sampled and analysed on this study (Antioquia and Cauca). Subgroups of the Cauca main group and their haplogroup (Hg) frequencies include linguistically associated individuals and are named Chibcha speaking group and Guambiano speaking group. Details on sampling can be found in 3. Materials and Methods chapter.

Table 5 - Haplogroup frequencies in two Colombian regions sampled (Cauca and Antioquia) and in the two linguistic subgroups from Cauca (Chibcha and Guambiano).

GROUP	N	HAPLOGROUP				
		A	B	C	D	OTHER
Antioquia (Embéra)	38	0,474	0,263	0,000	0,263	0,000
Cauca	60	0,183	0,083	0,633	0,083	0,017
Chibcha	15	0,400	0,133	0,267	0,133	0,067
Guambiano	33	0,091	0,061	0,756	0,091	0,000

From the mtDNA haplogroup frequencies in Antioquia and Cauca regions represented in Figure 9 and Table 5, it is clear that these two populations are markedly different. In fact, while the Antioquia sample reveals a high frequency of haplogroup A (47.4%) followed by haplogroups B (26.3%) and D (26.3%) and an absence of C haplogroup, the Cauca region reveals the opposite by presenting the haplogroup C (63.3%) as the most frequent, followed by A (18.3%), B (8.3%) and D (8.3%) at lower frequencies. Additionally, in Cauca a single case of a haplotype belonging to a non-Native American haplogroup was found - L2a1c1 (1.7%) which is originated from L2a that is typically African and very frequent in Bantu populations (Salas *et al.*, 2002; Salas *et al.*, 2004).

Because Cauca sample includes individuals from various ethnic and linguistic groups (see section 3.1. Sampling for details) it is possible to divide it into linguistic subgroups. Accordingly, two subgroups were defined, namely Chibcha speaking group (N=15) and Guambiano speaking group (N=33). Furthermore, there are evident differences in the haplogroup frequencies between the subgroups: while Guambiano group presents a high frequency of haplogroup C (75.6%) and minor frequencies of the remaining haplogroups (A-9.1%; B- 6.1%; D- 9.1%; Others- 0%); Chibcha group presents a high frequency of haplogroup A (40.0%) followed by haplogroup C (26.7%) and minor frequencies of haplogroups B (13.3%), D (13.3%) and Others - L2a1c1 (6.7%), which is visible in Figure 9 and Table 5.

In order to contextualize the previous results in a broader scale, representations of the distribution of the haplogroup frequencies in Native American populations throughout the American continent were performed (Figures 10 and 11).

In Figure 10, there is a representation of the frequencies of the major haplogroups in the populations analysed from literature as well as from the present study grouped in geographic regions. In the same figure, within the green box, there is a representation of the haplogroup frequencies in the Colombian territory.

Keyeux *et al.* (2002) found two profiles of mtDNA haplogroup distribution within Colombia, one characteristic of northern populations that presented a high frequency of haplogroup A, also seen in the present results (Antioquia, PS) and in other literature populations (Keyeux *et al.*, 2002; Melton *et al.*, 2007). The second profile was found in southern populations with a prevalence of haplogroup C also seen in the present results (Cauca, PS) and in Salas *et al.* (2008). There are notorious differences in the haplogroup frequencies from the groups sampled in the North of the country and those sampled in the southern areas.

The analyses of the groups dispersed along the American continent allow a broader perspective. In Figure 10, it is visible a decrease in haplogroup A frequencies from the Centre to the South of the American continent together with the increase in the haplogroups C and D frequencies which agrees with the results presented in Salas *et al.* (2009). Haplogroup B shows higher values in the Andean region, as previously observed (Fuselli *et al.*, 2003).

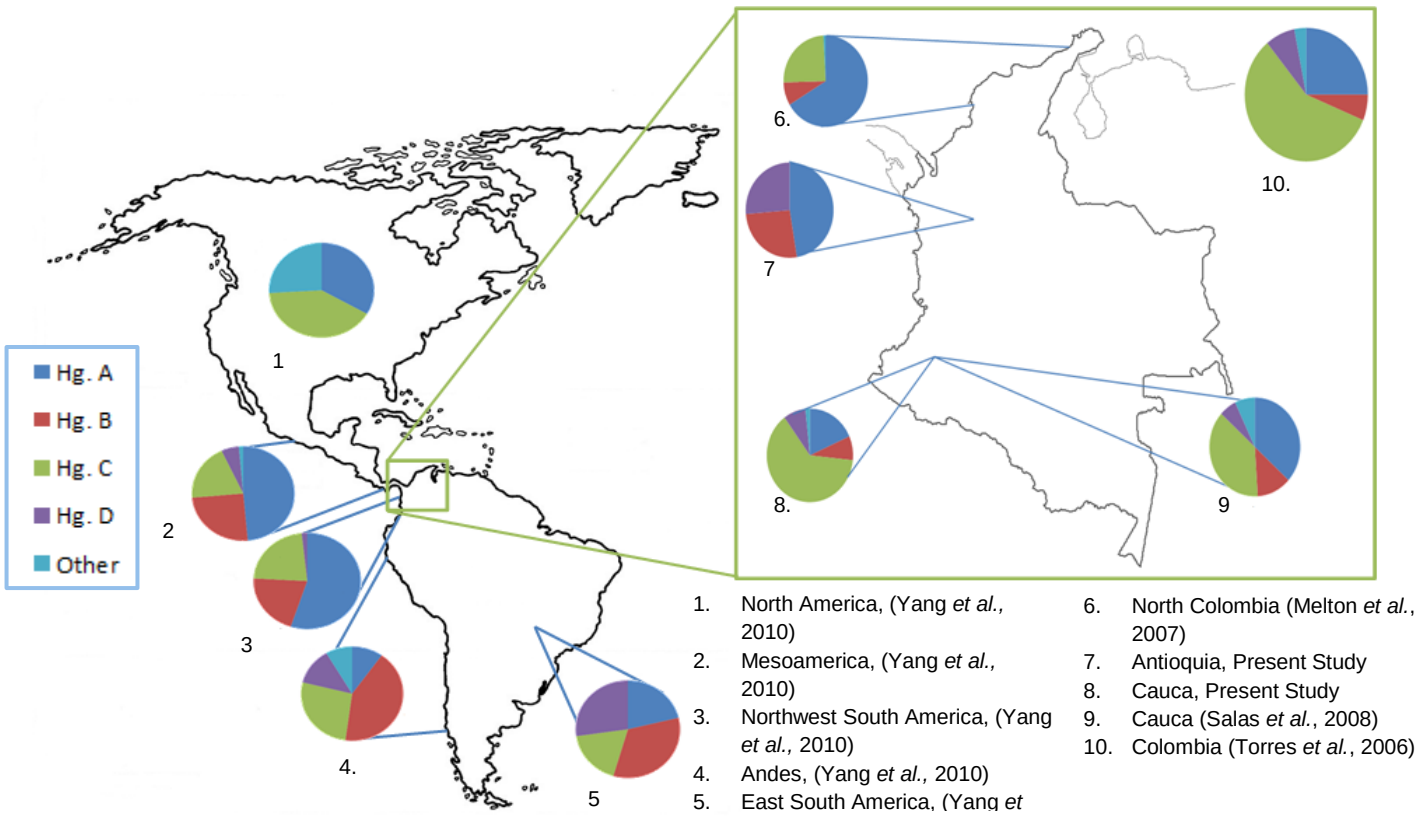
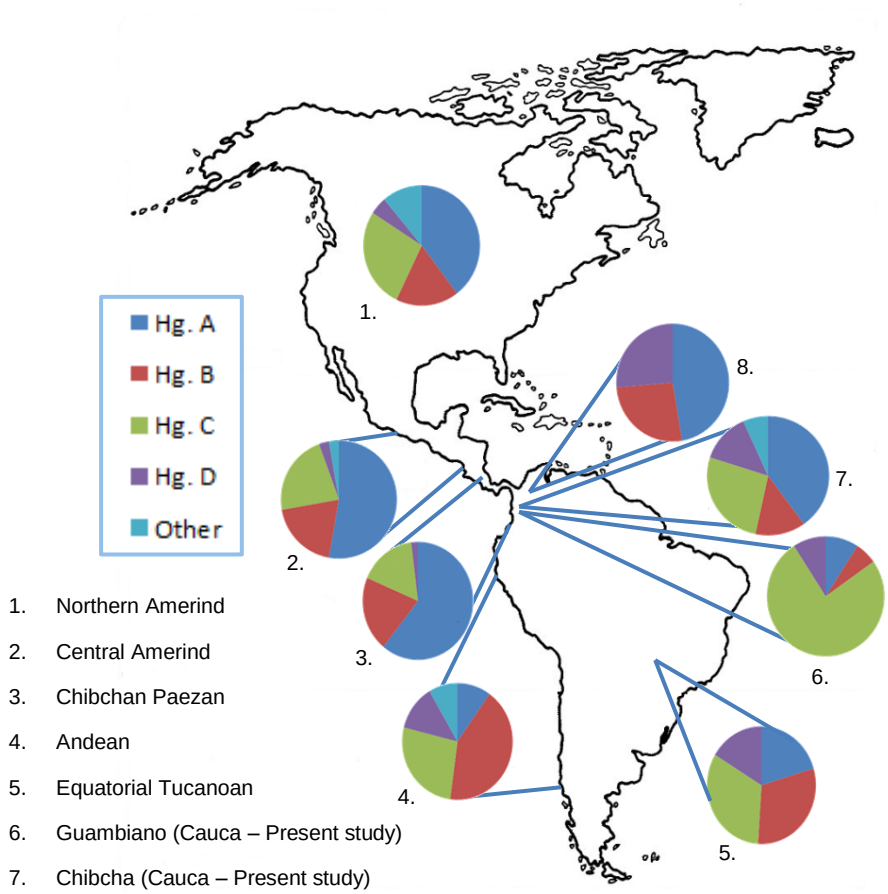


Figure 10 - Distribution of the haplogroup frequencies in the American continent and within the Colombian country (within green box), following a geographic criterion.

Applying a linguistic criterion (represented in Figure 11) and considering Embéra language – Choco - associated with Chibcha family as shown in the works of Campbell (2000) and Constenla & Margery (1991), it is visible that the three Chibchan speaking groups, namely the Chibchan-Paezan Embéra-Chamí population from Antioquia and Chibcha speaking group from Cauca, show a high frequency of haplogroup A. On the other hand these groups differ in haplogroups B and D frequencies, which are much less represented in the sample from Cauca. On the contrary, Guambiano presents a high frequency of haplogroup C which resembles Equatorial Tucanoan.



1. Northern Amerind
2. Central Amerind
3. Chibchan Paezan
4. Andean
5. Equatorial Tucanoan
6. Guambiano (Cauca – Present study)
7. Chibcha (Cauca – Present study)
8. Embéra-Chamí (Antioquia –Present study)

Figure 11 - Distribution of haplogroup frequencies in the American country, gathered from reference (Yang *et al.*, 2010), and the present study's data, following a linguistic criterion.

4.2. Genetic Distances

Population comparisons were made regarding two criteria: geographic regions and linguistic affiliation. Results regarding both criteria can be observed in Tables 6-8.

Table 6 - Pairwise F_{ST} genetic distances (below the diagonal) based on haplogroup frequencies for geographic groups within Colombia, analysed for HVRI.

	Cauca PS	Antioquia PS	North Colombia[2]	Colombia[1]	Cauca[3]
Cauca PS		0.00000	0.00000	0.70270	0.01802
Antioquia PS	0.30543		0.00000	0.00000	0.00000
North Colombia[2]	0.26515	0.14324		0.00000	0.00000
Colombia [1]	0.00000	0.26444	0.20768		0.01802
Cauca [3]	0.06252	0.12315	0.07964	0.03402	

PS stands for present study, groups with number 1 are from (Torres et al., 2006), with 2 from (Melton et al., 2007) and with 3 from (Salas et al., 2008) . P-values can be found above the diagonal. Significant values (P-value<0.05) are coloured in blue, significant values after Bonferroni correction (P-value<0.005) are coloured in red.

Table 7 - Pairwise F_{ST} genetic distances (below the diagonal) based on haplogroup frequencies for geographic groups in the American Continent, analysed for the CR.

	Antioquia PS	Cauca PS	North America[1]	Mesoamerica[1]	Northwest South America[1]	Andes[1]	East South America[1]
Antioquia PS		0.00000	0.00000	0.03223	0.00195	0.00000	0.05566
Cauca PS	0.31300		0.09668	0.00000	0.00000	0.00000	0.00000
North America[1]	0.24489	0.04983		0.00586	0.00781	0.00000	0.00000
Mesoamerica[1]	0.04049	0.20004	0.10847		0.59766	0.00000	0.01074
Northwest South America[1]	0.07386	0.21023	0.09455	0.00000		0.00000	0.00098
Andes[1]	0.15500	0.17126	0.22470	0.12287	0.17242		0.14844
East South America[1]	0.04527	0.18499	0.19445	0.06848	0.11870	0.01754	

PS stands for Present Study, groups with number 1 are from (Yang et al., 2010). P-values can be found above the diagonal. Significant values (P-value<0.05) are coloured in blue, significant values after Bonferroni correction (P-value<0.00238) are coloured in red.

Concerning a geographic criterion two analyses were performed, one only with Colombian regions and another one with main American geographic regions.

An overall analysis of Colombian regions (Table 6) shows that these regions are strongly differentiated. The higher value of F_{ST} is obtained for the Cauca PS and Antioquia PS pair which highlights the differences already discussed between these two groups (see 4.1. Haplogroup Frequencies and 4.3. Phylogeographic Analysis). These differences can result from the fact that the Antioquia's sampling is from one ethnic population that shows signs of isolation as can be also seen in other pairwise comparisons with high and significant F_{ST} values, whereas the Cauca's sample is a

mixture of individuals from various ethnic groups that are differentiated between them. Cauca PS is not statistically differentiated from Colombia (Torres *et al.*, 2006) which is a consequence of the high frequency of haplogroup C, but presents significantly high F_{ST} values with North Colombia, whereas the not significant F_{ST} value (after Bonferroni correction) obtained for the comparison with Cauca is probably a result of a small sampling size.

