Characterization of Y chromosome diversity in Colombian Populations

Rodrigo Campos Ribeiro

Genética Forense

Departamento de Biologia

2022

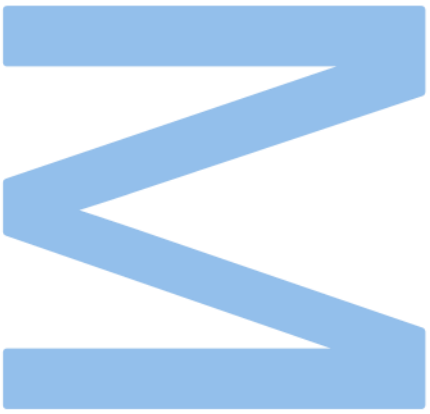
Orientador

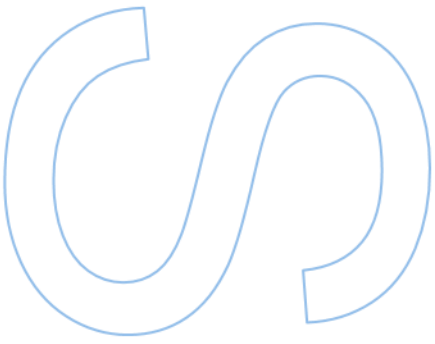
Maria João Prata, Professor Associado, Faculdade de Ciências da Universidade do Porto

Coorientador

Verónica Gomes, Junior Researcher, i3S, IPATIMUP


U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO





Declaração de Honra

Eu, Rodrigo Campos Ribeiro, inscrito no Mestrado em Genética Forense da Faculdade de Ciências da Universidade do Porto declaro, nos termos do disposto na alínea a) do artigo 14.º do Código Ético de Conduta Académica da U.Porto, que o conteúdo da presente dissertação de projeto reflete as perspetivas, o trabalho de investigação e as minhas interpretações no momento da sua entrega.

Ao entregar esta dissertação de projeto, declaro, ainda, que a mesma é resultado do meu próprio trabalho de investigação e contém contributos que não foram utilizados previamente noutros trabalhos apresentados a esta ou outra instituição.

Mais declaro que todas as referências a outros autores respeitam escrupulosamente as regras da atribuição, encontrando-se devidamente citadas no corpo do texto e identificadas na secção de referências bibliográficas. Não são divulgados na presente dissertação de projeto quaisquer conteúdos cuja reprodução esteja vedada por direitos de autor.

Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Rodrigo Campos Ribeiro

Porto, 30 de setembro de 2022

Agradecimentos

À Professora Doutora Maria João Prata, minha orientadora, pela oportunidade de orientação com rigor científico, e pela paciência, pelo interesse, pela disponibilidade e pela exigência, que foram imprescindíveis para o enriquecimento de todas as etapas subjacentes ao trabalho apresentado;

À Verónica, minha coorientadora, pela transmissão dos conhecimentos laboratoriais e *in silico* que foram essenciais para a realizam deste trabalho, pela paciência, pela disponibilidade e pelo interesse durante a realizarão deste trabalho;

À Leonor Gusmão, que mesmo estando noutro continente, ajudou-me através do seu conhecimento e dos seus conselhos na realização da tese;

A todos os elementos que constituem o grupo de Genética Populacional e Evolução do I3S, que durante este ano letivo se disponibilizaram a ajudar-me sempre que precisei e por me terem proporcionado um bom ambiente de trabalho;

À minha família, por me ter apoiado e mostrado durante todas as fases deste projeto, em especial à minha mãe e ao Américo, aos meus irmãos, ao meu pai e à minha avó por serem referências importantes para mim.

Aos meus amigos de faculdade e de escola, muitos que passaram pelo mesmo que eu durante este ano letivo e sempre apoiaram-me quando precisava. Um agradecimento especial à Inês Guedes, por também ter sido uma inspiração e durante o meu percurso académico todo, e por ser uma amiga que levo para a vida toda; à Beatriz Guimarães e à Inês Ferreira, por me darem mais um motivo para continuar a ser melhor e a orgulha-las; ao Zé Luís, o meu amigo mais antigo, que apesar de ainda não ter passado por isto sempre mostrou interesse e apoio no meu trabalho; e ao meu grupo – BioLamas - por terem estado sempre comigo durante estes anos todos e em muito que virão.

Ao meu mestre, Alberto Ramos, e a todos os membros e amigos da Academia Zen, cuja educação e crescimento na via marcial foi fundamental nos momentos mais difíceis e preocupantes.

Muito obrigado a todos!

Resumo

Estudos de genética populacional têm contribuído para um melhor entendimento da origem do homem moderno, da história das populações humanas e relações que estabeleceram ao longo dos anos. O conhecimento da composição genética das populações da América do Sul pode fornecer importantes pistas sobre o processo primordial de colonização deste subcontinente, assim como sobre as restruturações pelas quais passaram as populações sul-americanas em consequência do massivo influxo migratório que se registou na época colonial e pós-colonial.

A Colômbia atual é um país com uma população heterogénea e diversa, resultante da sua complexa história, geografia, topografia, demografia e diversidade de culturas. Dadas estas circunstâncias, a realização de estudos sobre a diversidade genética das populações deste país pode produzir informação valiosa para perceber a história dessas populações, nomeadamente no que respeita aos padrões de mistura que experimentaram num passado mais ou menos recente. As implicações desse conhecimento estendem-se a áreas como a genética clínica e forense. Tendo tudo isto em conta, o principal objetivo deste trabalho foi caracterizar e avaliar a diversidade de populações colombianas nativas e miscigenadas, através do estudo do cromossoma Y.

Dada a ausência de recombinação na extensão maioritária do cromossoma Y, a informação genética nela contida é transmitida de pai para filho sem alterações, salvo a ocorrência de mutação. Definem-se, assim, linhagens paternas que podem ser identificadas pelo estudo de polimorfismos presentes no cromossoma Y.

Este trabalho envolveu um estudo de 27 Y-STRs e 52 Y-SNPs num total de 121 amostras de indivíduos masculinos, não aparentados, pertencentes a 3 populações nativas, Guainía, Barí-Motilon e Paeces; e 4 populações não-nativas que surgiram de processos de mistura populacional que ocorreram pós-Colombo, nomeadamente das regiões administrativas de Cundinamarca, Huila, Norte de Santander e Tolima.

Quanto aos Y-STRs, foram identificados 108 haplótipos diferentes, nenhum partilhado entre populações distintas. Nos grupos nativos, a diversidade haplotípica era menor comparativamente à registada nas populações miscigenadas. Com base na informação obtida através dos Y-SNPs, identificaram-se 22 haplogrupos distribuídos pelas diferentes populações. O haplogrupo Q, característico de populações Nativo-Americanas, surgiu como

o mais frequente (41.322%), sendo seguido-pelo haplogrupo R, de origem Euroasiática (35.537%). Os restantes haplogrupos eram predominantemente de ancestralidade Euroasiática, E (6.61%), G (3.31%), I (1.65%), J (4.96%) e T (1.65%), embora também contemplassem linhagens de origem Africana, sempre com baixa frequência, nomeadamente as do haplogrupo A (0.83%) e de dois sub-haplogrupos do clado E, E-U174 e E-U290 (0.83% em ambos). De acordo com as ancestralidades inferidas, no conjunto das populações colombianas analisadas prevalecia a Euroasiática (56%), seguindo-se a Nativo-Americana (41%), enquanto que a Africana era apenas residual (3%). Estes valores diferiam consoante a população, sobressaindo, no entanto, a variação entre populações miscigenadas e nativas. Nas 4 populações miscigenadas as proporções de ancestralidade eram muito semelhantes, destacando-se claramente a fração de ~80% corresponde à Euroasiática, complementada por ~ 20% de ancestralidade Nativo-Americana (exceto na Cundinamarca, onde 10% das linhagens eram Nativo americanas e 10% Africanas). Nos grupos nativos, essas proporções divergiam muito, não só comparativamente a populações miscigenadas como também entre si. Nos Barí-Motilon linhagens de origem Ameríndia atingiram o valor máximo de 100%; nos Guainía, 72% das linhagens eram de origem Nativa, 22% Euroasiática e 6% Africana; nos Paeces a ancestralidade Nativo Americana era 33% e a Euroasiática 67%.

Análises comparativas envolvendo populações Nativas da América Latina revelaram que os Paeces se distinguiam bem das restantes populações Nativas, sendo entre os grupos nativos os mais próximos de populações Europeias.

Quanto às populações miscigenadas, quer da Colômbia quer de outros países da América do Sul, todas apresentavam grande afinidade com populações Europeias, o mesmo não acontecendo relativamente às Africanas.

No sentido de desfragmentar o componente euroasiático detetado na maioria das populações Colombianas, estas foram comparadas com outras populações, sobretudo da Europa e Médio Oriente, tendo ficado exposta a maior proximidade das populações Colombianas às da Europa Ocidental.

No global, os resultados obtidos neste trabalho mostram bem a heterogeneidade genética das populações colombianas, e atestam, por um lado, a forte influência que populações da Europa Ocidental tiveram na restruturação da diversidade populacional que atualmente caracteriza a Colômbia, e por outro, a escassa contribuição africana para essa remodelação, como, aliás, tem sido reportado para outas populeões da América Latina.

Palavras-chave: Genética populacional; Cromossoma Y; Y-STRs; Y-SNPs; Linhagens paternas; América Latina; Colômbia; Populações nativas; Populações miscigenadas.

Abstract

Population genetics studies are contributing to a better understanding of the origin of modern humans, the history of human populations and the relationships they established over the years. The knowledge of the genetic composition of populations from South America can provide important clues about the primordial process of colonization of the subcontinent, as well as about the restructuring that South American populations went through as a result of the massive migratory influx that occurred in the colonial and post-colonial.

Today's Colombia is a country with a heterogeneous and diverse population, which resulted from its complex history, geography, topography, demography and cultural diversity. Given this, carrying out studies on the genetic diversity of the country's populations can produce valuable information to understand the history of these populations, namely regarding the admixture processes they experienced in the more or less recent past. The implications of this knowledge extend to areas such as clinical and forensic genetics. Taking all this into account, the main objective of this work was to characterize and evaluate the diversity of native and admixed Colombian populations, through the study of the Y chromosome.

Given the absence of recombination on the majority length of the Y chromosome, the genetic information contained therein is transmitted from father to son unchanged, unless mutation occurs. Thus, paternal lineages defined that can be identified by the study of polymorphisms of the Y-chromosome.

This work involved the study of 27 Y-STRs and 52 Y-SNPs in a total of 121 male individuals, unrelated, belonging to 3 native populations, Guainía, Barí-Motilon and Paeces; and 4 non-native populations, namely from the administrative regions of Cundinamarca, Huila, Norte de Santander and Tolima.

In respect to Y-STRs, 108 different haplotypes were identified, none shared between the different populations. In native groups, haplotype diversity was lower compared to that registered in admixed populations. Based on the information from the Y-SNPs, 22 different haplogroups distributed across the different populations were identified. Haplogroup Q, characteristic of Native American populations, emerged as the most frequent (41.32%), followed by haplogroup R, of Eurasian origin (35.54%). The remaining haplogroups were

predominantly of Eurasian ancestry, E (6.61%), G (3.31%), I (1.65%), J (4.96%) and T (1.65%), although they also included low frequent lineages of African origin, namely those belonging to haplogroup A (0.83%) and to 2 sub-haplogroups of clade E, E-U174 and E-U290 (0.83% in both). According to the inferred ancestries, in the set of Colombian populations prevailed the Eurasian (56%), followed by the Native American (41%), while the African was only residual (3%). These values differed according to population, although the variation between admixed and native populations stood out. In the 4 admixed populations the proportions of ancestry were very similar, clearly prevailing the fraction of ~80% to Eurasian lineages, complemented by ~20% of others with Native American ancestry (except in Cundinamarca, where 10% of the lineages were Native American and 10% African). In native groups, these proportions differed a lot, not only compared to admixed populations but also among each other. In the Barí-Motilon, Amerindian lineages reached the maximum value of 100%; in the Guainía, 72% of the lineages were of Native origin, 22% Eurasian and 6% African; in the Paeces, Native American ancestry was 33% and Eurasian 67%.