Regarding the division in main American geographic regions, populations from Yang *et al.* (2010) were divided into 4 main groups: North America (Cree and Ojibwa), Mesoamerica (Mixe, Mixtec, Zapotec and Kaqchikel), Northwest South America (Arhuaco, Cabecar, Embera, Guaymi, Kogi, Waunana, Wayúu and Zenu), Andes (Aymara, Huilliche, Inga and Quechua) and East South America (Aché, Guarani, Kaingang and Ticuna). The present study's data was also organized under a geographic criterion, in Antioquia PS and Cauca PS. Pairwise comparisons F_{ST} are detailed on Table 7 and can be visualized under a MDS plot (Multidimensional Scaling) in Figure 12.A.

In Table 7 and Figure 12.A, it is observable that our study regions present high statistically significant F_{ST} values in almost all comparisons. There is a separation between Mesoamerica, northern South America and Antioquia (Figure 12.A) from the regions of Andes and East South America. Cauca and North America remain distant from all groups (Figure 12.A).

Table 8 - Pairwise F_{ST} genetic distances based on haplogroup frequencies for linguistic groups in the American Continent, analysed for the CR.

	Embéra PS	Chibcha PS	Guambiano PS	North Amerind[1]	Central Amerind[1]	Chibchan-Paezan[1]	Andean[1]	Equatorial Tucanoan[1]
Embéra PS		0.14746	0.00000	0.00586	0.02734	0.00977	0.00000	0.00195
Chibcha PS	0.03299		0.00195	0.89453	0.67969	0.27148	0.02539	0.34766
Guambiano PS	0.41090	0.22249		0.00000	0.00000	0.00000	0.00000	0.00000
North Amerind[1]	0.07704	0.00000	0.21749		0.63965	0.06152	0.00000	0.03809
Central Amerind[1]	0.05911	0.00000	0.31102	0.00000		0.72363	0.00000	0.00977
Chibchan-Paezan[1]	0.06880	0.01500	0.36947	0.02361	0.00000		0.00000	0.00000
Andean[1]	0.15500	0.09447	0.22898	0.11092	0.15570	0.21281		0.23730
EquatorialTucanoan[1]	0.10065	0.00702	0.15913	0.03799	0.07885	0.13840	0.00689	

PS stands for Present Study, groups with number 1 are from (Yang et al., 2010). P-values can be found in annex. Significant values (P-value<0.05) are coloured in blue, significant values after Bonferroni correction (P-value<0.00179) are coloured in red.

Another criterion adopted was the linguistic affiliation of the populations. Five distinct macro-families were considered: Northern Amerind (Cree, Ojibwa, Mixe and Kaqchikel), Central Amerind (Mixtec and Zapotec), Chibchan-Paezan (Arhuaco, Cabecar, Embera, Guaymi, Kogi, Waunana and Zenu), Andean (Aymara, Huilliche, Inga and Quechua) and Equatorial Tucanoan (Aché, Guarani, Ticuna and Wayúu) following Campbell's classification (Campbell, 2000). Present study's data were also organized under a linguistic criterion; therefore Antioquia's population Embéra PS and Cauca's subgroups Guambiano PS and Chibcha PS were considered. Pairwise genetic distances F_{ST} were performed and can be observed on Table 8 and Figure 12.B.

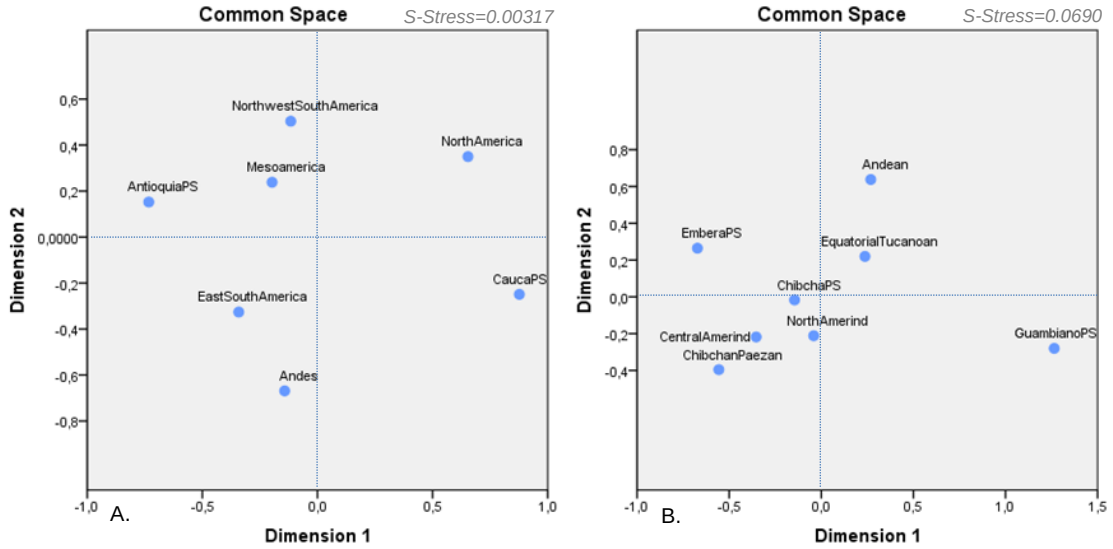


Figure 12 - A: MDS plot of the F_{ST} genetic distances between the 7 groups under a geographic criterion analysed for the haplogroup frequencies (S-Stress=0.00317). B: MDS plot of the F_{ST} genetic distances between the 8 groups under a linguistic criterion analysed for the haplogroup frequencies (S-Stress=0.00690).

The non-significant F_{ST} values are probably a consequence of the small number of individuals analysed in these populations. Embéra PS and Chibcha PS pair presents a not-significant F_{ST} value of 0.03299. Moreover, most of the not-significant F_{ST} values were found in pairs involving Embéra PS and Chibcha PS which have a small number of individuals analysed and therefore influence the statistically significance of the F_{ST} values.

Guambiano PS has revealed a strong genetic distance from the other groups with all values significant. The highest genetic distance is found with Embéra PS due to the high divergence in haplogroup frequencies (see 4.1 Haplogroup Frequencies section). Guambiano PS is also highly distant from Chibchan-Paezan (0.36323) and Central Amerindian (0.29480) groups. This is not a surprising result if we take into account that Guambiano language was classified in the Paezan group (Campbell, 2000; Greenberg

& Ruhlen, 2007) but the majority of populations inserted in the Chibchan-Paezan group are Chibchan or Chocoan speakers (and not Paezan speakers).

Additional population comparison analyses (Table 11, in appendix) and MDS (Figure 13) were performed detailing the relation between the study groups under a linguistic affiliation (Embéra PS, Chibcha PS and Guambiano PS) and populations sampled in Northwest South America with linguistic affiliations. In Figure 13 it is observable that Embéra PS remains close to Chibchan and Chocoan speaking groups (Waunana – Chocoan; Cabecar – Chibcha; Guaymi – Chibcha) and is closer to Central American populations as Cabecar and Guaymi. Chibcha PS has a central location between two differentiated groups and Guambiano PS stays close to populations with not well defined languages (Zenu, an extinct language, here classified as Chibcha-Paezan is also classified as Carib by some scholars, Mesa *et al.,* 2000) and Wayúu, inserted in Equatorial-Tucanoan macro-family. The cause for this divergence is the high frequencies of haplogroup A in all Chibchan speakers, which can be in fact a characteristic of this linguistic group (Casas-Vargas *et al.*, 2011) while in the other populations (Wayúu, Zenu and Guambiano PS) the haplogroup with highest frequency is C.

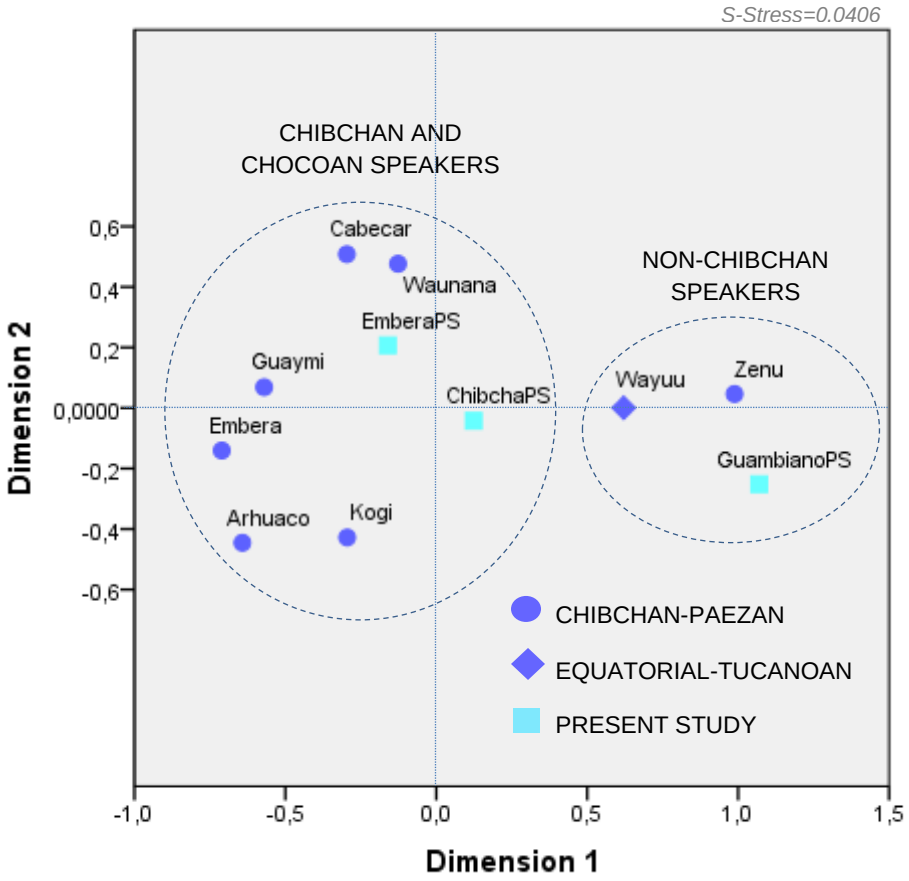


Figure 13 - MDS plot of the F_{ST} genetic distances with the populations from Northwest South America with linguistic affiliation, adapted from (Yang *et al.,* 2010). (S-Stress=0.0406).

4.3. Phylogeographic Analysis

In order to perform a phylogenetic analysis, reduced median and median joining networks were performed for A, B and C haplogroups and for the totality of haplogroups using the present samples and relevant samples from literature.

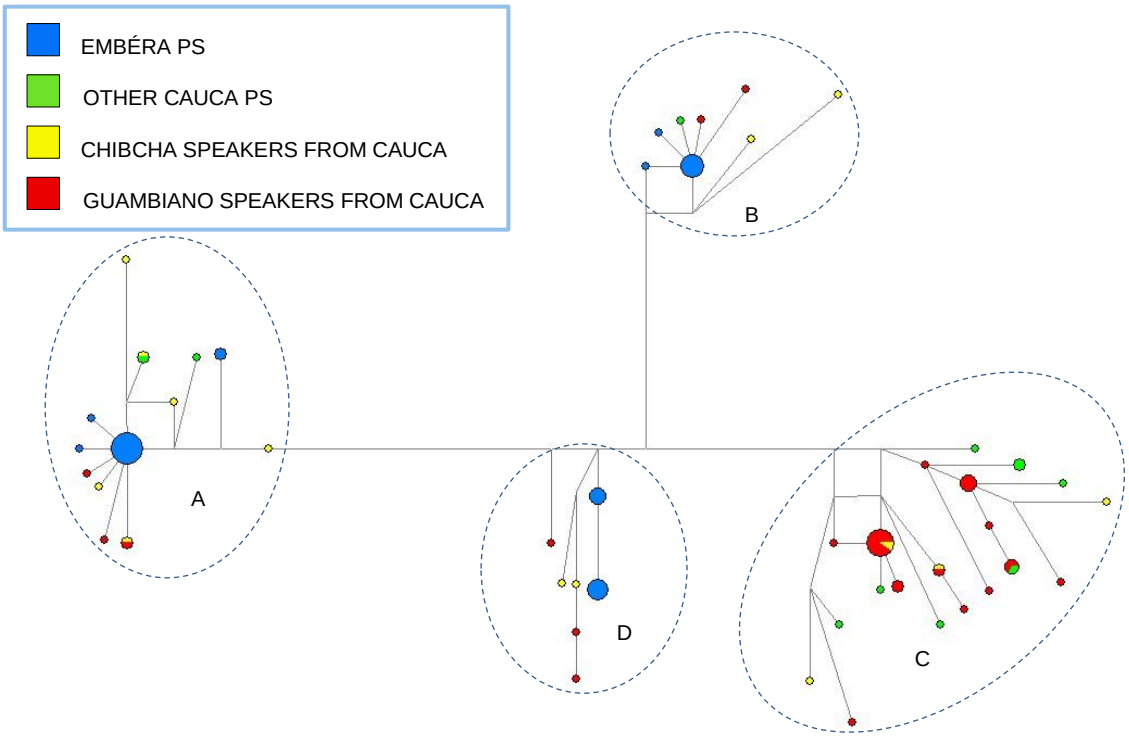


Figure 14 - Median joining network of CR data from the present study. Circle sizes are proportional to the haplotype frequencies.

Firstly a network analysis was done with the present data to confirm haplogroup classification and check haplotype sharing between the two regions sampled (Figure 14). It is visible that haplotypes group into haplogroups which allows confirming their classification. Antioquia region reveals to have fewer haplotypes than Cauca per haplogroup and also presents less diverse lineages. Furthermore, there is no haplotype sharing between both which can be explained by the small number of individuals analysed per sample. These results also indicate that the population within Antioquia group is probably under isolation. However, Cauca group is composed of individuals from several ethnicities and languages which could lead to a biased increase of diversity. When considering the Cauca sample divided into two major linguistic groups (Chibcha, N=15 and Guambiano, N=33) (Figure 14), private haplotypes seem to emerge associated to speaking groups and haplotype sharing is little between the different groups.

Afterwards, network analyses were made including the results from the present study and literature data from Colombia that was sampled in regions geographically close in order to check haplotype sharing with neighbouring populations. Two analyses were performed: one regarding northern Colombia and Antioquia region and the other one with southern Colombia and Cauca region from the present study (Figures 15 and 16, respectively).

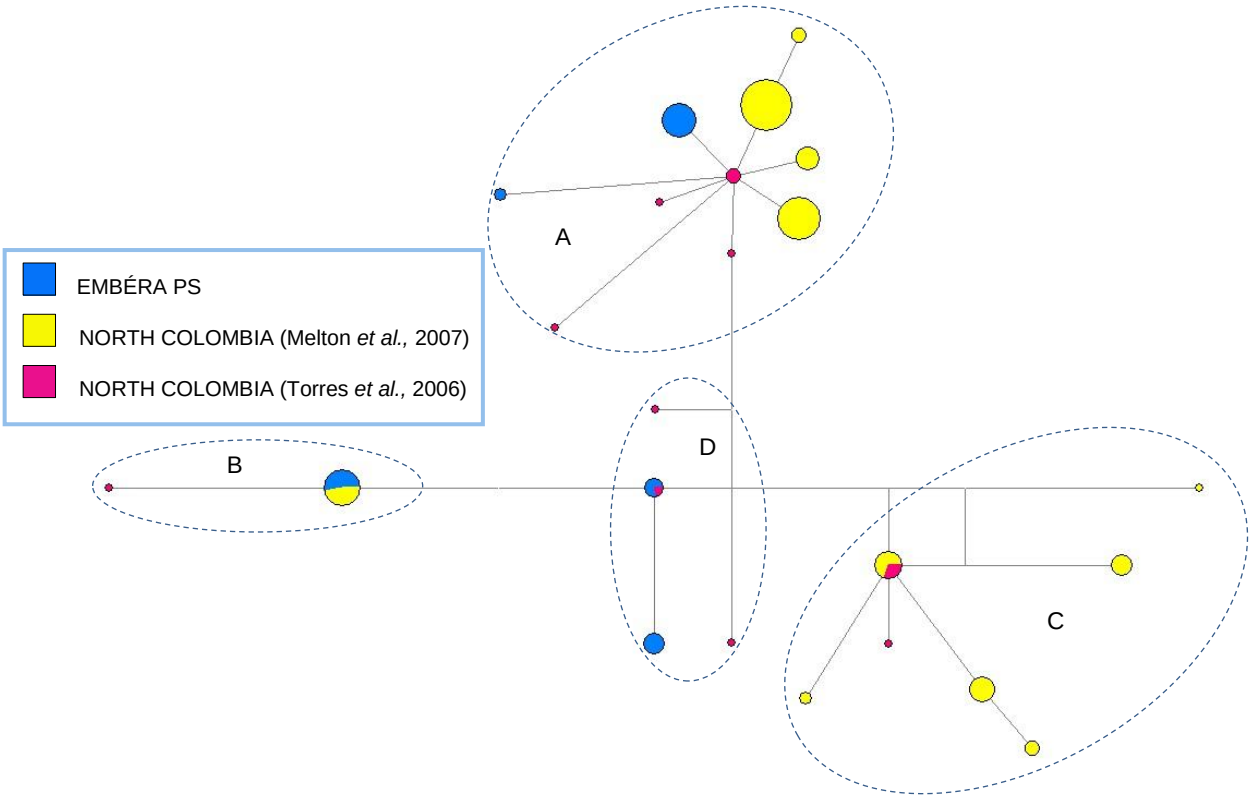


Figure 15 - Median joining network analysis based on HVRI of the Antioquia data from present study and data gathered from the literature from groups sampled in North Colombia. Circle sizes are proportional to the haplotype frequencies.

When analysing the northern region (Figure 15) it is possible to see that the same pattern of small populations is maintained with low number of shared haplotypes and low number of haplotypes per haplogroup. This pattern may corroborate that the Embéra-Chamí population within Antioquia group is under isolation, or it may just be the result of the small sample sizes.

In the case of the southern region (Figure 16) there is a higher number of haplotypes and many are shared with populations form the literature. Such results corroborate the diversity indices obtained in the chapter 4.4 where southern groups are more diverse than the northern ones.

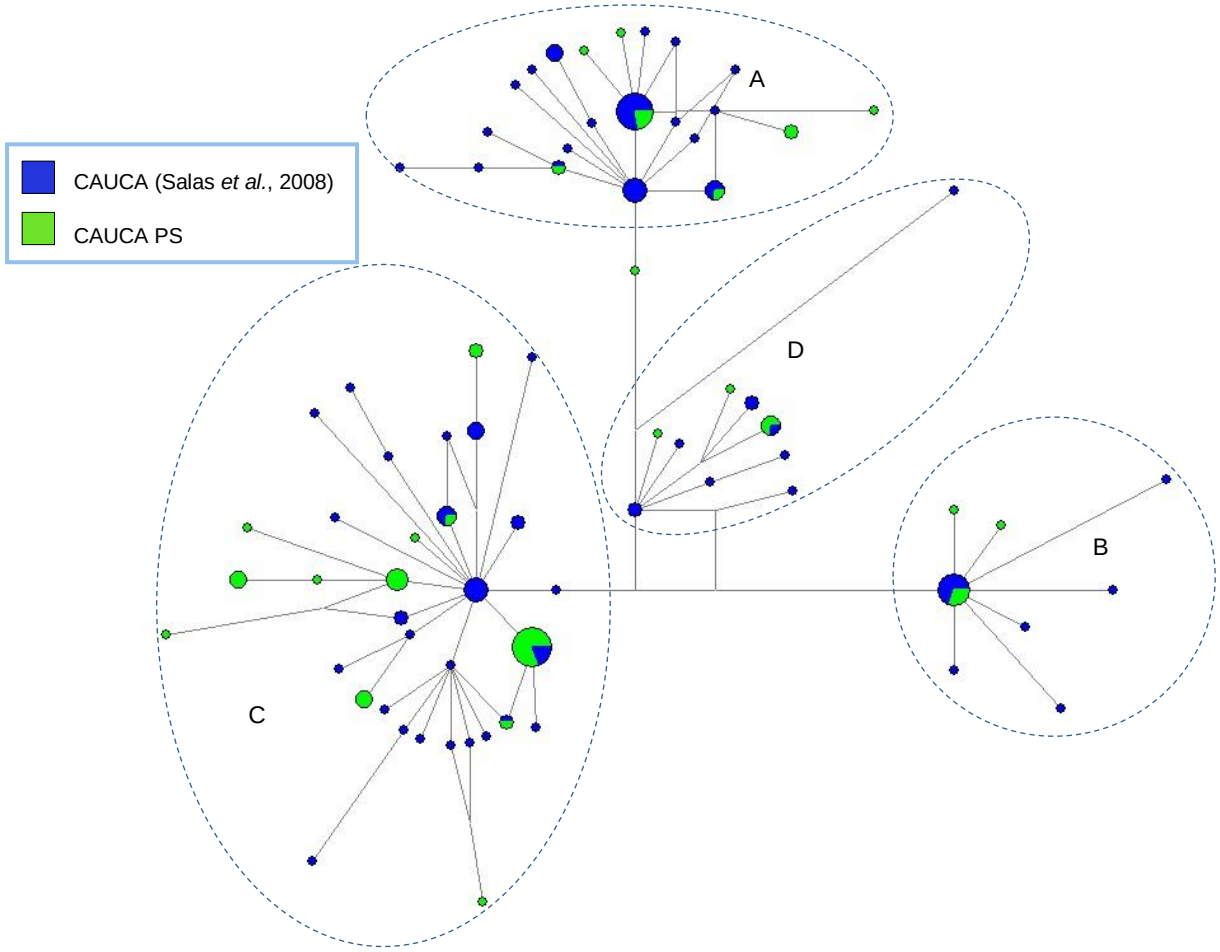


Figure 16 - Median joining network based on HVRI. Data from the Cauca region (Salas *et al.*, 2008) and present study's Cauca. Circle sizes are proportional to the haplotype frequencies.

In order to achieve a better insight into each haplogroup, network analyses of the different haplogroups were constructed with the present and literature data. Note that no network was constructed for haplogroup D because the number of samples belonging to this haplogroup is low.

Haplogroup A

Among Native American populations the haplogroup A frequency varies, showing higher values in northern artic zones and decreasing towards Central and South America. However, populations in Central and northern South America still maintain relatively high frequencies of haplogroup A (see 1.2.2.2.1. Mitochondrial DNA Evidence section). Indeed, Antioquia region from the present data reveals a frequency of 47.4% but only 4 distinct haplotypes. On the other hand Cauca region displays a haplogroup A frequency of 18.3% but a higher number of haplotypes (9). In order to gain a better understanding of haplogroup A diversity in American individuals, a network (Figure 17) based on the CR was constructed with samples from the present study and several

groups from the American continent from the work of Yang *et al.* (2010).The network obtained shows a starlike conformation with an ancestral node. The few haplotypes involving the groups under study are shared with groups from Andes, northwest South America and Mesoamerica which is explained by geographical proximity and by the contact established between the two major ancient civilizations in America – Maya (Mexico) and Inca (Peru).

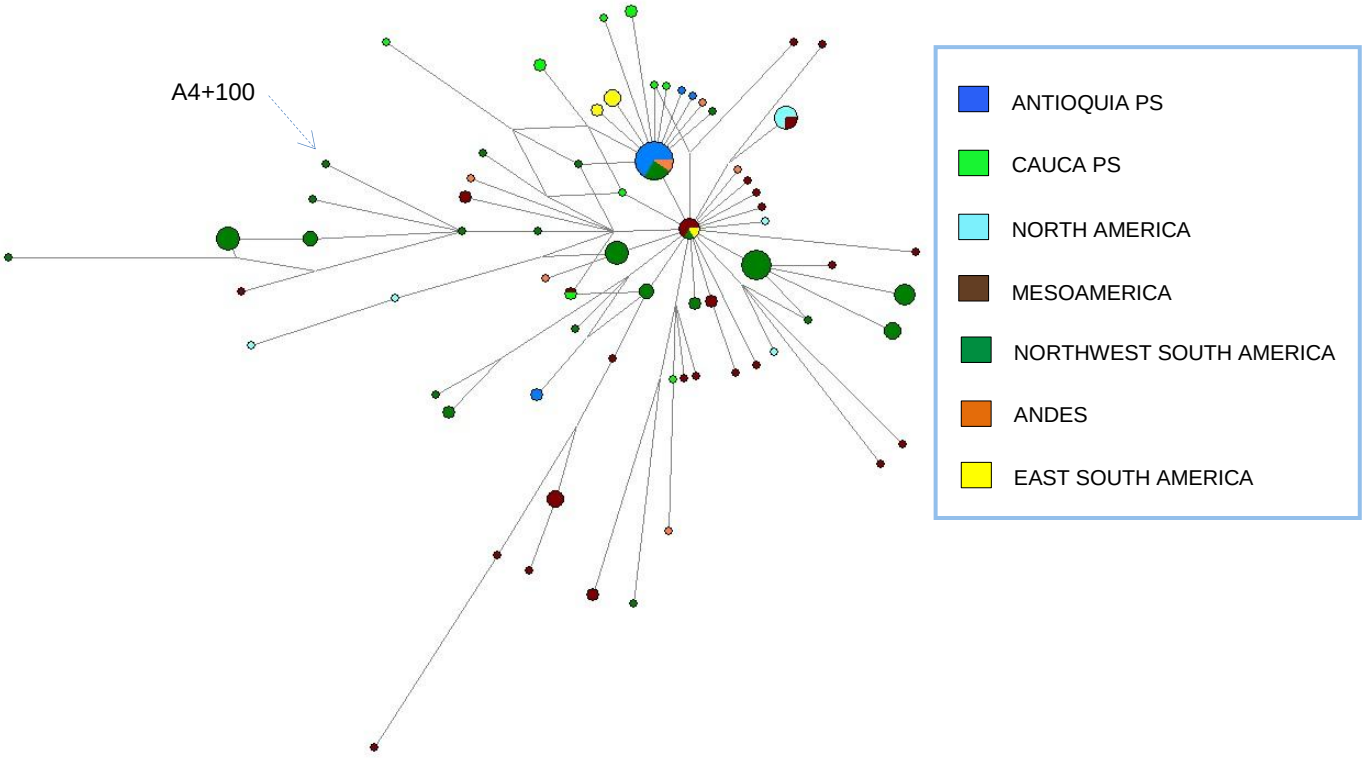


Figure 17 - Median joining network of A haplogroup, based on CR data. All literature data are gathered from Yang *et al.* (2010). All haplotypes belong to A2 branch, except the one marked with an arrow which is A4+100. Circle sizes are proportional to the haplotype frequencies.

Haplogroup B

In America, haplogroup B frequencies reach its highest values in Andean populations, followed by other South American populations (Fuselli *et al.*, 2003; Salas *et al.*, 2009; O'Rourke & Raff, 2010). In the present study, the results were as following: Antioquia revealed a frequency of 26.3% whereas the Cauca group reached 8.3%. Instead, Antioquia only presented 3 distinct haplotypes while Cauca showed 5. When Cauca is subdivided though, Chibcha and Guambiano groups reveal only 2 haplotypes each (Figures 14 and 18). A network based on CR of samples from the present study and various American regions (Yang *et al.*, 2010) was performed to improve the understanding of the haplogroup B (Figure 18). A starlike conformation was found with the central node being less frequent than some of the branches and a small number of shared haplotypes. A recent expansion is observed for the most frequent haplotype

that is shared by three different population samples. This haplotype is in the centre of a cluster of most mtDNA sequences in northwest populations from South America, including those from the present study. Most of the samples were classified as B4b due to the presence of 499A motif, a branch that originates B2 which is the most common branch in America. However B2 is characterized by motifs in the coding region which we have not typed. Other branches found in the literature data were B2c2a (16319A), B2c2 (146C), B2+16311 and B2h (16468C).