Comparative analyses involving Native populations of Latin America revealed that the Paeces were well distinguished from the other Native populations, being among the native groups the closest to European populations.

As for the admixed populations, either from Colombia or from other South American countries, whereas they all showed great affinity with European populations, the same was not the case with African populations.

In order to dissect the Eurasian component detected in most Colombian populations, the latter were compared with other populations, especially from Europe and the Middle East, which has exposed the affinities between Colombian populations and those from Western Europe.

Overall, the results obtained in this work highlighted the genetic heterogeneity of Colombian populations, and further attested, on the one hand, the strong influence that Western European populations had on the restructuring of the population diversity that currently characterizes Colombia, and, on the other hand, the scarce African contribution to that reshaping. This has often been reported for other populations in Latin America.

Keywords: Population Genetics; Y chromosome; Y-STRs; Y-SNPs; Paternal lineages; Latin America; Colombia; Native Populations; Admixed Populations.

Index

Figure Index

Table Index

Abbreviations

AIM	Ancestry Informative Markers
AZF	Azoospermia Factor
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
InDel	Insertion/Deletion
YHRD	Y Haplotype Reference Database
ISFG	International Society of Forensic Genetics
ISOGG	International Society of Genetic Genealogy
kya	Thousands Years Ago
MDS	Multi-Dimensional Scaling
mRNA	Messenger RNA
MSY	Male-specific Region of the Y Chromosome
mtDNA	Mitochondrial DNA
PAR	Pseudo-Autosomal Region
PCR	Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic Acid
SBE	Single Base Extension
SMM	Stepwise Mutation Model
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
VNTR	Variable number tandem repeat

1. Introduction

1.1. Population and Forensics Genetics

Population Genetics is a field of Biology concerned with the origin, amount and geographical distribution of the genetic variation present in populations and the dynamics of change of this variation throughout space and time (Templeton, A. R. 2021). Variation leads to diversity among different populations, which might arise due to a variety of evolutionary mechanisms, such as mutation, recombination, genetic drift, natural selection, gene flow, migration and non-random (assortative) mating (Class, B., & Dingemanse, N. J. 2022, Staub, N. L. 2002). Many scholars tend to focus on the evolutionary forces that shaped genetic diversity in our own species (Homo sapiens), aiming to obtain a better understanding of our past and evolution throughout history. This knowledge is very important in various areas, such as medicine, anthropology, forensic genetics, among others in which concepts and models from Population Genetics are commonly applied.

The first instance of a genetic tool being used in forensics to discriminate individuals was the ABO blood group system, which was discovered in the early 1900s (Butler, J. M. 2005). Over the years, an increasing number of markers was identified, accompanying the advances in molecular biology techniques. Blood groups and electrophoretic genetic markers were the conventional markers until the 1980. In the meanwhile, the ability to analyze directly the DNA molecule afforded the emergence of DNA markers, which gradually replaced the other kind of markers. At the beginning of the DNA era, markers like VNTRs (Variable Number of Tandem Repeats), typed through RFLP-based approaches, gained ground, but unfortunately, the analysis methods at the time required large amounts of DNA, which represented a serious limitation (Goodwin, W. et al. 2011).

The development of the Polymerase Chain Reaction (PCR) by Kary Mullis (Mullis, K. et al. 1986) made possible to obtain large number of copies of DNA fragments, even from small amounts of template DNA, by amplifying a specific segment in only a few hours (Butler, J. M. 2005). With this revolutionary technique, DNA could be examined with great detail, and soon two new DNA markers were introduced in the research setting: the Short Tandem Repeats (STRs) and the Single Nucleotide Polymorphism (SNPs).

These two types of markers started to be used as the main source of variation in forensic genetic studies (Budowle, B., & Van Daal, A. 2008), and persist nowadays with different sets of STRs and/or SNPs available for the forensic routine. The information obtained by genotyping these markers provides a DNA profile for each individual, which can be compared with profiles from other samples that vary from case to case. For the evaluation of the DNA evidence, likelihood ratios between the probabilities of the hypotheses formulated must be calculated, viewing which the frequency of the profiles in a reference population need to be known. Due to this, a major endeavor has been the development of high quality DNA databases containing information regarding the markers employed in forensic genetics, which has been done recruiting data from population genetic studies performed in distinct regions of the world (Goodwin, W. et al. 2011).

1.2. DNA Markers

At the DNA level, a genetic marker can be defined as a DNA sequence that differs between individuals or populations and has a known physical location, acting thus as a chromosomal landmark (Shabir, Ghulam, et al. 2017). Genetic markers can be present in specific genes and account or not to the phenotype variability, depending on the position where they occur and on the type of the variations harbored. Most of them, however, reside in non-coding regions and usually do not have direct recognized impact on the phenotype of an individual, despite seldom cases demonstrated to have different consequences at the phenotypic level (Teama, Salwa 2018). Like genes, a genetic marker occupies a certain genomic location that is designated by locus (Collard, B. C. et al. 2005).

DNA markers arise through mutation, which is the original source of DNA variation. When the minor allele is present with at least 1% frequency in a population, the variation is classified as a polymorphism (Brookes, Anthony J. 1999) otherwise is only referred to as a rare mutation/variation (Lander, Eric S et al. 2001). Taking into account the frequency criterion, each DNA polymorphism must have at least two variants.

There are different kinds of DNA polymorphisms; actually they might involve a wide range of variations, from SNPs, variable number of tandem repeats (VNTRs: mini- and microsatellites), transposable elements (e.g. Alu repeats), structural alterations, and copy number variations.

In the last decades, the study of DNA markers increased extraordinarily along with the technical and methodological advances in fields such as biotechnology and molecular biology.

Genetic markers have been widely and successfully used in molecular genetic studies aiming at to identify genes associated with inherited diseases or with non-disease phenotypes, to evaluate genetic distances between populations, to solve questions in the scope of forensic genetics, among other goals.

In population and forensic genetics, the most commonly used DNA markers are multiallelic markers of the type STR, and biallellic markers (SNPs or small Indels) that typically only have two alleles.

1.2.1. STRs

Short Tandem Repeats (STRs), also known as microsatellites, are small sequences of DNA distributed across the genome in all eukaryote species (Brinkmann, B., et al., 1998). These sequences vary in length and complexity, being constituted by a variable number of copies of specific motifs encompassing one to six base pairs. In humans, there are ~4,500,000 STRs covering up to 2.5% of the human genome (Avvaru, A. K., et al., 2020). Whereas most of them are located in non-coding regions and do not affect gene expression, length variation in STRs residing inside introns or exons can play a significant role in the etiology of various human genetic diseases (Pearson, Christopher E. et al 2005).

STRs are very unstable, exhibiting a higher mutation rate (μ) than other type of genomic variations, around 10^{-4} to 10^{-2} mutations per generation (Baeta, Miriam, et al. 2018). They are associated to a main mutational process that consists in the insertion/deletion of one repeat motif (Forster, P. et al. 2015, Wierdl, M., et al., 1997). A common mechanism underlying the generation of these changes is replication slippage, also known as slipped strand mispairing, an event that can occur in the phase of cell division (Figure 1). Evolution of STRs is considered to fit reasonably a mathematical model called Stepwise Mutation Model (SMM) (Kimura, M., & Ohta, T. 1978) according to which mutations in STRs happen mainly by adding or subtracting a repeat unit. Although STRs mutations are normally single-step, some multi-step mutations are also observed (Dauber, E. M. et al. 2012). The mutation rate is directly correlated with the length and with the number of uninterrupted repeat units (Amos, William 2010).

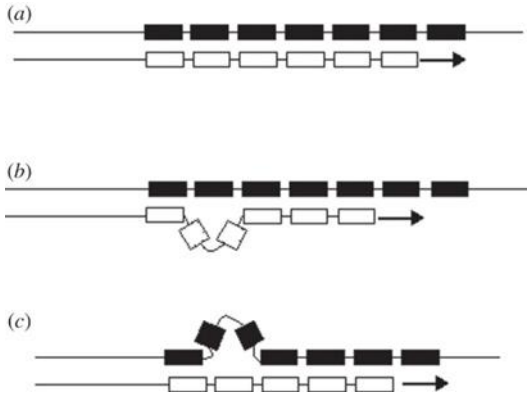


Figure 1: DNA strand slippage during replication of an STR locus. Boxes represent repetitive DNA units. Arrows indicate the direction in which a new DNA strand (white boxes) is being replicated from the original strand (black boxes). (a) Replication of the STR locus has proceeded without a mutation. (b) Replication of the STR locus has led to a gain of one motif owing to a loop in the new strand; the aberrant loop is stabilized by flanking units complementary to the opposite strand. (c) Replication of the STR locus has led to a loss of one unit owing to a loop in the template strand. (Forster et al. 2015).

STRs are typically multiallelic markers, with each specific allele being determined by the number of repeats in the sequence (Edelmann, Jeanett et al. 2002). This high variability together with the large number of STRs available, gives them an increased value in forensic genetics and population studies (Alves, Cíntia et al. 2003).

1.2.2. SNPs

Single Nucleotide Polymorphisms (SNPs) are the most common type of genetic markers. They arise by rare mutational events that alter a single nucleotide in the genome. As referred before, a polymorphism arise when a mutation is present with at least 1% frequency in a population (Brookes, Anthony J. 1999). Unlike the STRs, SNPs have a very low mutation rate, about 100 000 times lower that most STRs (around 10^{-9} mutations per generation) (Robino, Carlo et al. 2008, Xue, Yali, et al. 2009, Lang, M. et al. 2019), which means that SNPs are much more stable than STRs.

There are at least 10 million SNPs within the human genome, occurring approximately every 100–300 base pairs (Yue, P., & Moult, J. 2006). Similarly to STRs, the SNPs can be found throughout the genome, whether in regions of euchromatin or heterochromatin, in non-coding or protein-coding regions. It is well known that SNPs can influence all steps of gene expression depending on their genomic location. When present in non-coding regions, such as, per example, transcriptional regulatory elements, they can affect the mRNA structure and consequently its expression, increasing the risk to diseases, including some kinds of cancer (Lu, Yi-Fan et al. 2015, Li, Gongcheng et al. 2014).

SNPs within coding regions may also affect gene expression in numerous different ways, having potential to affect mRNA splicing, nucleo-cytoplasmic export, stability, and translation. Accordingly, an increasing bloom of studies continues to demonstrate the relationship between specific SNPs and many diseases (Robert, F. and Pelletier, J. 2018).

1.3. Uniparental Markers

Meiotic recombination is a process that greatly contributes to the level of genetic diversity in most sexually-reproducing organisms, including humans. By reshuffling the genetic material during meiosis, recombination can produce a vast number of different allele combinations between generations, accounting thus to the inter-individual genetic diversity (with the exception of monozygotic twins) (Ayub, Q. et al. 2016). While most of the human genome undergoes recombination, there are however two exceptions that are unique due to the lack of recombination: the male-specific region of the Y chromosome (MSY), and the mitochondrial DNA (mtDNA) (Pereira, V., & Gusmão, L. 2016, Butler, J. M. 2005).