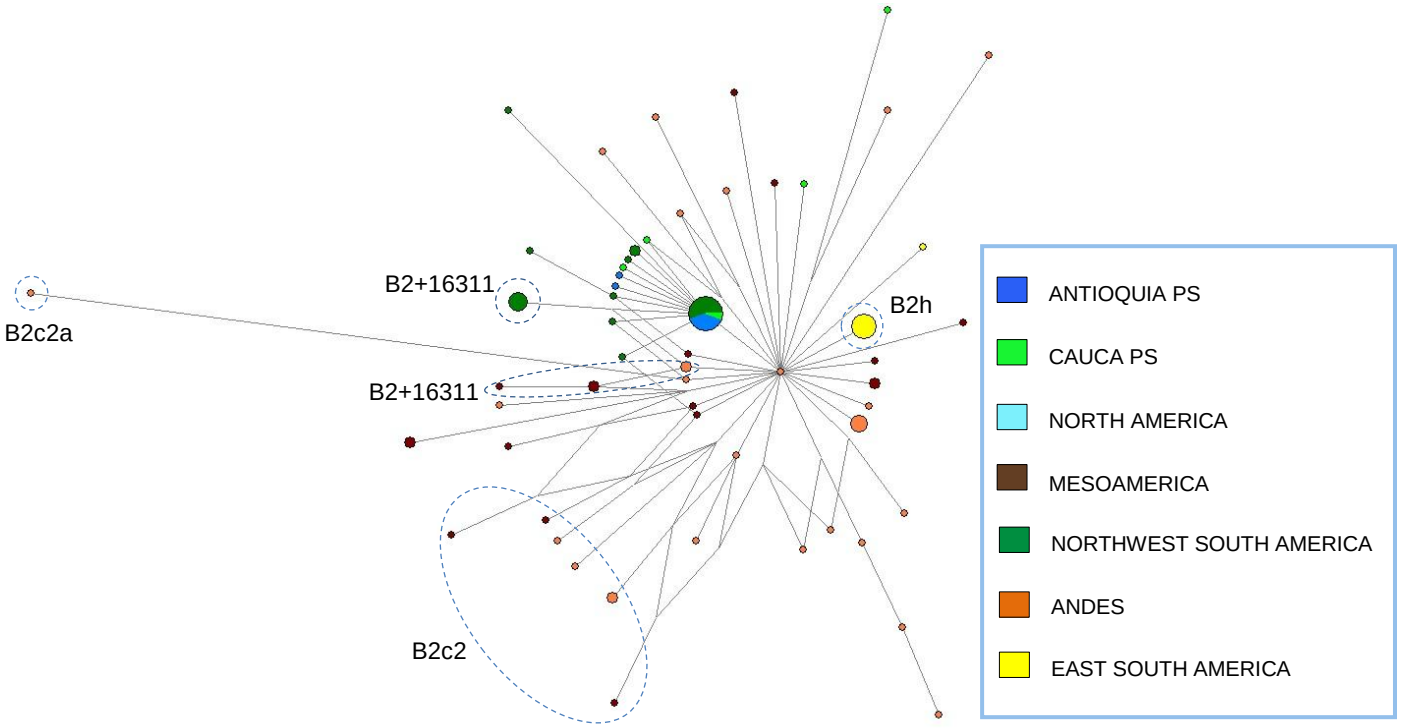


Figure 18 - Median joining network of B haplogroup, based on CR data. All literature data are gathered from (Yang *et al.*, 2010). Circle sizes are proportional to the haplotype frequencies. Samples were classified as belonging to sub-haplogroup B4b unless specified in the figure.

Haplogroup C

In this study, haplogroup C was found to be absent in Antioquia but present in a high frequency in Cauca, 63.3% in a total of 20 different haplotypes. When considering the subdivision of Cauca, Chibcha group presents 26.7% (4 haplotypes) while Guambiano presents a frequency of 75.0% (12 haplotypes) (Figure 14).

In order to provide a better insight into haplogroup C diversity, a network based on CR of present study samples and samples from several American regions was constructed (Figure 19). There are clearly two patterns of starlike configuration for the samples in the network, on the left belonging to branches C1, C1c and C1d and on the right belonging to the C1b branch. There is only one haplotype shared between Cauca and

Andes group, which is explained by the geographic proximity but also due to the fact that in the Cauca sample there are ethnic groups characteristic from the highlands.

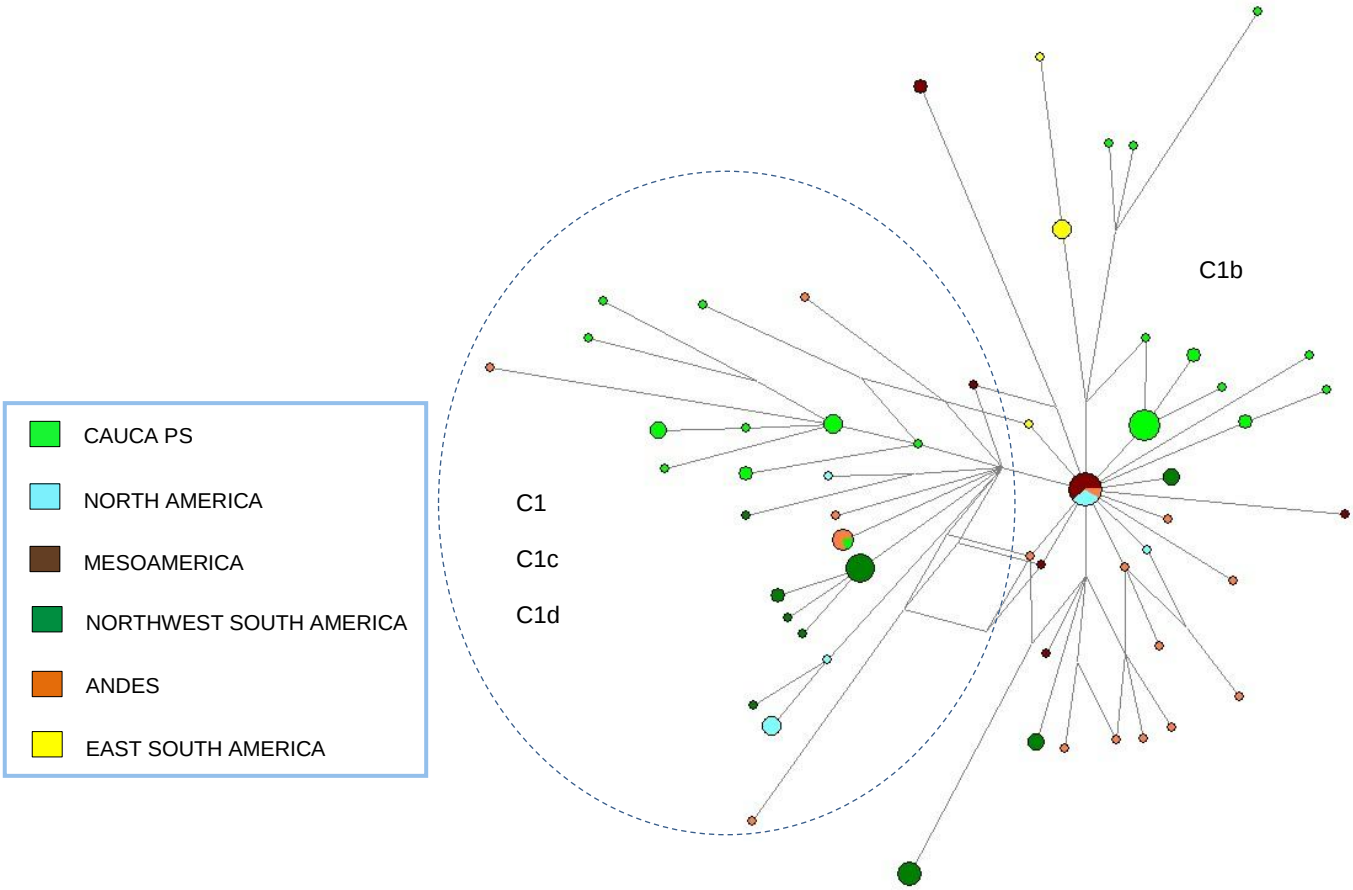


Figure 19 - Median joining network of C haplogroup, based on CR data. All literature data are gathered from (Yang *et al.*, 2010). Dashed circle separates C1b branch in the outside and the other haplotypes that belong to the C1, C1c and C1d minor haplogroups inside. Circle sizes are proportional to the haplotype frequencies.

4.4. Diversity Indices

Diversity indices were calculated using data from the present study and also for all the comparative groups. Results are displayed in Table 9.

Table 9 - Diversity indices calculated for the Present Study (PS) data and also for all the comparative groups in two levels of resolution: Complete CR and HVRI (16050-16383).

GROUP	LOCATION	N	K	CR			HVRI		
				S	H ± sd	π	S	H ± sd	π
Antioquia (PS)	**Colombia**	**38**	**11**	**26**	**0,805 ± 0,046**	**0,00818**	**14**	**0.745 ± 0,049**	**0,01642**
Cauca (PS)	**Colombia**	**60**	**40**	**74**	**0,961 ± 0,017**	**0,00944**	**39**	**0.939 ± 0.021**	**0,01849**
Chibcha speaking group (PS)	**Colombia**	**15**	**15**	**54**	**1,000 ± 0,024**	**0,01235**	**35**	**1,000 ± 0,024**	**0,01618**
Guambiano speaking group (PS)	**Colombia**	**33**	**19**	**45**	**0,900 ± 0,043**	**0,00709**	**26**	**0,848 ± 0,053**	**0,00935**
Colombia[1]	Colombia	64	31	-	-	-	40	0.904 ± 0,029	0,0155
North Colombia[2]	Colombia	110	11	-	-	-	22	0.815 ± 0,024	0,01458
Cauca[3]	Colombia	98	57	-	-	-	51	0.973 ± 0,007	0,01964
Arhuaco[4]	Colombia	16	4	17	0,442 ± 0,145	0,00332	-	-	-
Embera[4]	Colombia	9	4	18	0,750 ± 0,112	0,00493	-	-	-
Inga[4]	Colombia	16	7	24	0,850 ± 0,060	0,00869	-	-	-
Kogi[4]	Colombia	16	6	22	0,850 ± 0,047	0,00747	-	-	-
Ticuna[4]	Colombia	12	6	24	0,879 ± 0,060	0,00873	-	-	-
Waunana[4]	Colombia	18	8	23	0,830 ± 0,064	0,0076	-	-	-
Wayuu[4]	Colombia	18	8	32	0,856 ± 0,059	0,01055	-	-	-
Zenu[4]	Colombia	15	5	22	0,562 ± 0,143	0,00513	-	-	-
Ojibwa[4]	Canada	16	5	22	0,792 ± 0,064	0,00839	-	-	-
Cree[4]	Canada	11	11	11	1,000 ± 0,039	0,01054	-	-	-
Cabecar[4]	Costa Rica	15	8	22	0,829 ± 0,085	0,00884	-	-	-
Guaymi[4]	Costa Rica	13	9	21	0,949 ± 0,042	0,00704	-	-	-
Kaqchikel[4]	Guatemala	16	15	38	0,992 ± 0,025	0,00833	-	-	-
Mixtec[4]	Mexico	17	12	38	0,941 ± 0,043	0,00972	-	-	-
Mixe[4]	Mexico	20	11	41	0,895 ± 0,052	0,00969	-	-	-
Zapotec[4]	Mexico	19	15	44	0,977 ± 0,023	0,01076	-	-	-
Aymara[4]	Chile	17	16	42	0,993 ± 0,023	0,00861	-	-	-
Huilliche[4]	Chile	20	20	49	1,000 ± 0,016	0,009	-	-	-
Quechua[4]	Peru	18	18	45	1,000 ± 0,019	0,00932	-	-	-
Ache[4]	Brazil	11	2	9	0,182 ± 0,144	0,00147	-	-	-
Guarani[4]	Brazil	8	5	18	0,786 ± 0,151	0,00539	-	-	-

PS stands for present study, samples with number 1 are from (Torres et al., 2006) with 2 from (Melton et al., 2007), with 3 from (Salas et al., 2008)and groups listed with number 4 were collected from reference (Yang et al., 2010). 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 π is the nucleotide diversity. Sd is the standard deviation associated with the haplotypic diversity.

As expected, higher values of haplotypic diversity (H) were obtained when using the complete CR than those based only on HVRI. However, some works used for comparison 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.

Antioquia PS displays lower values of haplotypic diversity (H) than other Colombian regions, but when compared with other Colombian ethnic groups those values fall within the same interval. Some populations within Colombia show even lower values but due to the small sample size strong inferences cannot be taken. Considering nucleotide diversity (π) and the number of segregating sites (S), Antioquia PS values are similar to the majority observed in other Colombian regions and ethnic groups.

The diversity values observed in Cauca PS are higher than most values found in other Colombian regions and in agreement with those observed in Cauca by Salas *et al.* (2008). Comparing the Cauca subgroups with native populations from Colombia, Chibcha PS presents high values of haplotypic diversity and number of segregating sites as do some North American, Central American and Andean populations (Huilliche and Quechua). Nucleotide diversity (π) is also high in this group and only matched by one Colombian (Wayúu), one Central American (Zapotec) and one North American (Cree) populations, however the sample sizes must be taken into account (see 5. Discussion). However, the fact that Chibcha group is based on linguistic affinity, reuniting individuals from several villages, could also lead to an artificially increase of diversity.

Finally, Guambiano PS has more individuals sampled (N=33) and the majority of these come from the same villages of Sílvia and Puracé. Both diversity parameters values are in agreement with the values found in some Colombian populations. The nucleotide diversity (π) found in Guambiano PS is low and similar to the values found in some Colombian populations like Kogi and Waunana. Instead, the value of number of segregating sites found in Guambiano is higher than what is common to find in Colombian populations and resembles the values found in Central America and Andean regions.

5. DISCUSSION

The majority of haplogroups found in the present study belong to Native American lineages, which was expected since the sampling took into consideration the ethnical affinity of the individuals. The non-Native American contribution was restricted to a single mtDNA sequence from an African haplogroup, belonging to an individual from the Department of Cauca that also presents a non-Native American paternal input (J.J. Builles and L. Gusmão, personal communication). The Embéra-Chamí sample from Department of Antioquia reveals an absence of non-Native American input both in mtDNA and Y-Chromosome results (J.J. Builles and L. Gusmão, personal communication), meaning that while the region of Cauca was more subjected to non-Native American introduction, the sample from Antioquia was more preserved and isolated. When considering the Cauca sample divided into linguistic groups, the Chibcha group presents one non-Native American haplogroup while the Guambiano group presents only Native American haplogroups. These results were expected since the Guambiano people have inhabited the same regions since before the arrival of Europeans in the New World (Romoli, 1974; Llanos, 1981).

Both samples lack the presence of haplogroup X which was expected since this haplogroup is very rare and nearly absent in South American populations (Dornelles *et al.*, 2005) (see 4.1 Haplogroup Frequencies and 4.3. Phylogeographic Analysis for details).

Antioquia and Cauca regions revealed to be markedly differentiated, both in terms of haplogroup frequencies and haplotype diversities. Antioquia Embéra-Chamí population presents signs of strong founder effects, presenting five founder haplotypes (Figure 14, see 4.3. Phylogeograpic Analysis for details), and isolation, by showing a low number of different haplotypes and little haplotype sharing with other groups (Figures 14 and 15) which is expected since these individuals inhabit an Indian Reserve, however these results must be confronted with Y-chromosome data. The Cauca sample presents a higher number of haplotypes (Figure 14 and 16) and higher values of diversity indices (Table 9), however as this sample reunites individuals from various ethnic and linguistic groups, this diversity is probably artificially high.

When Cauca group is divided into two linguistic groups, the Chibcha group presents unique haplotypes (Figure 14) and high values of diversity indices (Table 9) when comparing with other Colombian populations. Nevertheless, these can result from the small sample size and the fact that this group is composed by individuals from several villages. Guambiano, on the other hand, presents lower values of the diversity indices and many shared haplotypes belonging to haplogroup C which can result from the fact

that these individuals came from in the same villages of Silvia and Puracé, where there are vestiges of Guambiano communities from before the arrival of the Europeans (Romoli, 1974; Llanos, 1981).

Considering a geographic approach, the haplogroup frequencies found in the present groups are in agreement with literature data as they go along with the two distributions found by Keyeux *et al.* (2002): the groups sampled in North Colombia show higher frequencies of haplogroup A whereas the groups sampled in South Colombia display higher frequencies of haplogroup C (Figure 10). The existence of these differences between North and South Colombia can be explained by two migration routes that occurred throughout the country: one following the Pacific Coast and the Andean chain and another one following the Amazonian plains into Brazil or backwards (Keyeux *et al.*, 2002).

Following into a broader geographic analysis, the haplogroup frequencies present in our groups are well contextualized with the haplogroup frequencies distribution throughout America (Figure 10) (Salas *et al.*, 2009). Diversity indices also agree with literature data as most of the groups reveal diversity parameters that fall within the interval of values seen in the Colombian groups (Table 9). Moreover higher values of diversity are found in North and Central American populations, and within South America, in Andean populations. Intermediate values were observed in Colombia and lower values in eastern South America. These results agree with Rothhammer & Dillehay (2009) that proposes a decrease in diversity from North to South America and another one from West to East within South America. The genetic distances point to a strong differentiation between American geographic regions that is explained by the strong effects of genetic drift and bottlenecks in small-sized populations (see sections 4.1. Haplogroup Frequencies, 4.4. Diversity Indices and 4.2. Genetic Distances for details). The genetic distance analysis shows the formation of two clusters, one composed by Mesoamerica, northern South America and Antioquia (Table 7 and Figure 12.A) which is probably due to the migrations from the Central America into northwest South America and backwards (Keyeux *et al.*, 2002; Melton *et al.*, 2007). The second cluster reunites East South America and Andes which are geographically close. The sample from Cauca remains distant from both groups (Figure 12.A, see section 4.2. Genetic Distances for details).

Following a linguistic criterion, an affinity was found between Chibchan speaking groups and Central American groups (Figure 11) that could also suggest the maintenance of contact between these two regions by the sequential migrations

performed by the Chibcha speakers, while non-Chibchan groups remain differentiated (Tables 8 and 11, Figures 12.B and 13, see 4.2. Genetic Distances for details). However, the presence of an Embéra-Chamí haplotype belonging to haplogroup A in populations from the Andean region indicates that the high frequency of this haplogroup in Embéra-Chamí population within Antioquia is a result from genetic drift and not a result of linguistic associated migrations (Figure 17, see section 4.3. Phylogeographic Analysis for details).

Finally, our results support the occurrence of two migratory waves within Colombia. One in the northern part of the country and into the Pacific coast and another migration in the southern part of the country that went through the Amazonian plains or backwards and maintained high frequencies of haplogroup C. Our results are well contextualized in a broader American perspective and in agreement with literature.

It is also worth mentioning that the pattern of differentiation found between groups and the fact that different groups display restricted haplotypes is of outmost relevance in the forensic genetic analyses in these populations.

6. CONCLUSIONS

98 mtDNA haplotypes were deposited in EMPOP©, mtDNA Forensic Database and their haplogroup classification confirmed;

i. The non-Native American maternal contribution to the populations under study was little or inexistent;

ii. The two Colombian regions studied are very differentiated with little haplotype sharing and small number of haplotypes probably as a consequence of strong drift effects and small population sizes, pattern very common in Amerindian populations:
    a. Antioquia's "Embéra-Chamí" population shows signs of isolation, however Y-chromosome data must be analysed;
    b. Cauca's subgroups reveal little haplotype sharing, indicating that a wider sampling is needed;

iii. Two different distributions of mtDNA haplogroup frequencies were found within Colombia:
    a. In North Colombia populations present higher frequencies of haplogroup A that is explained as a consequence of a migration following the Andean chain and the Pacific Coastline;
    b. In South Colombia there is a prevalence of haplogroup C which is explained by the occurrence of a migration through the Amazonian plains into Brazil;
    c. Nevertheless, studies with higher resolution and broader sampling are needed;

iv. Strong differentiation between geographic regions throughout the American continent have been found which reflect the strong genetic drift effects and the small population sizes of all Native American populations;

v. When a linguistic criterion was applied, Chibchan speakers revealed proximity, whereas non-Chibchan remained differentiated;
    a. Chibcha-related groups (like Embéra-Chamí from Antioquia) present a high frequency of haplogroup A in common with Central American populations;

b. Due to the small number of samples registered as Chibcha speaking group, broader sampling should be made;

vi. The differences found between Native Americans populations and the pattern of restricted haplotypes makes the genetic characterization of these groups of extreme relevance for forensic genetic studies.

BIBLIOGRAPHY

Achilli, A., Perego, U.A., Bravi, C.M., Coble, M.D., Kong, Q.-P., Woodward, S.R., Salas, A., Torroni, A. & Bandelt, H.-J. (2008) The Phylogeny of the Four Pan-American MtDNA Haplogroups: Implications for Evolutionary and Disease Studies. *PLoS ONE,* 3**,** e1764.

Acosta, J.D., Mangan, J.E., Mignolo, W. & López-Morillas, F.M. (2002) *Natural and Moral History of the Indies.* Duke University Press.

Adovasio, J.M. & Page, J. (2003) *The First Americans: In Pursuit of Archaeology's Greatest Mystery.* Random House Publishing Group.

Adovasio, J.M., Pedler, D., Donahue, J. & Stuckenrath, R. (1999) No Vestige of a beginning nor prospect for an end: two decades of debate on Meadowcroft Shelter. In: *Ice age peoples of North America* R.B.a.K. Turmire (ed.) *Ice age peoples of North America.* Oregon State University Press.

Amorim, A. (2007) Genetic markers: The interplay between concepts and technology in the Anthropological scene. In: *Recent Advances in Molecular Biology and Evolution: Applications to Biological Anthropology* C.S.a.M. Lima (ed.) *Recent Advances in Molecular Biology and Evolution: Applications to Biological Anthropology.* Research Signpost, Kerala.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics,* 23**,** 147.

Arango, R. & Sanchéz, E. (1998) *Los pueblos indígenas de Colombia 1997, desarrollo y territorio. Departamento Nacional de Planeación.* Unidad Administrativa Especial de Desarrollo Territorial.

Arango, R. & Sánchez, S.G. (2004) *Los pueblos indígenas de Colombia en el umbral del nuevo milenio: Población, cultura y territorio: bases para el fortalecimiento social y económico de los pueblos indígenas.* Departamento Nacional de Planeación.

Arnaiz-Villena, A., Parga-Lozano, C., Moreno, E., Areces, C., Rey, D. & Gomez-Prieto, P. (2010) The Origin of Amerindians and the Peopling of the Americas According to HLA Genes: Admixture with Asian and Pacific People. *Current Genomics,* 11**,** 481-481.

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., Forster, P., Sykes, B.C. & Richards, M.B. (1995) Mitochondrial portraits of Human-Populations using Median Networks. *Genetics,* 141**,** 743-753.

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.

Bodner, M., Perego, U.A., Huber, G., Fendt, L., Röck, A.W., Zimmermann, B., Olivieri, A., Gómez-Carballa, A., Lancioni, H., Angerhofer, N., Bobillo, M.C., Corach, D., Woodward, S.R., Salas, A., Achilli, A., Torroni, A., Bandelt, H.-J. & Parson, W. (2012) Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Research*.

Bradley, B. & Stanford, D. (2004) The North Atlantic ice-edge corridor: A possible Palaeolithic route to the New World. *World Archaeology,* 36**,** 459-478.

Brigham-Grette, J., Lozhkin, A.V., Anderson, P.M. & Gluskova, O.Y. (2004) Paleoenvironmental conditions in western Beringia before and during the last glacial maximum. In: *In Entering America: Northeast Asia and Beringia Before the Last Glacial Maximum* D.B. Madsen (ed.) *In Entering America: Northeast Asia and Beringia Before the Last Glacial Maximum.* Salt Lake City: University of Utah Press.

Brown, M.D., Hosseini, S.H., Torroni, A., Bandelt, H.J., Allen, J.C., Schurr, T.G., Scozzari, R., Cruciani, F. & Wallace, D.C. (1998) mtDNA haplogroup X: An ancient link between Europe/Western Asia and North America? *Am J Hum Genet,* 63**,** 1852-61.

Bryan, A.L. (1986) *New evidence for the Pleistocene peopling of the Americas.* Center for the Study of Early Man, University of Maine.

Budowle, B., Allard, M., Wilson, M. & Chakraborty, R. (2003) Forensics and Mitochondrial DNA: Applications, Debates, and Foundations. *Annual Review of Genomics and Human Genetics,* 4**,** 119-141.

Burger, G., Gray, M.W. & Franz Lang, B. (2003) Mitochondrial genomes: anything goes. *Trends in Genetics,* 19**,** 709-716.

Butler, J.M. (2005) *Forensic DNA Typing: Biology, Technology, and Genetics of Str Markers.* Elsevier Academic Press.

Callegari-Jacques, S.M., Salzano, F.M., Weimer, T.A., Hutz, M.H., Black, F.L., Santos, S.E., Guerreiro, J.F., Mestriner, M.A. & Pandey, J.P. (1994) Further blood genetic studies on Amazonian diversity--data from four Indian groups. *Annals of human biology,* 21**,** 465-81.

Campbell, L. (2000) *American Indian Languages: The Historical Linguistics of Native America.* Oxford University Press.

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.

Carvajal-Carmona, L.G., Soto, I.D., Pineda, N., Ortíz-Barrientos, D., Duque, C., Ospina-Duque, J., Mccarthy, M., Montoya, P., Alvarez, V.M., Bedoya, G. & Ruiz-Linares, A. (2000) Strong Amerind/White Sex Bias and a Possible Sephardic Contribution among the Founders of a Population in Northwest Colombia. *American journal of human genetics,* 67**,** 1287-1295.

Casas-Vargas, A., Gómez, A., Briceño, I., Díaz-Matallana, M., Bernal, J.E. & Rodríguez, J.V. (2011) High genetic diversity on a sample of pre-Columbian bone remains from Guane territories in northwestern Colombia. *American Journal of Physical Anthropology,* 146**,** 637-649.

Cavalli-Sforza & Feldman, M.W. (2003) The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.,* 33 Suppl**,** 266-275.

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

Clague, J.J., Matthews, R.W. & Ager, T.A. (2004) *Entering America: Northeast Asia and Beringia Before the Last Glacial Maximum.* University of Utah Press.

Clapperton, C.M. (1993) *Quaternary geology and geomorphology of South America.* Elsevier.

Constenla, U.A. & Margery, P.E. (1991) Elementos de fonología comparada Chocó. *Filología y lingüística,* 17**,** 137-191.