Unlike the autosomes and the X chromosome, the Y chromosome and the mtDNA have a uniparental transmission, meaning that they are passed down to the next generation by only one parent (Figure 2) – the Y chromosome is transmitted exclusively from father to son, whereas the mtDNA is transmitted from the mother to all offspring (Jobling and Tyler-Smith 2003). Due to the lack of recombination and the haploid nature, either MSY or mtDNA are inhered intact over generations (unless mutation occurs) as a single locus. This implies that each individual carries a haplotype transmitted in block from one generation to the next (Nilsson, M., et al., 2008). Given the characteristics of MSY and mtDNA, both have a low effective population size (¼ of the autosomes) (Pereira, V., & Gusmão, L. 2016, Butler, J. M. 2005), which, comparatively to other genomic regions, makes them more prone to genetic drift, and so more susceptible to the effects of demographic episodes like founder effects or bottlenecks. Furthermore, the lineages defined by MSY or mtDNA markers tend to be geographically restricted, turning them very suitable to infer geographical ancestry (Butler, J. M. 2005).

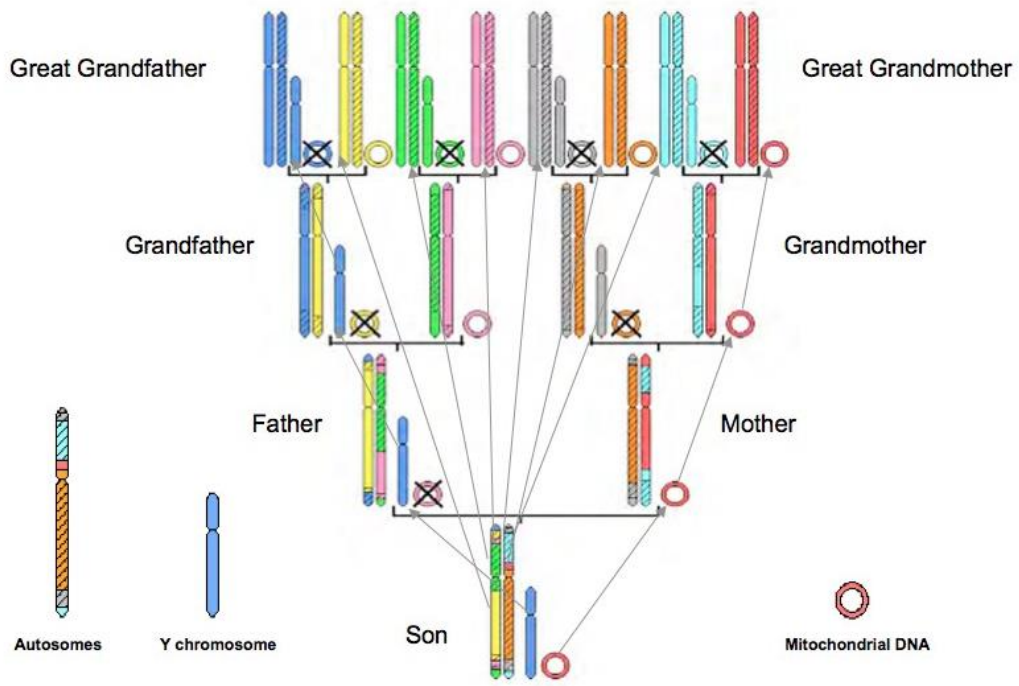


Figure 2: Transmission of biparental (autosomal) and uniparental markers throughout four generations. As we go down into younger individuals the autosomal markers get recombined between them, creating unique genomes, while uniparental markers are only inherited from the corresponding lineage (the Y chromosome from the father side of the family and the mtDNA from the mother side of the genealogy) without suffering any recombination, reaching the son unchanged. Adapted from: https://sites.google.com/site/bowessurnames/y-dna-results/types-of-dna-explained/visual-dna-transmission.

Uniparental markers have demonstrated to be very useful in population genetics and forensic studies. In the forensic field, the analysis of autosomal markers is the usual routine for DNA profiling, and is enough in most investigations. However, there are cases that require complementary information, namely from the X and Y-chromosomes or from mtDNA. Concerning uniparental markers, there are many instances where their study might be crucial. For over 30 years, analysis of mtDNA has been used routinely in forensic investigations dealing with challenging samples such as hair shafts (which do not contain nuclear DNA) and ancient or aged skeletal material, like poorly preserved bones and teeth, which are samples where DNA is highly degraded (Canale, L.C., et al., 2022). In these circumstances, mtDNA might be the unique source of DNA for analysis due to its high number of copies in each cell (and consequently high availability) compared to nuclear DNA (Butler 2005; Nilsson et al. 2008; Pereira, V. & Gusmão, L. 2016).

As for Y chromosome, its analysis has clear value in cases of sexual assault, particularly when the amount of material from a male perpetrator is scarce in a female background. It can also inform on multiple male contributors to a trace (de Knijff, P., 2022). Furthermore, the analysis of Y chromosome can assist in the inference of geographical ancestry and in familial

reconstruction, similarly to mtDNA, since both are lineage markers (Syndercombe Court, D. 2021). In paternity disputes of male offspring and other types of paternal kinship testing, where the complete DNA profiles are not available, the Y chromosome also can be a crucial tool (Jobling, M. A., & Gill, P. 2004),.although under the limitation of disclosing only paternal lineages. For example, if a putative father is deceased or missing, analysis of Y chromosome in a paternally related family member can help support the relationship (Syndercombe Court, D. 2021).

In population genetics, uniparental markers are giving since long important insights on the history of different population groups, being able to provide evidence on the origin, spread, admixture and many other events underling their demographic history (Petit, E. et al. 2002).

1.4. Y chromosome

1.4.1. Evolution

In Eutherian mammals, the biological sex is associated to an XX/XY sex chromosome system (Bachtrog, D., 2006). The two sex chromosomes were originated from a pair of ancestral autosomes that evolved into distinct chromosomes (X and Y). Females are homogametic with two X chromosomes, while males are heterogametic with one Y and one X chromosome (Rice, W. R., 1996, Ross, M.T., et al. 2005).

The evolution of the Eutherian sex chromosome started around 180 million years ago (Cortez, D., et al. 2014) to become nowadays highly differentiated in both size and gene content. A key step in the differentiation of the Y chromosome was the acquisition of a testis-determining gene. This was followed, assumedly since the early phases, by numerous inversions and transpositions in the chromosome that hindered recombination with the X. Only small telomeric regions of the Y escaped these kind of events, which were the short segments where the X and Y needed to pair during meiosis to guarantee correct chromosomal segregation. In the great extension of the proto-Y, recombination was progressively suppressed leading to the degeneration of the chromosome (Bickmore, W. A. and Cooke, H. J., 1987, Schwartz, A., et al. 1998).

As a result, the Y experienced a major loss in gene content and expression, contrary to the X chromosome that retained most of the original gene content and size. Along with gene decay, the Y chromosome has accumulated many DNA repetitive sequences (Bachtrog, D., 2013).

1.4.2. Structure

The human Y chromosome has around 57 million base pairs, encompassing 71 protein coding genes and 109 non-coding genes according to the Ensembl database (http://www.ensembl.org/). It is an acrocentric chromosome with a centromere separating a short arm (Yp), very small compared to the much larger long arm (Yq) (Krausz, Csilla et al. 2006).

There are two distinct regions in the human Y chromosome that can undergo recombination, the Pseudo-Autosomal Regions 1 and 2 (PAR1 and PAR2). These regions are located at the telomeres of the short and long arms, respectively. Between the PARs stands a large non recombining region - the MSY - that represents around 95% of the Y chromosome. In this region are present several genes that produce long and short noncoding RNAs (whose effects on regulation of gene expression remain not deciphered) and 27 protein-coding genes. From these coding genes, 14 are considered testis specific (such as the specific testis-determining factor SRY) or show predominant expression in specific tissues such as the brain or the thyroid; (Maan, A. A. et al. 2017) (Guo, Xihan et al. 2020). It is in MSY that are located 3 loci identified in the early 1990' as being involved in spermatogenesis, then named Azoospermia Factor 1, 2 and 3 (AZF 1, 2 and 3), in which deletions have been associated with varying degrees of spermatogenic failure in men (Figure 3).

The PARs are the only sections of the Y-chromosome with homologous regions in the X, and so the PARs of the sex chromosomes can pair and recombine during meiosis. However, while the recombination rate in PAR1 is around ~17-fold times higher than the average in autosomes, the PAR2 rate is much lower, being similar to the genome average (Hinch, A. G. et al. 2014).

The two PARs contain at least 29 genes with diverse roles in cell signaling, transcriptional regulation and mitochondrial function (Colaco, S., Modi, D. 2018, Raudsepp, Terje et al. 2015).

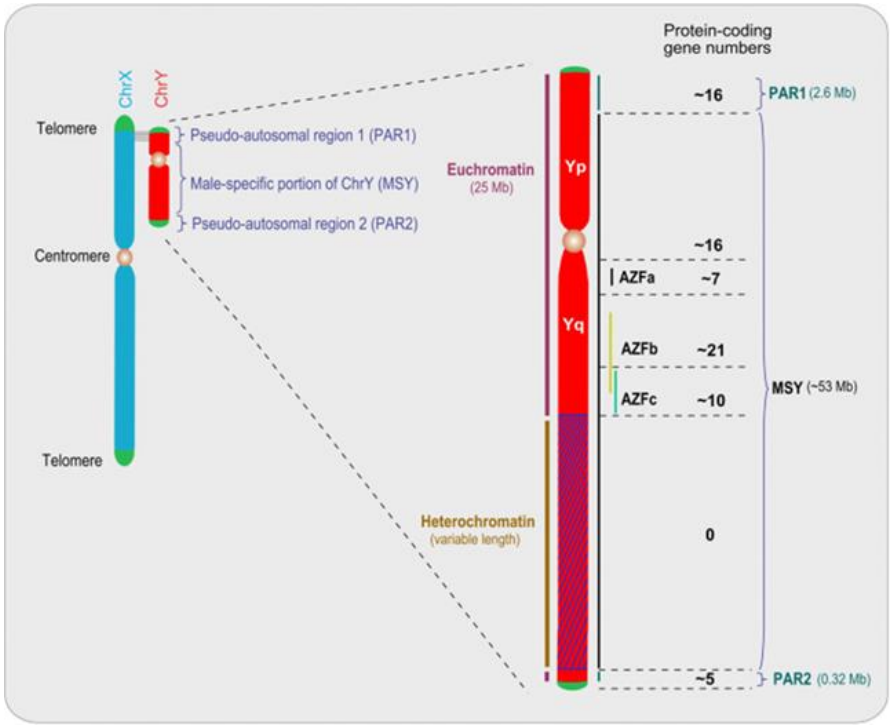


Figure 3: Structure of human Y chromosome (ChrY) and relative length comparison with the X chromosome. In the euchromatin of Yq are contained the Azoospermia factors (AZFa, AZFb and AZFc). Adapted from Guo, Xihan et al. 2020.

1.4.3. Y chromosome polymorphisms

Like in other genomic regions, the most commonly studied polymorphisms in the Y chromosome are multi-allelic markers of the kind STR; and bi-allelic markers mainly belonging to the SNP category, but also comprehending some small Insertions/Deletions (InDels) (Goodwin, W. et al. 2011; Pereira, V. & Gusmão, L. 2016).

Many STRs are located in the Y chromosome (Y-STRs). Those lying in the MSY since long are being used in a variety of forensic investigations or to address questions in the field of population genetics (Gusmão, L. et al. 2006, Roewer, L. 2019). Given that MSY is a haploid region, the different alleles identified for specific sets of Y-STRs define a Y-haplotype, which is inherited along the paternal lineage. Currently, a number of commercial Y-STR kits is available, each of one allowing the definition of Y-STR haplotypes based on the loci contained that can vary not only in number.

Estimating accurate frequencies of Y-haplotypes in populations has been a subject of great work that prompted its compilation in specific databases. Among them stands the Y Haplotype Reference Database (YHRD), which is currently recommended by the International Society of Forensic Genetics (ISFG) as the reference source to obtain the data needed to the interpretation of Y-STR results in forensic analysis (Roewer, L. et al. 2020).