Curieux, T.R., Unicef & Press, F.A. (2009) Atlas sociolingüístico de pueblos indígenas en América Latina.). FUNPROEIB Andes Press.
Dane, D.a.N.E. (2007) *Colombia una nación multicultural. Su diversidad étnica.*
Dillehay, T.D. (1997) *Monte Verde, a Late Pleistocene Settlement in Chile: The archaeological context and interpretation.* Smithsonian Institution Press.
Dillehay, T.D. (1999) The late Pleistocene cultures of South America. *Evolutionary Anthropology: Issues, News, and Reviews,* 7**,** 206-216.
Dillehay, T.D. (2000) *The Settlement of the Americas: A New Prehistory.* New York: Basic Books.
Dillehay, T.D. (2009) Probing deeper into first American studies. *Proc Natl Acad Sci U S A*.
Dixon, J.E. (2001) Human colonization of the Americas: timing, technology and process. *Quaternary Science Reviews,* 20**,** 277-299.
Dixon, J.E. (2006) Peopling of the Americas. How and when did people first come to North America? . *Athena Review,* 3.
Dornelles, C.L., Bonatto, S.L., De Freitas, L.B. & Salzano, F.M. (2005) Is haplogroup X present in extant South American Indians? *American Journal of Physical Anthropology,* 127**,** 439-448.
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.).
Ebenesersdóttir, S.S., Sigurðsson, Á., Sánchez-Quinto, F., Lalueza-Fox, C., Stefánsson, K. & Helgason, A. (2011) A new subclade of mtDNA haplogroup C1 found in icelanders: Evidence of pre-columbian contact? *American Journal of Physical Anthropology,* 144**,** 92-99.
Elson, J.L., Andrews, R.M., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. & Howell, N. (2001) Analysis of European mtDNAs for Recombination. *American journal of human genetics,* 68**,** 145-153.
Estrada-Mena, B., Estrada, F.J., Ulloa-Arvizu, R., Guido, M., Méndez, R., Coral, R., Canto, T., Granados, J., Rubí-Castellanos, R., Rangel-Villalobos, H. & García-Carrancá, A. (2010) Blood group O alleles in Native Americans: Implications in the peopling of the Americas. *American Journal of Physical Anthropology,* 142**,** 85-94.
Estrada, E., Meggers, B.J. & Evans, C. (1962) Possible Transpacific Contact on the Coast of Ecuador. *Science,* 135**,** 371-372.
Excoffier, L. & 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**,** 564-567.
Fernández, C. (1999) La arqueología molecular aplicada a la solución de problemas prehistoricos: análisis de ADN mitocondrial en momias y restos óseos prehispánicos (tésis).). Bogotá: Universidad Nacional de Colombia.
Fisher, R.A. & Bennett, J.H. (1930) *The Genetical Theory of Natural Selection: A Complete Variorum Edition.* Oxford University Press.
Fix, A.G. (2005) Rapid deployment of the five founding Amerind mtDNA haplogroups via coastal and riverine colonization. *American Journal of Physical Anthropology,* 128**,** 430-436.
Forster, P., Harding, R., Torroni, A. & Bandelt, H.J. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *American journal of human genetics,* 59**,** 935-45.
Fox, C.L. (1996) Mitochondrial DNA haplogroups in four tribes from Tierra del Fuego-Patagonia: inferences about the peopling of the Americas. *Human Biology,* 68**,** 855–871.
Friedemann, N.S.D. (1993) *Presencia Africana en Colombia. La saga del Negro.*: Instituto de Genetica Humana. Facultad de Medicina. Pontificia Universidad Javeriana.
Fuselli, S., Tarazona-Santos, E., Dupanloup, I., Soto, A., Luiselli, D. & Pettener, D. (2003) Mitochondrial DNA Diversity in South America and the Genetic History of Andean Highlanders. *Molecular Biology and Evolution,* 20**,** 1682-1691.
Goebel, T., Waters, M.R. & O'rourke, D.H. (2008) The Late Pleistocene Dispersal of Modern Humans in the Americas. *Science,* 319**,** 1497-1502.
Greenberg, J.H. & Ruhlen, M. (2007) *An Amerind etymological dictionary.* Stanford: Stanford University Press.
Greenberg, J.H., Turner, C.G. & Zegura, S.L. (1986) The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. *Current Anthropology,* 27**,** 477-497.
Griffiths, A., Wessler, S., Lewontin, R. & Carroll, S. (2008) *Introduction to Genetic Analysis (Introduction to Genetic Analysis (Griffiths)).* New York: W. H. Freeman.
Haldane, J.B.S. (1932) *The causes of evolution, by J.B.S. Haldane.* London, New York [etc.]: Longmans, Green and co.
Haynes, G. (2002) *The Early Settlement of North America: The Clovis Era.* Cambridge University Press.
Herrnstadt, C., Elson, J.L., Fahy, E., Preston, G., Turnbull, D.M., Anderson, C., Ghosh, S.S., Olefsky, J.M., Beal, M.F., Davis, R.E. & Howell, N. (2002) Reduced-Median-Network Analysis of Complete Mitochondrial DNA Coding-Region Sequences for the Major African, Asian, and European Haplogroups.). DigitalCommons@University of Nebraska - Lincoln.
Hirszfeld, L. & Hirszfeld, H. (1919) Essai d'ápplication des methods au probléme des races. *Anthropologie,* 29**,** 505-537.
Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature,* 408**,** 708.
Jackson, L.E., Phillips, F.M., Shimamura, K. & Little, E.C. (1997) Cosmogenic 36Cl dating of the Foothills erratics train, Alberta, Canada. *Geology,* 25**,** 195-198.
Keefer, D.K., Defrance, S.D., Moseley, M.E., Richardson, J.B., Satterlee, D.R. & Day-Lewis, A. (1998) Early Maritime Economy and El Niño Events at Quebrada Tacahuay, Peru. *Science,* 281**,** 1833-1835.
Keyeux, G., Rodas, C., Gelvez, N. & Carter, D. (2002) Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations. *Human Biology,* 74**,** 211-233.
Kitchen, A., Miyamoto, M.M. & Mulligan, C.J. (2008) A Three-Stage Colonization Model for the Peopling of the Americas. *PLoS ONE,* 3**,** e1596.
Klein, H. (1999) *The Atlantic Slave Trade.* Press Syndicate of the University of Cambridge.
Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G. & Kronenberg, F. (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Human Mutation,* 32**,** 25-32.
Lahr, M.M. (1995) Patterns of modern human diversification: Implications for Amerindian origins. *American Journal of Physical Anthropology,* 38**,** 163-198.

Lalueza, C., Pérez-Pérez, A., Prats, E., Cornudella, L. & Turbón, D. (1997) Lack of Founding Amerindian Mitochondrial DNA Lineages in Extinct Aborigines from Tierra del Fuego-Patagonia. *Human Molecular Genetics,* 6**,** 41-46.

Lavallee, D. (2000) *The First South Americans.* Salt Lake City: University of Utah Press.

Librado, P. & Rozas, J. (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics,* 25**,** 1451-1452.

Llanos, H. (1981) *Los Cacicazgos de Popayán a la llegada de los Conquistadores.* Bogotá: FIAN.

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N.K., Raja, J.M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.-J., Oppenheimer, S., Torroni, A. & Richards, M. (2005) Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes. *Science,* 308**,** 1034-1036.

Mandryk, C.a.S., Josenhans, H., Fedje, D.W. & Mathewes, R.W. (2001) Late Quaternary paleoenvironments of Northwestern North America: implications for inland versus coastal migration routes. *Quaternary Science Reviews,* 20**,** 301-314.

Manfredi, G., Thyagarajan, D., Papadopoulou, L.C., Pallotti, F. & Schon, E.A. (1997) The Fate of Human Sperm-Derived mtDNA in Somatic Cells. *American journal of human genetics,* 61**,** 953-960.

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

Mazières, S. (2011) Towards a reconciling model about the initial peopling of America. *Comptes Rendus Biologies,* 334**,** 497-504.

Mcavoy, J.M., Mcavoy, L.D., Resources, V.D.O.H. & Research, N.R.S.A. (1997) *Archaeological investigations of site 44SX202, Cactus Hill, Sussex County, Virginia.* Department of Historic Resources.

Melton, P.E., Briceño, I., Gómez, A., Devor, E.J., Bernal, J.E. & Crawford, M.H. (2007) Biological relationship between central and South American Chibchan speaking populations: Evidence from mtDNA. *American Journal of Physical Anthropology,* 133**,** 753-770.

Meltzer, D.J. (2004) Peopling of North America. In: *The quaternary period in the United States* S.C.P.a.B.a.E. A. Gillespie (ed.) *The quaternary period in the United States.* New York: Elsevier Science.

Meltzer, D.J. (2006) Paleoamerican origins: beyond Clovis. *Journal of Field Archaeology,* 31**,** 441-443.

Merriwether, D.A., Rothhammer, F. & Ferrell, R.E. (1995) Distribution of the four founding lineage haplotypes in native Americans suggests a single wave of migration for the New World. *American Journal of Physical Anthropology,* 98**,** 411-430.

Mesa, N.R., Mondragón, M.C., Soto, I.D., Parra, M.V., Duque, C., Ortíz-Barrientos, D., García, L.F., Velez, I.D., Bravo, M.L., Múnera, J.G., Bedoya, G., Bortolini, M.-C. & Ruiz-Linares, A. (2000) Autosomal, mtDNA, and Y-Chromosome Diversity in Amerinds: Pre- and Post-Columbian Patterns of Gene Flow in South America. *American Journal of Human Genetics,* 67**,** 1277-1286.

Millstein, R.L. & Skipper, R.A. (2006) Population Genetics.).

Monsalve, M.V., Cardenas, F., Guhl, F., Delaney, A.D. & Devine, D.V. (1996) Phylogenetic analysis of mtDNA lineages in South American mummies. *Annals of Human Genetics,* 60**,** 293-303.

Moraga, M.L., Rocco, P., Miquel, J.F., Nervi, F., Llop, E., Chakraborty, R., Rothhammer, F. & Carvallo, P. (2000) Mitochondrial DNA polymorphisms in Chilean aboriginal populations: Implications for the peopling of the southern cone of the continent. *American Journal of Physical Anthropology,* 113**,** 19-29.

Morell, V. (1990) Confusion in Earliest America. *Science,* 248**,** 439-441.

Mourant, A.E. (1985) *Blood Relations: Blood Groups and Anthropology.* Oxford University Press.

Nei, M. (1987) *Molecular Evolutionary Genetics.* Columbia University Press.

Neves, W.A. & Hubbe, M. (2005) Cranial morphology of early Americans from Lagoa Santa, Brazil: Implications for the settlement of the New World. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 18309-18314.

Neves, W.A., Prous, A., González-José, R., Kipnis, R. & Powell, J. (2003) Early Holocene human skeletal remains from Santana do Riacho, Brazil: implications for the settlement of the New World. *Journal of Human Evolution,* 45**,** 19-42.

Neves, W.A. & Pucciarelli, H.M. (1990) The origin of the first Americans: an analysis based on the cranial morphology of early South American human remains. *American Journal of Physical Anthropology* 81.

Neves, W.A. & Pucciarelli, H.M. (1991) Morphological affinities of the first Americans: an exploratory analysis based on early South American human remains. *Journal of Human Evolution,* 21**,** 261-273.

O'rourke, D.H. & Raff, J.A. (2010) The Human Genetic History of the Americas: The Final Frontier. *Current biology : CB,* 20**,** R202-R207.

Parson, W., Brandstätter, A., Pircher, M., Steinlechner, M. & Scheithauer, R. (2004) EMPOP—the EDNAP mtDNA population database concept for a new generation, high-quality mtDNA database. *International Congress Series,* 1261**,** 106-108.

Pauling, L., Itano, H.A., Singer, S.J. & Wells, I.C. (1949) Sickle Cell Anemia, a Molecular Disease. *Science,* 110**,** 543-548.

Perego, U.A., Achilli, A., Angerhofer, N., Accetturo, M., Pala, M., Olivieri, A., Kashani, B.H., Ritchie, K.H., Scozzari, R., Kong, Q.-P., Myres, N.M., Salas, A., Semino, O., Bandelt, H.-J., Woodward, S.R. & Torroni, A. (2009) Distinctive Paleo-Indian Migration Routes from Beringia Marked by Two Rare mtDNA Haplogroups. *Current biology : CB,* 19**,** 1-8.

Pietschmann, R. & De Gamboa, P.S. (1906) *Geschichte des Inkareiches.* Kraus Reprint.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., Garcia, L.F., Triana, O., Blair, S., Maestre, A., Dib, J.C., Bravi, C.M., Bailliet, G., Corach, D., Hunemeier, T., Bortolini, M.C., Salzano, F.M., Petzl-Erler, M.L., Acuna-Alonzo, V., Aguilar-Salinas, C., Canizales-Quinteros, S., Tusie-Luna, T., Riba, L., Rodriguez-Cruz, M., Lopez-Alarcon, M., Coral-Vazquez, R., Canto-Cetina, T., Silva-Zolezzi, I., Fernandez-Lopez, J.C., Contreras, A.V., Jimenez-Sanchez, G., Gomez-Vazquez, M.J., Molina, J., Carracedo, A., Salas, A., Gallo, C., Poletti, G., Witonsky, D.B., Alkorta-Aranburu, G., Sukernik, R.I., Osipova, L., Fedorova, S.A., Vasquez, R., Villena, M., Moreau, C., Barrantes, R., Pauls, D., Excoffier, L., Bedoya, G., Rothhammer, F., Dugoujon, J.-M., Larrouy, G., Klitz, W., Labuda, D., Kidd, J., Kidd, K., Di Rienzo, A., Freimer, N.B., Price, A.L. & Ruiz-Linares, A. (2012) Reconstructing Native American population history. *Nature,* advance online publication.

Rivet, P. (1943) Recherches anthropologiques sur la Basse-Californie. *J Soc Amer Paris* 3**,** 6-109.

Roewer, L., Nothnagel, M., Gusmão, L., Gomes, V., González, M., Corach, D., Sala, A., Alechine, E., Palha, T., Santos, N., Ribeiro-Dos-Santos, A., Geppert, M., Willuweit, S., Nagy, M., Zweynert, S., Baeta, M., Núñez, C., Martínez-Jarreta, B., González-Andrade, F., Carvalho, E.F.D., Silva, D.a.D., Builes, J.J., Borrega, D.T., Parra, A.M.L., Arroyo-Pardo, E., Toscanini, U., Borjas, L., Barletta, C., Ewart, E., Santos, S. & Krawczak, M. (2012) Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. *American Journal of Human Genetics, in press*.

Romoli, K. (1974) Nomenclatura y población indígena de la Antigua jurisdicción de Cali a mediados del Siglo XVI. *Revista Colombiana de Antropología,* 16.

Roostalu, U., Kutuev, I., Loogväli, E.-L., Metspalu, E., Tambets, K., Reidla, M., Khusnutdinova, E., Usanga, E., Kivisild, T. & Villems, R. (2007) Origin and Expansion of Haplogroup H, the Dominant Human Mitochondrial DNA Lineage in West Eurasia: The Near Eastern and Caucasian Perspective. *Molecular Biology and Evolution,* 24**,** 436-448.

Rothhammer, F. & Dillehay, T.D. (2009) The Late Pleistocene Colonization of South America: An Interdisciplinary Perspective. *Annals of Human Genetics,* 73**,** 540-549.

Rothhammer, F. & Silva, C. (1992) Gene geography of South America: Testing models of population displacement based on archeological evidence. *American Journal of Physical Anthropology,* 89**,** 441-446.

Rothhammer, F., Silva, C., Callegari-Jacques, S.M., Llop, E. & Salzano, F.M. (1997) Gradients of HLA diversity in South American Indians. *Annals of human biology,* 24**,** 197-208.

Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H. & Arnheim, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science,* 230**,** 1350-1354.

Salas, A., Acosta, A., Álvarez-Iglesias, V., Cerezo, M., Phillips, C., Lareu, M.V. & Carracedo, Á. (2008) The mtDNA ancestry of admixed Colombian populations. *American Journal of Human Biology,* 20**,** 584-591.

Salas, A., Bandelt, H.J., Macaulay, V. & Richards, M.B. (2007) Phylogeographic investigations: The role of trees in forensic genetics. *Forensic Science International,* 168**,** 1-13.

Salas, A., Lovo-Gómez, J., Álvarez-Iglesias, V., Cerezo, M., Lareu, M.V., Macaulay, V., Richards, M.B. & Carracedo, Á. (2009) Mitochondrial Echoes of First Settlement and Genetic Continuity in El Salvador. *PLoS ONE,* 4**,** e6882.