The information available at YHRD can be retrieved searching six different haplotype formats that vary in loci number ranging from the least discriminative with 9 loci (Minimal haplotype) to the most discriminative with 29 loci (Y-Max haplotype). The remaining four haplotype formats accommodate the following commercial kits: PowerPlex® Y12 (Promega©) (12 loci), YFiler® (Applied Biosystems™) (17 loci), PowerPlex® Y23 (Promega©) (23 loci), YFiler® Plus (Applied Biosystems™) (27 loci) (Table 1), which are the kits nowadays more standardized. Besides, the YHRD also allows the search of partial Y-haplotypes defined by even more enriched Y-STR kits, such as Argus® Y-28 (QIAGEN®) (28 loci), GoldenEye® (Goldeneye Technology Ltd.) (37 loci), PathFinder® Plus (SureID®) (37 loci), AGCU® Y37 (AGCU ScienTech Incorporation) (37 loci) or the YFiler® Platinum (Applied Biosystems™) (37 loci). These later are kits containing a few loci without data yet not deposited in YHRD, which up to now just contemplates information for 29 Y-STR markers.

Some of those kits include rapidly mutating Y-STRs (RM Y-STRs), which are characterized by higher mutation rates than the average estimated for Y-STRs ($\mu > 10^{-2}$ per generation) (Ballantyne, Kaye N., et al. 2010, Alghafri, R., et al. 2015, Adnan, A., et al. 2016). It's the case of the PowerPlex® Y23 and YFiler® Plus kits, that contain two and seven RM Y-STRs respectively (Turrina, S., et al. 2015, Rapone, C., et al. 2016, D'Atanasio, et al. 2019). Although RM Y-STRs are typically associated with high discrimination power in forensics caseworks involving male individuals, they were only recently introduced in forensic or population investigations, justifying the absence or scarcity of data in current databases (Ballantyne, K. N., and Manfred, K., 2012).

Table 1: Haplotype formats presented in the YHRD database with the respective number of loci and specific Y-STRs.

HAPLOTYPE FORMAT	LOCI NUMBER	Y-STR LOCI
MINIMAL	9	DYS19 DYS389I DYS389II DYS390 DYS391 DYS392 DYS393 DYS385
POWERPLEX® Y12	12	DYS391 DYS389I DYS439 DYS389II DYS438 DYS437 DYS19 DYS392 DYS393 DYS390 DYS385
YFILER®	17	DYS456 DYS389I DYS390 DYS389II DYS458 DYS19 DYS385 DYS393 DYS391 DYS439 DYS635 DYS392 YGATAH4 DYS437 DYS438 DYS448
POWERPLEX® Y23	23	DYS576 DYS389I DYS448 DYS389II DYS19 DYS391 DYS481 DYS549 DYS533 DYS438 DYS437 DYS570 DYS635 DYS390 DYS439 DYS392 DYS643 DYS393 DYS458 DYS385 DYS456 YGATAH4
YFILER® PLUS	27	DYS576 DYS389I DYS635 DYS389II DYS627 DYS460 DYS458 DYS19 YGATAH4 DYS448 DYS391 DYS456 DYS390 DYS438 DYS392 DYS518 DYS570 DYS437 DYS385 DYS449 DYS393 DYS439 DYS481 DYF387S1 DYS533
YMAX	29	DYS19 DYS389I DYS389II DYS390 DYS391 DYS392 DYS393 DYS385 DYS439 DYS438 DYS437 DYS456 DYS458 DYS635 YGATAH4 DYS448 DYS576 DYS481 DYS549 DYS533 DYS570 DYS643 DYS627 DYS460 DYS518 DYS449 DYF387S1

Just like Y-STRs, there are also many biallellic markers in the Y chromosome (Y-SNPs and Y-Indels) being used as a powerful tool both in population genetic studies and in the forensic setting (de Knijff, P. 2022). Since SNPs share with small insertion/deletions a similar low mutation rate, both kind of markers are commonly assumed to have arisen by unique mutation events that allow to define Y-haplogroups. The phylogenetic relationships between different Y-haplogroups have been used to construct hierarchical trees of the Y-chromosome diversity (Figure 4). These trees are permanently updated as long as new Y-markers appear, mainly Y-SNPs, mostly discovered due to the growing use of next-generation sequencing approaches in recent times (Larmuseau, M. H., et al. 2015).

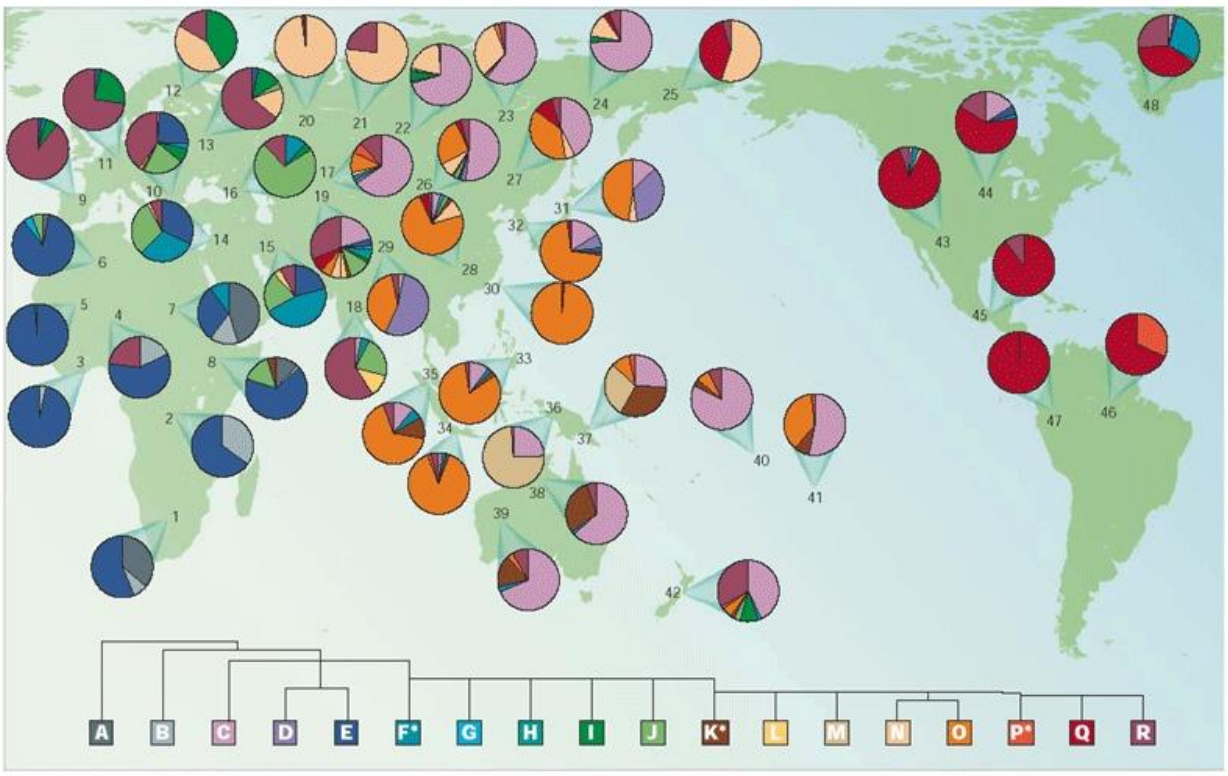


Figure 4: Global distribution of Y haplogroups. Each circle represents a population sample with the frequency of 18 Y haplogroups (shown here in simplified form) identified by the Y Chromosome Consortium (YCC) indicated by the coloured sectors. Adapted from: Jobling and Tyler-Smith 2003.

Initially, Y-haplogroup nomenclature was a confusing issue, because different systems were proposed and co-existed for some time, complicating the comparison of data from different studies. The problem was partially overcome with the work of the Y Chromosome Consortium, which in 2002 published a seminal paper addressing the nomenclature system for the tree of human Y-chromosomal haplogroups (Y Chromosome Consortium, 2002), and from then on it was widely adopted a standardized nomenclature (Redd, A. J. 2002). The original tree included 245 Y-SNPs defining 153 haplogroups, with the major clades being labeled from A to R. Soon after, new Y-SNPs were characterized increasing the resolution of the tree and expanding the repertoire of major Y-haplogroup clades to S and T (Karafet, T. M et al. 2008).

The number of novel Y-SNPs continued, and continues, to increase extraordinarily, adding complexity to the ever-expanding Y-tree. However, many Y-SNPs are redundant due to being phylogenetically equivalent to other defining markers, which again represents a problem in comparative studies (Van Oven, M., et al. 2014).

Over the years, several online resources where created providing reference phylogenies for the human Y-chromosome. Among the public databases more used currently, is that maintained by the International Society of Genetic Genealogy (ISOGG; http://www.isogg.org/tree) that is a source very complex that encompasses a large number of yet to be confirmed Y-SNPs; and the PhyloTree (http://www.phylotree.org/Y/) (Van Oven, M., et al. 2014), which contains only a minimal tree (Table 2) .While in both these and other online versions of the Y-phylogeny a general agreement exists in the definition of the basic haplogroups (A, B, C, etc.), parallel nomenclatures for some sub-clades are in use.

Table 2: Major clades, their mutations and location within the Y chromosome and geographic distributions. Adapted from http://www.phylotree.org/Y/.

Haplogroup	Location on the Y chromosome	Mutation	Geographic Distribution
A00 – L1086	2826312	A>T	Central Africa
A0-V148	6788191	G>A	Central Africa, West Africa
A1-M31	21739754	G>C	West Africa, North Africa
A2-V50	6845936	T>C	Southern Africa, Central Africa
A3-M32	21740436	T>C	East Africa, Southern Africa
B-M60	21878072..21878073	ins T	Central Africa, Southern Africa, East Africa
D-M174	14954280	T>C	East Asia
E-M96	21778998	C>G	Africa, West Asia, Southern Europe
C-M130	2734854	C>T	Central Asia, Northern Asia, North America, East Asia, Southeast Asia, Wallacea, Near Oceania, Australia, Remote Oceania
G-M201	15027529	G>T	West Asia, Europe, Central Asia
H-L901	17844304	C>T	South Asia
I-M170	14847792	A>C	Europe, West Asia
J-M304	22749853	A>C	West Asia, North Africa, Horn of Africa, Southern Europe, Central Asia, South Asia
M-P397	28759645	C>T	Near Oceania, Wallacea, Australia, Remote Oceania
Q-M242	15018582	C>T	Northern Asia, Central Asia, Americas
R-M207	15581983	A>G	Europe, West Asia, Central Asia, South Asia, North Africa, Central Africa
N-M231	15469724	G>A	Northern Asia, Northern Europe
O-M175	15508706..15508710	del TTCTC	East Asia, Southeast Asia
L-M20	21733454	A>G	South Asia, West Asia
T-M184	14898163	G>A	West Asia, Horn of Africa, North Africa, Southern Europe, South Asia

Evidence accumulated since the first population screenings based on sets of Y-SNPs, clearly demonstrate a strong geographical structure in the distribution of many Y-haplogroups. Actually, there are Y-haplogroups very frequent in specific populations or even restricted to some populations, meaning thus that they are rarely found or absent elsewhere. For example, while the haplogroup Q (or Q-M242) is associated with Natives from the American continent, the haplogroup O appears more frequently in Southeast Asia. So, information extracted from Y-haplogroups can help to recuperate many episodes underlying the demographic history of populations, including bottlenecks, founder effects, migration influences, among others (Lang, M. et al. 2019).

It is nowadays rather well-known the world pattern of distribution of Y-haplogroups, making them very informative in analyses involving biogeographic ancestry inference (de Knijff, P. 2022), a kind of analysis that is becoming an important part of the routine in forensic investigations (Tvedebrink, T. 2022).

The available Y-chromosome trees and the use of appropriate nomenclature, are indispensable to perform classification into haplogroups. Y-haplogroup assignment might be based on the manually integration of a given set of Y-SNPs analyzed in the most updated Y-chromosome tree. Alternative to this approach, haplogroups can be predicted taking into account the haplotypes defined by Y-STRs. Viewing that, different software packages were implemented in the last years, such as the Haplogroup Predictor (http://www.hprg.com/hapest5/) or the NevGen (https://www.nevgen.org/). These and others Y-STR-based tools permit to predict Y-haplogroups, albeit with variable degrees of accuracy depending on the different algorithms and databases used (Lee, E.Y. et al. 2014, Muzzio, M., et al. 2011).

1.5. Human Colonization of the Americas

1.5.1. Early settlements

After modern humans started the expansion out of Africa the last land mass to be colonized was the American continent, believed to have happened when Siberian hunters were able to cross the Bering Strait around 16'000 years ago or even earlier (Dixon, E.J., 2001, Bortolini, M.C., et al. 2003, Meltzer, D.J., 2013) (Figure 5). America is divided in 2 subcontinents, North and South America, connected by the narrow Panama Isthmus, a small strip of land that allowed the sharing and enrichment of terrestrial flora and fauna in the two enormous land masses (Marshall, L.G., et al. 1982, Leigh, E.G., et al. 2014, O'Dea, A., et al. 2016). The Isthmus also represented a fundamental crossroad of the initial migrations of Native Americans who expanding from the North subcontinent rapidly reached the southern South America by ~ 14'500 years ago (Capodiferro, et al. 2021).

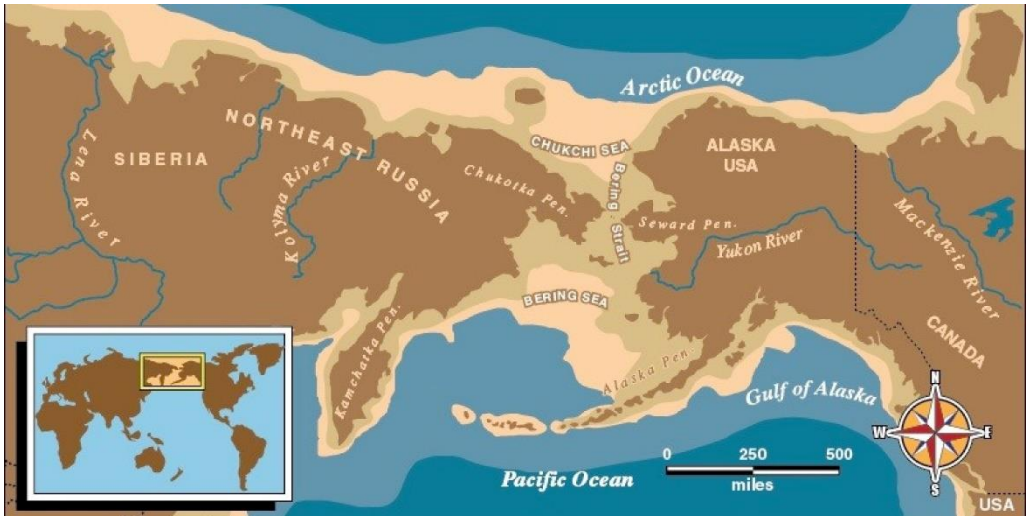


Figure 5: Bering Land Bridge around 16 000 years ago. Adapted from: https://www.nps.gov/bela/learn/historyculture/the-bering-land-bridge-theory.htm

There are still many doubts regarding the initial phases of the human settlement in South America, which remain open despite the recent data coming from analyses of ancient DNA (Casas-Vargas, A., et al. 2011, Roca-Rada, X., et al. 2021). However, several lines of evidence indicate that soon after the first Paleoindian settlers entered South America, the expansion towards south took place through two main paths: some groups followed the Pacific costal route, while others went eastwards, crossed the Amazon basin and proceeded the journey south along the Atlantic coastline (Gómez-Carballa et al. 2018).

Interior river ways also played a major role in this dispersion. It was a period of intense and dynamic human movements, which ultimately led to a very rapid peopling of South America. The whole process was accompanied by a series of bottlenecks that deeply accounted to the low genetic diversity typically found nowadays in Native American populations comparatively to other populations from around the world (Castro e Silva el at. 2022, Sutter 2021).

Besides the traces assumed to have been left by the early wave(s) of migrants that crossed the Bering Strait and reached America, a signal was also detected of genetic affinity between extant or/and ancient natives from South America and present-day indigenous groups of South Asia, Australia, and Melanesia (Castro, M. A et al. 2021). Many uncertainties persist concerning this intriguing Australasian ancestry (also referred as the “Y pikuéra population” or “Y population” component) (Raghavan, Maanasa, et al. 2015), which was first identified exclusively in the present-day Amazonian populations (Matisoo-Smith, E. and J.M., Ramirez 2010, Malaspinas, A.S., et al. 2014). Recently, however, this component was recently also found in the Pacific coast region, indicating a more widespread signal distribution within South America (Castro, M. A et al. 2021).

1.5.2. Post-Columbus America

Since the earliest days of the Age of Exploration, the increasing arrival of European and African individuals to America unprecedentedly disrupted the Native-American societies, causing a deep reconfiguration of the human populations from the continent. Complex patterns of intermarriage and cohabitation were at the origin of the numerous admixed populations that currently exist in America (Wang, S., et al. 2007, Salzano, F.M., and Sans, M., 2014). The admixture processes unfolded differently in distinct areas due to factors such as: geographical/topographical features, which could provide more or less conditions for the isolation of the ancestral indigenous populations; sources and patterns of migration, which varied considerably from region to region depending on temperature; fauna and flora; presence of volcanic activity; among others (Ossa, H., et al. 2016).

1.6. Colombia

1.6.1.1. A brief history

Of all current-day countries from South America, the only one that has contact with the Panama Isthmus is Colombia (Figure 6). This geolocation favored the settlement of various natives in a region that corresponds in great part to modern-day Colombia, contributing thus to partly explain the great heterogeneity typically found in present-day Native populations from the country (Xavier, C., et al. 2015, Ossa, H., et al. 2016).

Figure 6: Colombia's location (in green) within the South America subcontinent and its contact with the Panama Isthmus (From: https://www.dol.gov/agencies/ilab/country/ilab-colombia).

In pre-Columbus times, many aboriginal groups settled in the territory that corresponds to nowadays Colombia. According to different lines of evidence, they integrated a patchwork of several cultures and subcultures, among which seemingly no one dominated. It is assumed that those groups belonged to one of 3 major linguistic groups: Arawak, Carib, and Chibcha (Greenberg, J. H., 1987).

Suddenly, the history of these native groups was abruptly reconfigured with the discovery of the New World by Christopher Columbus in the late 15th century.

Colombia, where most of the first European settlers were Spanish, was declared a Spanish Colony by mid-16th century (Demy, T. J., & Shaw, J. M. 2019). Among the Spanish arriving to the region, a substantial proportion came from the Basque region (Gitelman, HM 1978, Pastor, JMA 2004). It is documented that Basques integrating the Spanish immigrants in the XVI and XVII centuries had an important impact in the development of the country's economy and administration, as well as in the change of the population structure in certain urbanized regions, such as Bogotá (Gitelman, HM 1978, Pastor, JMA 2004).

Conflicts between colonists and indigenous people (usually violent and bloody) started right from the first contacts, but another major consequence of the New World discovery was the introduction of infectious disease-causing agents to which indigenous Americans had not been previously exposed. Overall, many native groups were push into collapse, leading to a dramatic reduction of Native American ancestry (Merbs, C.F., 1992, Cook, N.D., 2002).

The sporadic attempts of the Spanish government to protect the Native Americans from the colonizers' mistreatments, including physical protection and granting of land-owing rights in the XVI century (LaRosa and Mejía 2017), were innocuous to alter the fate of the Native-American living in Colombia.

In the beginning of the XIX century, with the Spanish political unrest caused by Napoleon's invasion of the Iberian Peninsula and subsequent arrest of the king Ferdinand VII in France, all colonies demanded for more autonomy in their own government while still staying loyal to the Spanish Crown (LaRosa and Mejía 2017). This prompted the formation of the "juntas", which were assemblies created by authorities and prominent citizens that oversaw the administration of a specific city or region. However, rebellions against the colonial dominion in the Americas intensified and independence movements erupted everywhere, culminating in 1819 when the General Simón Bolívar declared the independence of the Gran-Colombia state, which encompassed the territories of modern-day Colombia, Ecuador, Venezuela and parts of other countries such as New Granada (in Panamá), Peru, Brazil and Guyana (Figure 7). This new state only lasted until 1831, when Venezuela and Ecuador seceded from it (Brown, Matthew 2012). Modern Colombia is one of the countries that emerged after the dissolution of Gran Colombia.

Figure 7: Territories of modern-day countries that were encompassed in the Gran Colombia state (From: https://commons.wikimedia.org/wiki/File:Gran_Colombia.GIF).

By the end of the XIX century and early new century, oversea migration has increased massively, with a large number of immigrates making his way over to South America (King, R., et al. 2000), where people from around the world searched for better work and living conditions. Colombia was no exception; many Europeans flowed into there, including Spanish, Italian, French and especially German immigrants in the late XIX and early XX centuries. After World War II., Colombia also received many refugees, once the Colombian government welcomed both Jews and Nazi escapees banned from different countries, including from northern America, namely from the USA (Rausch, Jane M 2021).

Along European immigrants, people from the Levant also found its way into Colombia. This wave was largely fueled by people from Arab countries, embracing refugees from Syria, Lebanon, Jordan and Palestine escaped from the Turkish Ottoman overlords and/or from the various independence conflicts occurring during this period (Fawcett, L et al. 1997, Bruckmayr, Philipp 2010). The arrival of Arab Middle Easterners triggered a considerable reconfiguration of the populations already settled in Colombia, not only genetically, but also social and religiously. These new communities were primarily established in coastal cities (Bruckmayr, Philipp 2010). Still in the early 20th century, small groups of Asian immigrants arrived into Colombia. They came mainly from Japan, most often settled in rural areas and many excelled in agricultural activities (Martinez-Martin, AF 2017).

A second wave of migration into Colombian territories occurred between the 1970's and 1990's, mostly from southern Europe countries (such as Spain, Portugal or Italy), the Levant and Asia (Weinstein, Asher 2021). These new incomers together with intense inter-country migrations contributed to an increase of the urbanization rate in Colombia, making it one of the most urbanized countries in Latin America by the years 2000. The accelerated urbanization demanded the industrialization of the agriculture activities, and especially in urban areas the industrialization process forced the energy sector in such a way that Colombia turned into one of the countries with the biggest emissions of greenhouse gas (Vélez-Henao, JA et al. 2020).

In recent years, Colombia has been receiving a large number of immigrants coming from its neighboring country Venezuela. It is estimated that 1.8 million people have crossed the border between 2015 and 2019, representing around 6% of the Venezuelan population (Freier, LF et al. 2019). This massive exodus, driven by drastic increases in violence, poverty and loss of social and economic freedoms in Venezuela, gave rise in Colombia to an outburst of anti-immigrant movements, reflected in protests, negative portrayal in media and even in the political discourse. While in fact, trust towards foreigners decreased in Colombia over the last years, notably, the recent study of Lebow, Jeremy et al. 2021 did not found evidence that proximity to migrants had a causal effect on trust, but instead suggested that under appropriate conditions, proximity to immigrants could improve cooperative attitudes (Lebow, Jeremy et al. 2021).