Salas, A., Richards, M., De La Fe, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V. & Carracedo, Á. (2002) The Making of the African mtDNA Landscape. *The American Journal of Human Genetics,* 71**,** 1082-1111.

Salas, A., Richards, M., Lareu, M.-V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V. & Carracedo, Á. (2004) The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade. *The American Journal of Human Genetics,* 74**,** 454-465.

Sánchez, C. (2007) Secuenciación de ADN mitocondrial a partir de fragmentos óseos prehispánicos hallados en el sector de Candelaria La Nueva en Bogotá (Tesis).). Bogotá: Pontificia Universidad Javeriana.

Sandweiss, D.H., Mcinnis, H., Burger, R.L., Cano, A., Ojeda, B., Paredes, R., Sandweiss, M.a.D.C. & Glascock, M.D. (1998) Quebrada Jaguay: Early South American Maritime Adaptations. *Science,* 281**,** 1830-1832.

Schurr, T.G. (2004) The peopling of the New World: Perspectives from Molecular Anthropology. *Annual Review of Anthropology,* 33**,** 551-583.

Schurr, T.G. & Sherry, S.T. (2004) Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: Evolutionary and demographic evidence. *American Journal of Human Biology,* 16**,** 420-439.

Schwartz, M. & Vissing, J. (2002) Paternal Inheritance of Mitochondrial DNA. *New England Journal of Medicine,* 347**,** 576-580.

Sichra, I., Unicef & Andes, F. (2009) Atlas sociolingüístico de pueblos indígenas en América Latina.). Cochabamba, Bolivia: FUNPROEIB Andes.

Silva, A., Briceño, I., Burgos, J., Torres, D., Villegas, V., Gómez, A., Bernal, J.E. & Rodríguez, J.V. (2008) Análisis de ADN mitocondrial en una muestra de restos óseos arcaicos del periodo Herrera en la sabana de Bogotá. *Biomédica,* 28**,** 569-577.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., Salas, A., Oppenheimer, S., Macaulay, V. & Richards, M.B. (2009) Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *American journal of human genetics,* 84**,** 740-759.

Spss, I. (2001) SPSS for Windows.). 11.0.1 ed. Chicago: IBM

Stanford, D.a.B., B. (2002) Ocean trails and prairie paths? Thoughts about Clovis Origins. In: *The First Americans: The Pleistocene colonization of the New World.* N. Jablonski (ed.) *The First Americans: The Pleistocene colonization of the New World.*: Memoirs of the California Academy of Sciences

Stothert, K. (1998) An early holocene maritime adaptation in southwest Ecuador: New perspectives on the Las Vegas evidence. In: *63rd Ann. Meeting of the SAA*) *63rd Ann. Meeting of the SAA.* Seattle.

Szathmary & E., E.J. (1993) Genetics of aboriginal north Americans. *Evolutionary Anthropology: Issues, News, and Reviews,* 1**,** 202-220.

Szathmary, E.J. (1993) mtDNA and the peopling of the Americas. *American journal of human genetics,* 53**,** 793-9.

Tajima, F. (1983) Evolutionary relationship of DNA sequencies in finite populations. *Genetics,* 105**,** 437-460.

Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D.G., Mulligan, C.J., Bravi, C.M., Rickards, O., Martinez-Labarga, C., Khusnutdinova, E.K., Fedorova, S.A., Golubenko, M.V., Stepanov, V.A., Gubina, M.A., Zhadanov, S.I., Ossipova, L.P., Damba, L., Voevoda, M.I., Dipierri, J.E., Villems, R. & Malhi, R.S. (2007) Beringian Standstill and Spread of Native American Founders. *PLoS ONE,* 2**,** e829.

Tarazona-Santos, E., Carvalho-Silva, D.R., Pettener, D., Luiselli, D., De Stefano, G.F., Labarga, C.M., Rickards, O., Tyler-Smith, C., Pena, S.D.J. & Santos, F.R. (2001) Genetic Differentiation in South Amerindians Is Related to Environmental and Cultural Diversity: Evidence from the Y Chromosome. *American journal of human genetics,* 68**,** 1485-1496.

Torres, M.M., Bravi, C.M., Bortolini, M.-C., Duque, C., Callegari-Jacques, S., Ortiz, D., Bedoya, G., Groot De Restrepo, H. & Ruiz-Linares, A. (2006) A revertant of the major founder Native American haplogroup C common in populations from northern South America. *American Journal of Human Biology,* 18**,** 59-65.

Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J.V., Larsen, M., Smith, D.G., Vullo, C.M. & Wallace, D.C. (1993) Asian affinities and continental radiation of the 4 founding Native-American mtDNAs. *American Journal of Human Genetics,* 53**,** 563-590.

Turner, C.G. (1987) Late Pleistocene and Holocene population history of east Asia based on dental variation. *American Journal of Physical Anthropology,* 73**,** 305-321.

Valboa, M.C. (1951) *Miscelánea antártica: una historia del Perú antiguo.* Universidad Nacional Mayor de San Marcos, Facultad de Letras, Instituto de Etnología.

Van Oven, M. & Kayser, M. (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation,* 30**,** E386-E394.

Wallace, D.C. & Torroni, A. (1992) American Indian prehistory as written in mitochondrial DNA: a review. *Human Biology,* 64**,** 403-416.

Wang, S., Lewis, C.M., Jr., Jakobsson, M., Ramachandran, S., Ray, N., Bedoya, G., Rojas, W., Parra, M.V., Molina, J.A., Gallo, C., Mazzotti, G., Poletti, G., Hill, K., Hurtado, A.M., Labuda, D., Klitz, W., Barrantes, R., Bortolini, M.C., Salzano, F.M., Petzl-Erler, M.L., Tsuneto, L.T., Llop, E., Rothhammer, F., Excoffier, L., Feldman, M.W., Rosenberg, N.A. & Ruiz-Linares, A. (2007) Genetic Variation and Population Structure in Native Americans. *PLoS Genet,* 3**,** e185.

Waters, M.R. & Stafford, T.W. (2007) Redefining the Age of Clovis: Implications for the Peopling of the Americas. *Science,* 315**,** 1122-1126.

Wiesner, G. (1999) Early Ecuador people were maritime adapted. *Mammoth Trumpet* 14**,** 4-11.

Wiuf, C. (2001) Recombination in Human Mitochondrial DNA? *Genetics,* 159**,** 749-756.

Wright, S. (1931) Evolution in Mendelian Populations. *Genetics,* 16**,** 97-159.

Yang, N.N., Mazières, S., Bravi, C., Ray, N., Wang, S., Burley, M.-W., Bedoya, G., Rojas, W., Parra, M.V., Molina, J.A., Gallo, C., Poletti, G., Hill, K., Hurtado, A.M., Petzl-Erler, M.L., Tsuneto, L.T., Klitz, W., Barrantes, R., Llop, E., Rothhammer, F., Labuda, D., Salzano, F.M., Bortolini, M.-C., Excoffier, L., Dugoujon, J.M. & Ruiz-Linares, A. (2010) Contrasting Patterns of Nuclear and mtDNA Diversity in Native American Populations. *Annals of Human Genetics,* 74**,** 525-538.

ANNEXES

Table 10 - Haplotypes and Haplogroup classification of all samples from both Colombian regions studied (Antioquia and Cauca).

SAMPLE	CR MUTATED POSITIONS	HAPLOGROUP
Antioquia		
L01	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L08	16183C 16189C 16193.1C 16223T 16239T 16286T 16325C 16362C 73G 263G 309.2C 315.1C 489C	D1
L09	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.2C 315.1C 523DEL 524DEL	A2+64
L11	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L18	16183C 16189C 16193.1C 16217C 16519C 73G 263G 315.1C 498DEL 499A	B4b
L21	16182C 16183C 16189C 16223T 16239T 16286T 16325C 16362C 73G 263G 315.1C 489C	D1
L24	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L25	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L30	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L31	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L34	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L35	16183C 16189C 16217C 16519C 73G 263G 315.1C 498DEL 499A	B4b
L36	16183C 16189C 16217C 16519C 73G 263G 315.1C 498DEL 499A	B4b
L42	16126C 16223T 16290T 16319A 16362C 16399G 16519C 64T 73G 146C 153G 235G 263G 315.1C 523DEL 524DEL	A2+64+16111!
L43	16183C 16189C 16217C 16519C 73G 263G 315.1C 498DEL 499A	B4b
L48	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L49	16183C 16189C 16193.1C 16217C 16519C 73G 263G 315.1C 498DEL 499A	B4b
L90	16182C 16183C 16189C 16223T 16239T 16286T 16325C 16362C 73G 263G 309.2C 315.1C 339C 489C	D1
L51	16183C 16189C 16193.1C 16223T 16325C 16362C 73G 263G 309.1C 315.1C 489C	D1
L52	16183C 16189C 16193.1C 16223T 16325C 16362C 73G 263G 309.1C 315.1C 489C	D1
L53	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 152Y 153G 235G 263G 309.2C 315.1C 523DEL 524DEL	A2+64
L55	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L57	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L60	16183C 16189C 16193.1C 16217C 16519C 73G 263G 315.1C 498DEL 499A	B4b
L63	16126C 16223T 16290T 16319A 16362C 16399G 16519C 64T 73G 146C 153G 235G 263G 315.1C 523DEL 524DEL	A2+64+16111!
L64	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L65	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
L70	16183C 16189C 16223T 16325C 16362C 73G 263G 309.1C 315.1C 489C	D1
L74	16182C 16183C 16189C 16193.1C 16223T 16239T 16286T 16325C 16362C 73G 263G 309.2C 315.1C 489C	D1
L76	16182C 16183C 16189C 16223T 16239T 16286T 16325C 16362C 73G 263G 315.1C 489C	D1
L77	16183C 16189C 16217C 16519C 73G 263G 315.1C 498DEL 499A	B4b
L79	16183C 16189C 16223T 16325C 16362C 73G 263G 315.1C 489C	D1
L81	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.2C 315.1C 523DEL 524DEL	A2+64
L82	16183C 16189C 16193.1C 16223T 16239T 16286T 16325C 16362C 73G 263G 489C	D1
L85	16111T 16213A 16223T 16290T 16319A 16362C 16519C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL 556DEL	A2+64
L88	16183C 16189C 16217C 16519C 73G 263G 315.1C 498DEL 499A	B4b

L89	16183C 16189C 16217C 16519C 73G 152Y 263G 315.1C 498DEL 499A	B4b
L91	16183C 16189C 16217C 16519C 73G 263G 315.1C 498DEL 499A	B4b
Cauca		
P1531	16051G 16094C 16223T 16298C 16325C 16327T 16526A 73G 194T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d+194
P6208	16108T 16111T 16213A 16223T 16290T 16319A 16356C 16362C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
P2941	16111T 16213A 16223T 16290T 16319A 16362C 16391A 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2g
H2941	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
P3634	16142T 16188T 16223T 16325C 16362C 16519C 73G 152C 263G 309.1C 315.1C 489C	D1f
H3634	16086C 16223T 16278T 16294T 16309G 16390A 73G 143A 146C 152C 195C 198T 263G 315.1C	L2a1c1
P3791	16142T 16147T 16223T 16325C 16362C 16519C 73G 195C 263G 309.1C 315.1C 489C	D1f
P3654	16051G 16223T 16298C 16325C 16327T 16519C 73G 194T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d+194
H3654	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
P3666	16086C 16223T 16295T 16298C 16325C 16327T 16519C 73G 150T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
H3666	16111T 16213A 16223T 16290T 16294T 16319A 16362C 16519C 64T 73G 146C 153G 155C 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
P3699	16183C 16189C 16193.1C 16217C 16519C 73G 207A 263G 309.1C 315.1C 498DEL 499A	B4b
H3699	16051G 16129A 16223T 16234T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d
P3728	16051G 16172C 16189C 16223T 16298C 16325C 16327T 16362C 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL 573.3C	C1d
H3728	16169T 16209C 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
H3791	16129A 16185T 16223T 16325C 16327T 73G 146C 249DEL 263G 290DEL 291DEL 309.2C 315.1C 382A 489C 493G 514T	C1b
P3839	16223T 16290T 16319A 16362C 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2
P3792	16051G 16209C 16223T 16325C 16327T 16357C 16519C 73G 194T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d2a
H3792	16183C 16189C 16193.1C 16217C 16519C 73G 263G 269T 309.2C 315.1C 498DEL 499A	B4b
P3807	16051G 16093C 16223T 16256T 16298C 16325C 16327T 16519C 73G 194T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d+194
H3807	16051G 16223T 16256T 16298C 16325C 16327T 16519C 73G 194T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d+194
P3826	16182C 16183C 16189C 16217C 16519C 73G 263G 309.1C 315.1C 498DEL 499A 524.1A 524.2C	B4b
P5279	16111T 16213A 16223T 16290T 16319A 16362C 64T 73G 125C 127C 146C 153G 235G 263G 309.2C 315.1C 523DEL 524DEL	A2+64
P4213	16051G 16223T 16298C 16325C 16327T 16519C 73G 194T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d+194
H4213	16051G 16223T 16298C 16325C 16327T 16519C 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d
P5324	16111T 16223T 16290T 16319A 16356C 16362C 64T 73G 146C 153G 235G 263G 309.2C 315.1C 523DEL 524DEL	A2+64
P5679	16051G 16092C 16209C 16223T 16298C 16300G 16325C 16327T 16519C 73G 194T 249DEL 263G 290DEL 291DEL 315.1C 489C 523DEL 524DEL	C1d2a
P4288	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
H4288	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
H4453	16185T 16223T 16325C 16327T 16549.1C 73G 263G 290DEL 291DEL 309.1C 315.1C 382A 489C 493G	C1b
P4557	16223T 16325C 16357C 16362C 16519C 73G 109A 146C 263G 315.1C 489C	D1e
PP4645	16051G 16093C 16223T 16256T 16298C 16325C 16327T 16519C 73G 194T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d+194
H5679	16086C 16092C 16111T 16213A 16223T 16290T 16319A 16356C 16362C 73G 146C 153G 235G 263G 309.1C 309.2C 315.1C 523DEL 524DEL	A2
P5942	16086C 16223T 16295T 16298C 16325C 16327T 16519C 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
P4803	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
P4985	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
P5076	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
H5076	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
H6564	16108T 16111T 16213A 16223T 16290T 16319A 16356C 16362C 64T 73G 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64
P6750	16182C 16183C 16189C 16217C 16266T 16519C 73G 103A 200G 263G 309.1C 315.1C 318C 499A	B4b
P5325	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
P2 4668	16183C 16189C 16193.1C 16217C 16497G 16519C 73G 263G 309.1C 315.1C 499A 573.1C	B4b
P3633	16223T 16298C 16325C 16327T 16504A 73G 249DEL 252C 263G 276G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b