Current Colombia is the 25th largest country in the world, with an area around 1 140 000 km^2 divided in 32 administrative departments (Figure 8). Its capital is Bogotá, located in the Cundinamarca department. The different Colombian departments vary greatly in area, population and urban centers. The largest and less populated ones being located at the South and Southeast regions encompassing the fringes of the Amazon rainforest, whereas the most urbanized and populated departments stand at the Center (crossed by the Andean Mountains), West (along the Pacific coast) and North (along the Atlantic shore).

Figure 8: Administrative map of Colombia. From: https://co.pinterest.com/pin/619596861215942222/.

1.6.2. Population Composition

Colombia is also one of the most population diverse countries in the South America, which surmounts a total around of 50 million people. It is a multi-ethnic population with major ancestry contributions from Europe, Africa and America. According to Colombian census data, 86% of the citizens self-report to be of mixed ancestry of European and Native Americans (Mestizos), 10.5% of African-Colombian ancestry and 3.4% of Native American ancestry (Rojas, Winston et al. 2010). These general percentages vary from region to region, depending on topographical, meteorological and natural fauna/flora conditions that hinder or facilitate human settlement and development of urban or rural centres. The distribution of the ethnic composition also varies across the country. Groups with high African ancestry are most common in the departments located along the Pacific side, in the west and in the Northern coast (Ibarra, A., et al. 2014, Ossa, H., et al. 2016), whereas those with predominant European ancestry are mainly found in the North and Center of Colombia (Ossa, H., et al. 2016).

1.6.3. Ethnic Groups

Nowadays the number of indigenous people who live in Colombia is very low. As previously mentioned, only around 3% of the total population is considered to comprise Native Colombians, whereas the remaining proportion consists of admixed people between Amerindian, European and African ancestries (Rojas, W., et al. 2010, Ossa, H., et al. 2015, Xavier, C., et al. 2015). Despite the low percentage of indigenous people, there are over 80 Native groups in Colombia that maintain alive 65 indigenous languages, classified in 13 linguistic families and 7 isolated languages. Most of those languages are only spoken in isolated communities with very little admixture with non-Natives (Figure 9) (Arango and Sánchez 2004, de Pérez, M. S. G. 2010).

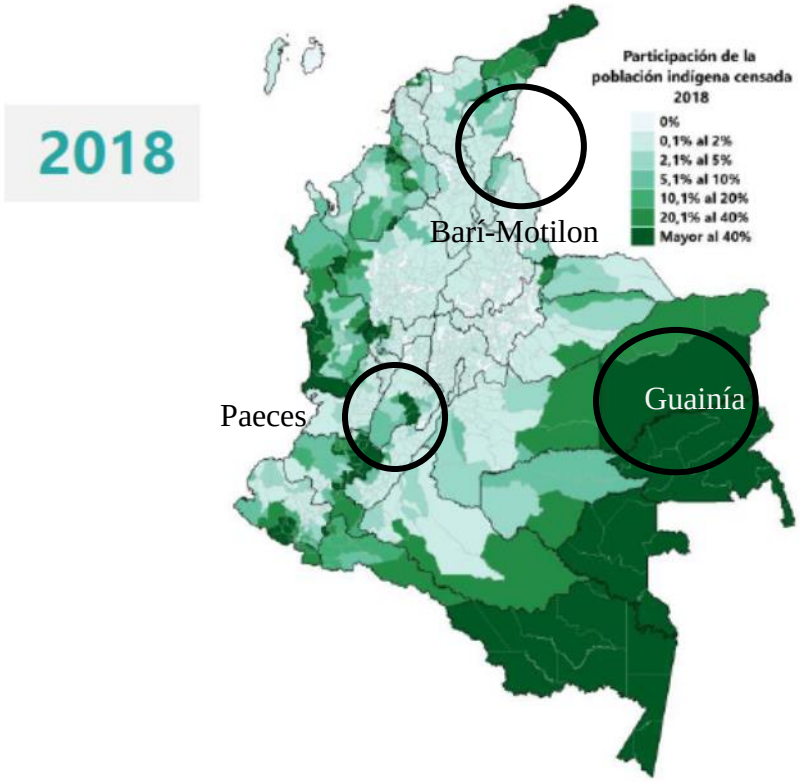


Figure 9: Map of Colombia's administrative regions with the relative percentage of indigenous individuals living in the distinct areas in 2018. The general area where live the three native groups, the Guainía, the Barí-Motilon and the Paeces, used in this study is also highlight. (Adapted from: https://elmorichal.com/esta-es-la-poblacion-indigena-de-Guainía-segun-el-censo-2018/)

The Departments that host most indigenous people are La Guajira, Cauca, and Nariño, where approximately 50% of the native population lives.

However, it is in Guainía that the indigenous population accounts to one of the highest proportions, around 75%, relatively to the total population of the department. Guainía, located in the eastern part of Colombia that borders Venezuela and Brazil, is one of the most isolated, scarcely populated and poorest regions of the country. For that, likely accounted the fact of being a region of the Colombian periphery that lies in a zone of the Amazon rainforest where the presence of rocky mountains hardly permits river navigation (Anzola and Luis 2004, Puerto, H., 2018). Most of the population of Guainía consists of indigenous people who resisted to the hostile environment, showing very low mixture with outsiders. Furthermore, there are many different groups scattered throughout the region, most of them speaking languages that belong to the same linguist group (Ossa, H., et al. 2015).

One of the native groups with which European sailors and settlers first contacted (Layrisse, Z., et al. 2001) were the Barí, who inhabited the northern part of Colombia and Venezuela, near the Lake Maracaibo (Quesada, J. D. 2004, Martinez-Laso, J., et al. 1995). They were so called because they spoke the Barí, a language of the linguistic family Chibcha also used by other natives who lived in the same area. The colonists called them “Motilones”, (from the Spanish word *motilar*— meaning to cut or crop), probably because traditionally men wore shaved head hair (Lizarralde, M., Lizarralde R., 2015). Identical custom was usual among neighbour people living around the Lake Maracaibo, explaining why the Spaniards, especially missionaries, also applied the term Motilones to refer to other tribes unrelated to the Barí, as the Yucpa, which didn't even shared the same language but instead were speakers of a Cariben language (Layrisse, Z., et al. 2001). Later on, the missionaries began to classify the Motilones as "Motilones mansos" to the Yucpa ethnic group and "Motilones bravos" to the Barí ethnic group, depending on the response of the indigenous people to the missionary presence and the level of pacification (Hernández, M.G. 2019). Since centuries the resistance to “pacification” maintained by Barí group was possible because they live in rather isolated communities, showing apparently low levels of admixture with non-Barí populations (Arango, R. and Sánchez, E., 2006, Lizarralde, M., Lizarralde R., 2015). In recent years, the number of individuals from this community has greatly increased due to better sanitation and health conditions (Lizarralde, M., Lizarralde R., 2015).

During the colonial period, many Native groups entered in conflict with the newly arrived colonists, striving for the preservation of resources and territories that the recent settlers wanted to expropriate from them forcedly. These conflicts drove many autochthonous groups

to near/total extinction (Gedicks 2003). Some of the most violent encounters involved the Paez and the Pijão, who lived in Center Colombia (Rappaport, J., 1990). They fought persistently against the invaders alongside other neighbouring natives, whether in colonial or post-colonial times (Zwisler 2018). Remarkably, they managed to survive, although persisting only in small communities that remain rather isolated in the Southern Andean region (Xavier, C. et al. 2015).

Despite the timid policies protecting the Colombian native populations during the colonial period under the rule of the Spanish monarchy and later during the governments established after the independence in the XIX century, indigenous individuals were reduced to a minority status without political protagonism, having been excluded from public and societal institutional representations. The unique contribution allowed to these groups in the construction of nowadays Colombia, was cheap labour enforced by a European elite. Their culture, social and religious practices were blatantly persecuted, promoting the assimilation of the canon beliefs and practices of the European elites that prevailed in the region (García, William et al.).

Only in the 1990's the Native Colombians were granted with full citizenship, becoming de facto citizens with rights identical to all other individuals with Colombian nationality (Escobar, C. 2015).

1.6.4. Genetic Diversity in Colombian Populations

As previously referred, the arrival of European settlers and African slaves to Colombia caused a massive rearrangement of the genetic diversity in the established pre-Columbian populations. Admixed populations arose throughout all country from interbreeding between Native and non-Native individuals, following patterns that varied widely from region to region. Consequently, each population is rooted in a-specific pattern of admixture, as is being revealed by distinct kinds of genetic markers (Salas, A., et al. 2008, Rojas, W., et al. 2010). In general, most of the Colombian Mestizo populations holds components of Native American and European descendant impregnated with a fraction of African ancestry that typically is considerably smaller (Salas, A., et al. 2008, Rojas, W., et al. 2010, Ibarra, A., et al. 2014). However, the proportions of Native American, European and African ancestries diverge greatly from population to population (Ossa, H., et al. 2016).

In urban areas, Native and European ancestries tend to be similar (reaching each ~40%), while the African one is much lower (usually ~10%). Taking into account administrative regions, admixed populations from the Caribbean and Southwest Coasts tend to exhibit larger African ancestry (between 20% and 30%) than in those from the Northeast, where often the Amerindian ancestry exceeds the European one, in opposition to the commonly observed in the Centre of the country (Rojas, W., et al. 2010, Ibarra, A., et al. 2014).

Importantly, the comparative study of uniparental markers revealed a distinct ancestry distribution in the mtDNA or Y chromosome lineages, adding valuable insights on the origin of women and men involved in the admixture processes, which as a rule were complex demographic events highly influenced by the social and historical context. It appears that most Mestizos resulted from admixture involving predominantly Native women and non-Native men, who more often were of European origin than of African one (Rojas, W., et al. 2010). Actually, such a sex bias underlying Colombian admixed populations is known to be a common pattern across Latin America (Wang, S., et al. 2008).

2. Aims

The population history of Colombia was very rich, complex and still today continues to be challenging. Yet, genetic information on Colombian populations is scarce, representing a limitation to understand the structure of the human genetic variation in the region.

In this context, the main aim of this work is to enlarge the knowledge on the paternal ancestry in distinct native groups and admixed populations from Colombia through the assessment of Y chromosome diversity, in order to obtain more insights into the main demographic events that shaped the current patterns of human genetic diversity.

Three native groups - the Guainía, the Barí-Motilion and the Paeces - and four admixture populations from the regions of Cundinamarca, Santander del Norte, Huila and Tolima will be considered.

To achieve the main goal, the following objectives were set:

- Characterize male lineages by genotyping STR as well as SNP markers present in Y chromosome;
- Evaluate the competence of some software that predict Y haplogroups based in Y haplotypes information;
- Collect data from other South American populations to obtain an extensive Y-chromosome genetic information to contextualize the obtained results;
- Contribute to a better understanding of the demographic processes that occurred in pre- and post- Columbus times that led to human colonization of South America.

3. Materials and Methods

3.1. Samples

A total of 121 unrelated men living in Colombia were considered in this work. From those, 51 were native Americans living in 3 different communities in Colombia (18 from Guainía, 21 belonging to Barí-Motilón community and 12 Paeces natives), and 70 were admixed individuals from 4 separated administrative regions (20 from Cundinamarca, 12 of Huila, 22 from Santander del Norte and 16 from Tolima). The present study is included in a project that further aims to characterize the same populations for Autosomal Ancestry Informative Markers (AIMs) and mitochondrial DNA. All participants have voluntarily given written informed consent to cooperate in this work, which was conducted under strict confidential conditions.

DNA extraction from blood samples had been already performed as described in Ossa, H. 2021, who dealt with samples collected into 4 mL flasks mixed with EDTA (Ethylenediaminetetraacetic Acid) to avoid blood coagulation DNA concentration of all samples was measured with a NanoDrop 1000 Spectrophotometer, and the results were considered to adapt the Y-STR and Y-SNP genotyping conditions.

3.2. Y-STR Genotyping

The commercial Yfiler™ Plus kit (Applied Biosystems™) was used to amplify simultaneously 27 Y-STRs (DYS576, DYS389I, DYS635, DYS389II, DYS627, DYS460, DYS458, DYS19, YGATAH4, DYS448, DYS391, DYS456, DYS390, DYS438, DYS392, DYS518, DYS570, DYS437, DYS385 a/b, DYS449, DYS393, DYS439, DYS481, DYS533, DYF387S1 and DYS533) following the manufacturer protocol. The PCR reaction comprised 2 μL of Reaction Mix (Applied Biosystems™), 1 μL of Primer Mix (Applied Biosystems™), 1 μL of DNA (± 0.5 μL depending of the NanoDrop results) and water to attain 5 μL of final volume. The samples were then submitted to an initial denaturation at 95ºC for 1 minute, followed by 30 cycles at 94ºC for 4 seconds and 61.5ºC for 1 minute, ending with a final extension step at 60ºC for 22 minutes.

After this, 1 μL of each PCR product was mixed with 9.6 μL of Formamide Hi-Di (Thermo Fisher™) and 0.4 μL of Gene Scan™ 600 LIZ™ Size Standard and then submitted to capillary electrophoresis in an ABI3500 sequencer (Applied Biosystems™). The results were analyzed with the GeneMapper® v.5.0 software and the alleles attributed in accordance with allelic ladder provided with Yfiler™ Plus kit (Applied Biosystems™).

Two different haplogroup prediction websites, the Haplogroup Predict (http://www.hprg.com/hapest5/) and the NevGen (https://www.nevgen.org/), were used to predict Y haplogroups based on haplotype data obtained.

3.3. Y-SNP Genotyping

52 Y-SNPs were studied using seven different PCR and SNaPshot Multiplexes already present in the Population Genetics and Evolution group from i3S (Table 3). The specific Multiplex system applied in each sample was chosen based on the haplogroup predictions by the tool NevGen (https://www.nevgen.org/).

Table 3: Multiplexes utilized and respective SNPs

Multiplex	SNPs
Multiplex 1 + M13 (Brion, M., et al. 2005, Gomes, V., et al. 2010)	P25, 92R7, SRY10831.1, SRY10831.2, M70, M173, Tat, M213, M9, M13
Multiplex 2 (Brion, M., et al. 2005)	M170, M62, M172, M26, M201, 12f2a
Multiplex E3 (Gomes, V., et al. 2010)	M96, M33, P2, M2, M191, M154, M35, M78, M81, M123, V6, M293, M85
Multiplex E1 (Campos, A. C., 2018)	M96, M33, P2, M2, U209, M154, U290, M191, U174, M35, M75, M85
Multiplex Q (Leite, G.A.D.F.P. 2018)	M242, P36.2, M346, M3, M19, Z19319, SA01, Z19483, SA05, M557
Multiplex R1 (Resque, R., et al. 2016)	M167, L23, M153, M529, U106, M207, U152, S116
Multiplex R2 (Resque, R., et al. 2016)	M167, M529, M207, M269, S116

The PCR reaction was prepared with 2.5 μL of QIAGEN® Multiplex PCR Master Mix, 0.5 μL Primer Mix (of the corresponding Multiplex, and each primer at 2 μM), 1 μL of DNA (with a ± 0.5 μL adjustment depending on results) and water to complete 5 μL of final volume. The PCR reaction began with an initial denaturation at 95ºC for 15 minutes, followed by 35 cycles at 94ºC for 30 seconds, 60ºC (except to the Multiplex Q where 62ºC were applied) for 90 seconds and 72ºC for 1 minute, ending with a final extension of 72ºC for 10 minutes.

To confirm the amplification success, the PCR products (also with negative control) were submitted to electrophoresis in a polyacrylamide gel. The results were visualized using the silver staining method as described in Budowle, B. 1991 with the exception of the time placing the gel in the ethanol solution (in our study was 10 minutes instead of 5) and in the nitric acid solution (changing to 5 minutes).

Once confirmed the presence of amplified target DNA fragments, a purification step was conducted through adding 0.5 μL of a mixture of two enzymes – Exonuclease I (25 μL) and FastAP (100 μL) (Thermo Scientific™) - to 1 μL of the PCR product. Those proportions were adjusted increasing up to 50% more of both components, when the PCR rate was weak, which was assessed through the results observed in the polyacrylamide gel. This step allows cleaning the amplified PCR product by hydrolysis of excess primers and nucleotides that could compromise the sequencing reaction. The conditions used were incubation at 37ºC for 30 minutes, followed by 15 minutes at 85ºC (to inactivate the enzyme).

Considering the single base extension (SBE) reaction, a mixture was prepared with 1.5 μL of the purified PCR product plus 1 μL of SBE primer mix (from specific multiplex), 1 μL of SNaPshot® Multiplex Kit (Applied Biosystems™). Water was added to complete the final volume to 5 μL. The mix was then submitted to 25 cycles (30 cycles for the Multiplex R1 and R2) of a denaturation cycle at 96ºC for 10 seconds, followed by an annealing at 50ºC for 5 seconds and ending with an extension at 60ºC for 30 seconds. Immediately after these reactions, the SBE products were purified by adding 1 μL of FastAP (Thermo Scientific™). The conditions of this purification were incubation during 60 minutes at 37ºC for the removal of the unwanted products, and finally a step of 15 minutes at 85ºC for the inactivation of the enzyme.

Afterwards, 0.63 μL of purified SBE products were added to 12 μL of mix of Formamide Hi-Di (Thermo Fisher™) with Liz120 Size Standard (Applied Biosystems™) (mix proportion: 9.45 μL of Formamide Hi-Di with 0.05 μL of Size Standard Liz120), and then submitted to capillary electrophoresis in an ABI3500 sequencer (Applied Biosystems™). Results were analysed with the GeneMapper® v.5.0 software.

3.4. Statistical Analysis

The haplogroup frequencies were counted directly. The Arlequin v3.5.2.2 software (Excoffier and Lischer, 2010) was used to calculate Y-STR and Y-SNP genetic diversities, and pairwise F_{ST} and R_{ST} genetic distances between our samples and other samples from South and Central America, Africa, Europe, Middle East and Southeast Asia. Given that not all studies had the same number of Y-STR and/or Y-SNP, a reduction towards a common number of markers was made. For pairwise R_{ST} genetic distances analysis null and duplicate alleles were removed, the repeat number of DYS389II was subtracted from the repeat number of DYS389I and the DYS385 a/b and DYF387S1 loci were removed.

The program IBM SPSS Statistics 26 was used to construct two-dimensional plots through Multidimensional Scaling (MDS) method, using both the pairwise R_{ST} and F_{ST} genetic distances obtained.

Phylogenetic networks were constructed with the NETWORK v.10.2.0.0 software using the Y-STR data, removing the DYS385 a/b and DYF387S1 loci. In order to present a less reticulated network, the reduced-median method (Bandelt et al. 1995) was applied before the median-joining method (Bandelt et al. 1999). Differential microsatellite weighting was applied in accordance with Qamar, R., et al. (2002) system: variance 0 - 0.09: weight 5; variance 0.1 - 0.19: weight 4; variance 0.2 - 0.49: weight 3; variance 0.5 - 0.99: weight of 2; variance $\geqslant$1.00: weight 1.

4. Results and Discussion

4.1. Y-STR Data

In the total of 121 samples of Colombian men belonging to 7 different groups, 108 different haplotypes were identified (Appendix Tables 12 – 19). None of these haplotypes was shared between groups, meaning that identical haplotypes were only detected within specific groups. All native groups presented a higher rate of haplotype sharing than the admixed groups. Among the latter, in Huila, Santander del Norte and Tolima all haplotypes were singleton, whereas in Cundinamarca the number of singletons was 18 in a total of 20 haplotypes. In the native groups, the Barí-Motilon showed the highest intra-group sharing (3 different haplotypes shared by 2 individuals, a distinct one shared by 4, in a total of 21 chromosomes), followed by the Guainía (2 haplotypes shared by two individuals, 3 shared by 3; in a total 18 chromosomes), and lastly the Paeces (only one haplotype shared out of 12).

The pattern of haplotype sharing in each population was naturally captured by the corresponding levels of Y-STRs genetic diversity, which were calculated excluding DYS385 a/b and DYF387S1 from the total set of markers analyzed (Table 4). All the native groups show smaller diversities than admixed populations, particularly the Barí-Motilon where it only reaches 0.9143 ± 0.0488. Diversity is maximum in all admixed populations, excepting Cundinamarca (0.9947 ± 0.0178).

Table 4: Genetic diversity for the different native and admixture populations using 23 Y-STRs.

	Group	N	Diversity
Native	Guaínia	18	0.9673 ± 0.0298
	Barí-Motilon	21	0.9143 ± 0.0488
	Paeces	12	0.9848 ± 0.0403
Admixed	Cundinamarca	20	0.9947 ± 0.0178
	Huila	12	1.0000 ± 0.0340
	Santander del Norte	22	1.0000 ± 0.0137
	Tolima	16	1.0000 ± 0.0221

High diversities in the 4 admixed populations here studied fits well the expectations for Colombian populations living in urban areas of country, since they have arose in the last few centuries with the waves of large-scale immigration that led to intense admixture, mainly between people of European, African and Native American ancestry. In addition, it is not surprising the reduced diversities in the 3 Native groups comparatively to the admixed ones.

The great majority of the indigenous populations from Colombia is concentrated in rural and sparsely populated zones; often living in a situation of great isolation. They are descendent of the people settled in Colombia before the European colonial occupation, and since then their history can be summarised as a struggle for autonomy and resistance to encroachment on culture and territory. Furthermore, admixture with non-indigenous people was quite restricted. All this sustains the predicted low levels of diversity in native groups compared with other Colombian populations. Even so, Y-STR diversity in the 3 native groups can be considered high in absolute terms; and do not show signs of founder or bottleneck effects, contrarily to what could be as well anticipated given the demographic past of these populations.

Anyway, levels of Y-STR genetic diversity here estimated in 3 Native Colombian groups are similar to those reported by Xavier, C., et al. 2015 for other Amerindian Colombian populations, as well are levels here found in 4 admixed Colombian populations compared to other admixed Colombian groups studied by Alonso Morales, L. A., et al. (2018).

Pairwise R_{ST} genetic distances between the 7 different populations examined in the present study were calculated using the Y-STRs contained in the Yfiler™ Plus kit excepting DYS385 a/b and DYF387S1, and the values obtained (Appendix Table 20) were plotted in a two-dimensional MDS (Figure 10).

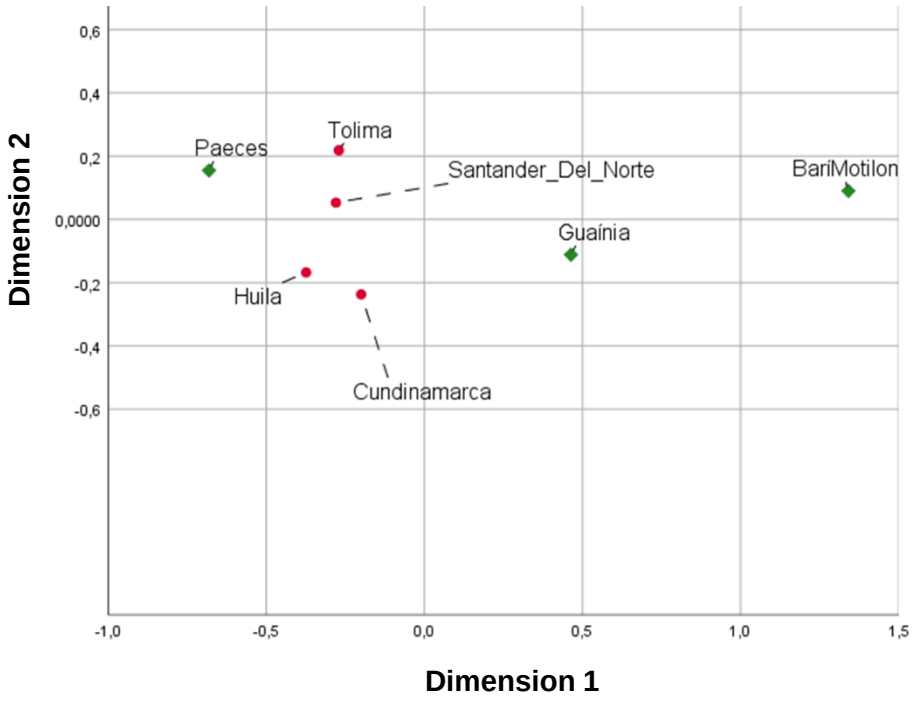


Figure 10: MDS plot based on R_{ST} genetic distances results between the 7 different populations studied. Green diamond represent native populations, whereby red dots are admixed groups (Stress=0.00189).