H3633	16051G 16129A 16223T 16234T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d
P6013	16169T 16223T 16298C 16325C 16327T 73G 185A 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
P6056	16051G 16223T 16298C 16325C 16327T 16519C 73G 194T 249DEL 263G 290DEL 291DEL 315.1C 489C 523DEL 524DEL	C1d+194
P4214	16086C 16185T 16223T 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 382A 489C 493G	C1b
P6336	16169T 16223T 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
H6336	16142T 16188T 16223T 16325C 16362C 16519C 73G 152C 225A 263G 309.1C 315.1C 489C	D1f
P6337	16169T 16223T 16298C 16325C 16327T 73G 185A 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
P6427	16142T 16188T 16223T 16325C 16362C 16519C 62T 73G 152C 225A 263G 309.1C 315.1C 489C	D1f
H6427	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.2C 315.1C 489C 493G 523DEL 524DEL	C1b
P6428	16051G 16223T 16298C 16325C 16327T 16519C 73G 194T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d+194
H6428	16169T 16223T 16298C 16325C 16327T 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
P6429	16111T 16213A 16223T 16290T 16319A 16362C 64T 73G 146C 235G 263G 309.2C 315.1C 523DEL 524DEL	A2+64
P4286	16223T 16274A 16298C 16325C 16327T 73G 185A 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C	C1
P1 4668	16051G 16093C 16223T 16256T 16298C 16325C 16327T 16519C 73G 194T 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 523DEL 524DEL	C1d+194
P6751	16086C 16223T 16295T 16298C 16325C 16327T 16519C 73G 249DEL 263G 290DEL 291DEL 309.1C 315.1C 489C 493G 523DEL 524DEL	C1b
H6767	16111T 16223T 16290T 16311C 16319A 16362C 64T 73G 143A 146C 153G 235G 263G 315.1C 523DEL 524DEL	A2+64
P6793	16111T 16213A 16223T 16290T 16319A 16362C 64T 73G 125C 127C 146C 153G 235G 263G 309.1C 315.1C 523DEL 524DEL	A2+64

Table 11 - F_{ST} genetic distances between groups sampled in northwest South America (Yang *et al.*, 2010) and the present study´s speaking groups.

	Embéra PS	Chibcha PS	Guambiano PS	Arhuaco	Cabecar	Embera	Guaymi	Kogi	Waunana	Wayuu	Zenu
Embéra PS		0.09910	0.00000	0.00000	0.11712	0.04505	0.04505	0.00000	0.32432	0.00000	0.00000
Chibcha PS	0.03543		0.00000	0.01802	0.05405	0.04505	0.03604	0.10811	0.117120	0.17117	0.02703
Guambiano PS	0.41090	0.21911		0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.09910	0.72973
Arhuaco	0.20084	0.18196	0.58019		0.00901	0.63063	0.171170	0.58559	0.00000	0.00000	0.00000
Cabecar	0.04348	0.09290	0.49774	0.28488		0.18018	0.25225	0.00901	0.80180	0.00000	0.00000
Embera	0.13831	0.16607	0.60623	0.00000	0.17489		0.64865	0.31532	0.01802	0.00000	0.00000
Guaymi	0.07610	0.11360	0.55655	0.05969	0.04576	0.00000		0.16216	0.18018	0.00901	0.00000
Kogi	0.16268	0.06960	0.44648	0.00000	0.22345	0.04977	0.07233		0.00000	0.00000	0.00000
Waunana	0.00306	0.06045	0.45030	0.28827	0.00000	0.19295	0.08163	0.22606		0.00000	0.00000
Wayuu	0.23187	0.04129	0.04870	0.40736	0.25179	0.40671	0.33800	0.25674	0.21810		0.49550
jgZenu	0.36295	0.15584	0.00000	0.56694	0.42857	0.58386	0.51250	0.40265	0.38055	0.00000	

PS stands for Present Study, all the other populations were gathered from reference (Yang et al., 2010). Significant values (P-value<0.05) are coloured in blue, significant values after Bonferroni correction (P-value<0.00090) are coloured in red.

Abstract for oral presentation in *IJUP*, Porto, February 2012

Insights into South-American colonization through mtDNA analysis in Colombian populations

C. Xavier [1, 2], J. Builles[3, 4], V. Gomes[1], J.M. Ospino[3], A. Amorim[1, 2], L. Gusmão[1], and A. Goios[1]

[1]IPATIMUP, Institute of Pathology and Molecular Immunology of the University of Porto, Portugal.
[2] Department of Biology, Faculty of Sciences, University of Porto, Portugal.
[3] Laboratorio Genes Ltda, Medellín, Colombia.
[4] Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

America was the last continent to be colonized by humans and it has been suggested that the first inhabitants arrived this continent at least 15000 years before present. The number of migrations and the dispersion routes through North America are still not clarified: while some studies propose a unique colonization route by land through Beringia and an ice-free corridor towards the south, others indicate the possibility of a sea-route entrance via Siberia and the Pacific coast.

The peopling of South America is even more controversial due to the persisting doubts on the number and relevance of the migrations associated with the dispersion of native populations in this subcontinent. However, the most accepted hypothesis is that the entrance was made through Colombia and then subdivided into two different routes, one following the Andes chain and the other into the Amazonian plains. Five centuries ago, when people from the old world, such as Europeans and lately the Africans, arrived in Colombia, the territory was already colonized by several ethnic groups of various linguistic families.

Aiming to add some clues on the above mentioned issues, studies with lineage DNA markers such as mitochondrial DNA (mtDNA) and Y chromosome have been performed. Lineage markers allow tracing back the history of populations because they are transmitted without recombination to the descendants; these lineages are grouped into haplogroups, distinguished by specific polymorphisms that tend to be geographically restricted. In the present study we analyzed the mtDNA of two populations in Colombia – Emberá-Chami (highland) and Cauca (lowland) – in order to determine the ancestry of the observed lineages, including those belonging to European and African haplogroups that could be explained by recent migrations, during and after the colonial period. Based on the analysis of the Native American haplogroups in both populations, we also intended to perceive if there are differences that could indicate different migrations towards the South of the continent.

The mtDNA control region was sequenced for 98 samples from the two populations (38 Emberá and 60 Cauca) and compared with the revised Cambridge Reference Sequence. Haplogroup frequencies were calculated and phylogenetic analyses were performed.

The majority of haplogroups found in both Colombian populations are typically Native American (A, B, C and D), which contrast with Y-Chromosome data for Cauca group, suggesting that the preferred mode of miscegenation was through Native American women and non-Native American men. Furthermore, our preliminary results show that

there are differences between the two populations, which may result from distinct ancient courses.

Abstract for oral presentation in *Portugaliae Genetica 16th*, Porto, March 2012

Insights into South-American colonization through mtDNA analysis in Colombian populations

C. Xavier [1, 2], J. Builles[3, 4], V. Gomes[1], J.M. Ospino[3], A. Amorim[1, 2], L. Gusmão[1], and A. Goios[1]

[1]IPATIMUP, Institute of Pathology and Molecular Immunology of the University of Porto, Portugal.
[2] Department of Biology, Faculty of Sciences, University of Porto, Portugal.
[3] Laboratorio Genes Ltda, Medellín, Colombia.
[4] Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

America was the last continent to be colonized by humans and it has been suggested that the first inhabitants arrived this continent at least 15000 years before present. The number of migrations and the dispersion routes through North America are still not clarified: while some studies propose a unique colonization route by land through Beringia and an ice-free corridor towards the south, others indicate the possibility of a sea-route entrance via Siberia and the Pacific coast.

The peopling of South America is even more controversial due to the persisting doubts on the number and relevance of the migrations associated with the dispersion of native populations in this subcontinent. However, the most accepted hypothesis is that the entrance was made through Colombia and then subdivided into two different routes, one following the Andes chain and the other into the Amazonian plains. Five centuries ago, when people from the old world, such as Europeans and lately the Africans, arrived in Colombia, the territory was already colonized by several ethnic groups of various linguistic families.

Aiming to add some clues on the above mentioned issues, studies with lineage DNA markers such as mitochondrial DNA (mtDNA) and Y chromosome have been performed. Lineage markers allow tracing back the history of populations because they are transmitted without recombination to the descendants; these lineages are grouped into haplogroups, distinguished by specific polymorphisms that tend to be geographically restricted. In the present study we analyzed the mtDNA of two populations in Colombia – Emberá-Chami (highland) and Cauca (lowland) – in order to determine the ancestry of the observed lineages, including those belonging to European and African haplogroups that could be explained by recent migrations, during and after the colonial period. Based on the analysis of the Native American haplogroups in both populations, we also intended to perceive if there are differences that could indicate different migrations towards the South of the continent.

The mtDNA control region was sequenced for 98 samples from the two populations (38 Emberá and 60 Cauca) and compared with the revised Cambridge Reference Sequence. Haplogroup frequencies were calculated and phylogenetic analyses were performed.

The majority of haplogroups found in both Colombian populations are typically Native American (A, B, C and D), which contrast with Y-Chromosome data for Cauca group,

suggesting that the preferred mode of miscegenation was through Native American women and non-Native American men. Furthermore, our preliminary results show that there are differences between the two populations with the Cauca population presenting higher levels of diversity than the Emberá, which could result from the isolation of the latter or from different sampling strategies. Both populations show differences from the data published in the literature, however our results corroborate the distinct patterns visible between northern and southern populations within Colombia that may result from distinct ancient courses.

Abstract for poster presentation in *I3s*, Póvoa de Varzim, May 2012

Insights into South-American colonization through mtDNA analysis in Colombian populations

C. Xavier [1, 2], J. Builles[3, 4], V. Gomes[1], J.M. Ospino[3], A. Amorim[1, 2], L. Gusmão[1], and A. Goios[1]

[1]Population Genetics Group, IPATIMUP, Institute of Pathology and Molecular Immunology of the University of Porto, Portugal, [2] Department of Biology, Faculty of Sciences, University of Porto, Portugal, [3] Laboratorio Genes Ltda, Medellín, Colombia, [4] Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

Aiming to add some clues on the colonization of the American continent, more precisely the entrance points and dispersion routes taken, studies with lineage DNA markers such as mitochondrial DNA (mtDNA) and Y chromosome have been performed. Lineage markers allow tracing back the history of populations because they are transmitted without recombination to the descendants; these lineages are grouped into haplogroups, distinguished by specific polymorphisms that tend to be geographically restricted.

In the present study we analyzed the mtDNA of two regions in Colombia – Antioquia composed of samples from one Emberá-Chami ethnic group population and Cauca constituted of several ethnic groups – in order to determine the ancestry of the observed lineages, including those belonging to European and African haplogroups that could be explained by recent migrations, during and after the colonial period. Based on the analysis of the Native American haplogroups in both populations, we also intended to perceive if there are differences that could indicate different migrations towards the South of the continent.

The complete mtDNA control region was sequenced for 98 samples from the two groups (38 Emberá from Antioquia and 60 samples from various ethnic groups from Cauca) and compared with the revised Cambridge Reference Sequence. Haplogroup frequencies were calculated and phylogenetic analyses were performed.

The vast majority of haplogroups found in both Colombian populations are typically Native American. Our results show that while in the Antioquia region, the Emberá population presents a very reduced number of haplotypes belonging to haplogroups A, B and D, the Cauca region is more diverse and has a significant percentage of C

haplogroup lineages. When dividing the Cauca group into smaller speaking groups it is visible that they are obviously distinct and behave as small populations that have suffered evolutionary forces along time such as genetic drift and bottlenecks. When comparing with other populations from literature, there is a notable proximity between Chibchan speaking groups, whereas non-Chibchan remain differentiated. Regarding a geographic separation, there is no visible substructure. Instead, distinct patterns are visible both in northern and southern populations within Colombia which may result from distinct ancient courses.

Abstract for poster presentation in *DNA in Forensics: Exploring the Phylogenies*, Innsbruck – Austria, September 2012

Insights into South-American colonization through mtDNA analysis in native Colombian populations

C. Xavier [1, 2], J. Builles[3, 4], V. Gomes[1], J.M. Ospino[3], A. Amorim[1, 2], L. Gusmão[1], and A. Goios[1]

[1]Population Genetics Group, IPATIMUP, Institute of Pathology and Molecular Immunology of the University of Porto, Portugal, [2] Department of Biology, Faculty of Sciences, University of Porto, Portugal, [3] Laboratorio Genes Ltda, Medellín, Colombia, [4] Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

Aiming to add some clues on the colonization of the American continent, more precisely the entrance points and dispersion routes taken, studies with lineage DNA markers such as mitochondrial DNA (mtDNA) and Y chromosome have been performed, which allow tracing back the history of populations because they are transmitted without recombination to the descendants.

In the present study we determined the matrilineal ancestry of samples from two regions in Colombia through mtDNA analysis. Based on the observation of Native American haplogroups in both populations, we also intended to perceive if there are differences that could indicate different migrations towards the South of the continent.

The complete mtDNA control region was sequenced for 98 samples from the two groups (38 Emberá from Antioquia and 60 samples from various ethnic groups from Cauca) and compared with the revised Cambridge Reference Sequence. Haplogroup frequencies were calculated and phylogenetic analyses were performed.

The vast majority of haplogroups found in both Colombian populations are typically Native American. Our results show that while in the Antioquia region, the Emberá population presents a very reduced number of haplotypes, all belonging to haplogroups A, B and D, the Cauca region is more diverse and has a significant percentage of C haplogroup lineages. When dividing the Cauca group into smaller speaking groups it is visible that they are obviously distinct and behave as small populations that have suffered evolutionary forces along time such as genetic drift and bottlenecks. When comparing with other populations from literature, there is a notable proximity between Chibchan speaking groups, whereas non-Chibchan remain differentiated. Regarding a geographic separation, there is no visible substructure.

Instead, distinct patterns are visible both in northern and southern populations within Colombia which may result from distinct ancient routes.