None of the pairwise genetic distances between the 4 admixed populations was statistically significant, which makes sense given that all were originated by similar recent admixture processes. The Paeces, the native group with the highest level of Y-STR diversity, also did not diverged from any of the admixed populations, while being significantly different from the other two natives (Guainía and Barí-Motilon). The Barí-Motilon stand out as the most differentiated population, showing significant differences from all other groups, either native or admixed.

4.2. Y-SNPs Data

Out of the 121 samples typed for Y-STRs, 98 were successfully characterized for Y-SNPs through the use of an hierarchical approach contemplating a total of 52 Y-SNPs. Y-haplogroup assignments assuming the Y-SNPs results were compared with those predicted by the Athey's Haplogroup Predictor v5 (Athey, T. W., 2006) and the NevGen (Cetkovic Gentula, M., and A. Nevski 2015) softwares, which are both based in Y-STR profiles. Since all the 98 samples typed for SNPs were allocated in the same haplogroups predicted by NevGen, whereas identical consistency was not observed with Athey's Haplogroup Predictor v5, the remaining 23 samples were assumed to belong to the haplogroup class assigned by NevGen (Appendix Tables 12-19, 22).

A total of 22 different Y-SNP defined haplogroups were identified (Figures 11 and 12). Y-haplogroup diversity varies in different populations, attaining the lower value in the native Barí-Motilon (0.0000) and the higher in the admixed population from Cundinamarca (0.9474). Unlike values assessed through Y-STRs, the native Paeces have an Y-haplogroup based diversity similar to that in the admixed populations from Santander del Norte and Tolima (0.8636, 0.8788 and 0.8000 respectively), being even higher than the one found in the admixed from Huila (0.75758) (Table 5).

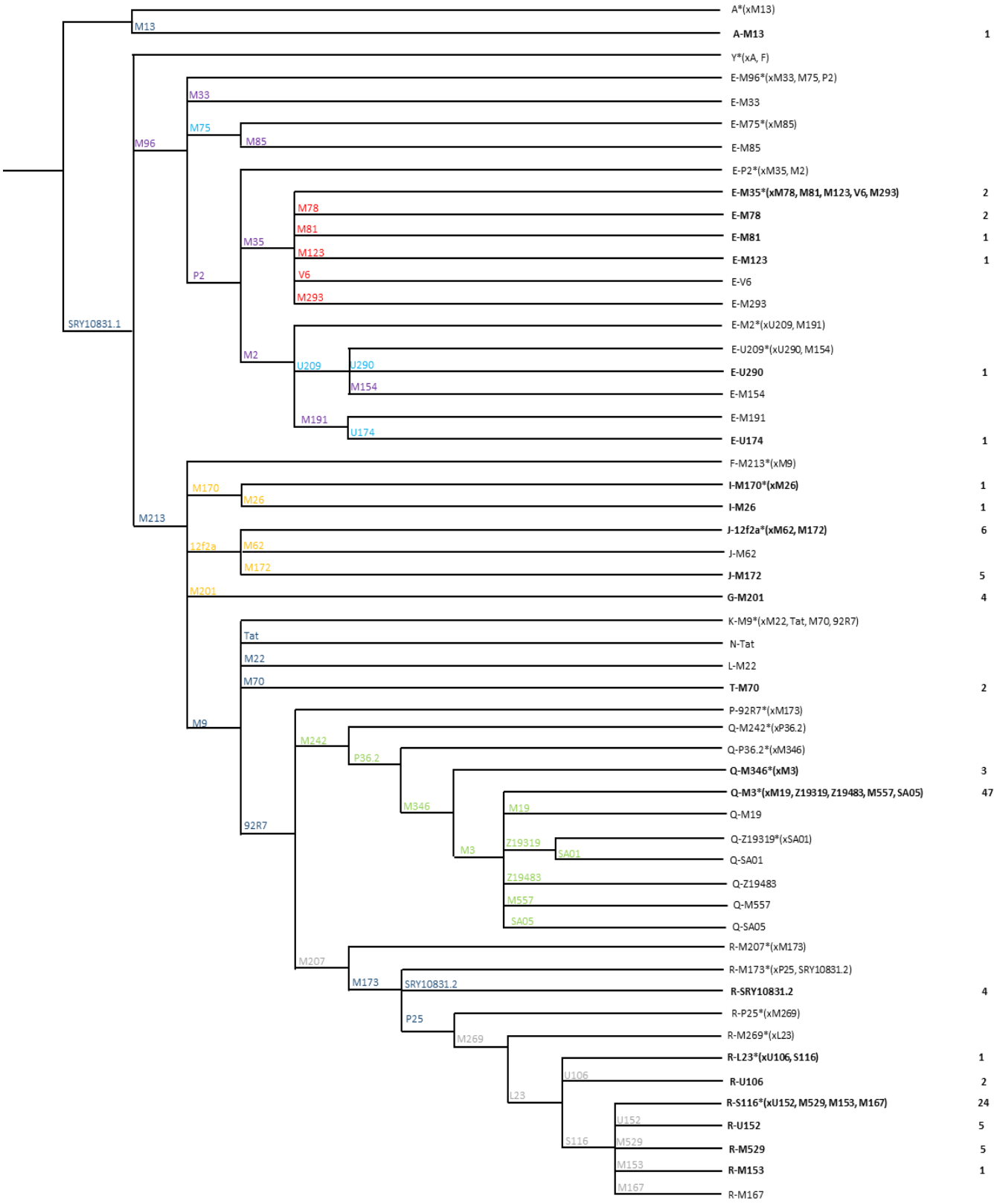


Figure 11: Phylogenetic tree with the Y chromosome haplogroups studied through the 52 Y-SNPs genotyping. The haplogroups identified in this study are in bold as well are the specific number of occurrences in a total of 121 samples. The colour scheme used in the SNP identifiers highlights the specific Multiplexes applied: Dark blue: Multiplex 1+M13; Yellow: Multiplex 2; Red: Multiplex E3; Light blue: Multiplex E1; Purple: SNPs shared by the Multiplexes E1 and E3; Grey: Multiplex R; Green: Multiplex Q.

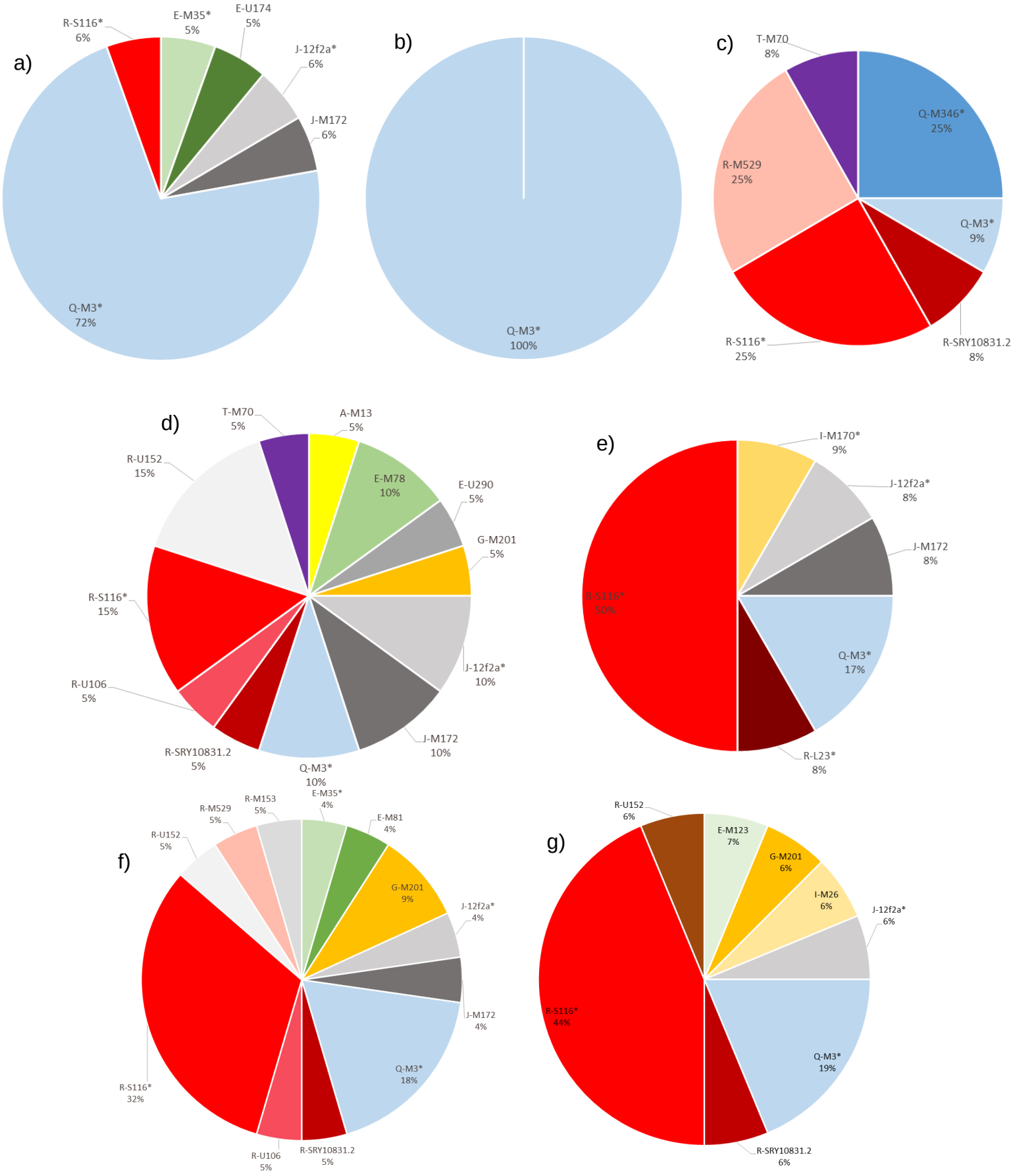


Figure 12: Pie charts displaying the distribution of Y-haplogroup frequencies in Colombian populations: a) Guainía; b) Barí-Motilon; c) Paeces; d) Cundinamarca; e) Huila;f) Santander Del Norte; g) Tolima.

Table 5: Genetic diversity found in the different native and admixture population's haplogroups.

	Group	N	Diversity
Native	Guaínia	18	0.4902 ± 0.1424
	Barí-Motilon	21	0.0000 ± 0.0000
	Paeces	12	0.8636 ± 0.0639
Admixed	Cundinamarca	20	0.9474 ± 0.0275
	Huila	12	0.7576 ± 0.1221
	Santander del Norte	22	0.8788 ± 0.0546
	Tolima	16	0.8000 ± 0.0916

Using the Y-SNPs results, the F_{ST} pairwise genetic distances between populations were calculated (Appendix Table 21) and then plotted in a two-dimensional MDS (Figure 13).

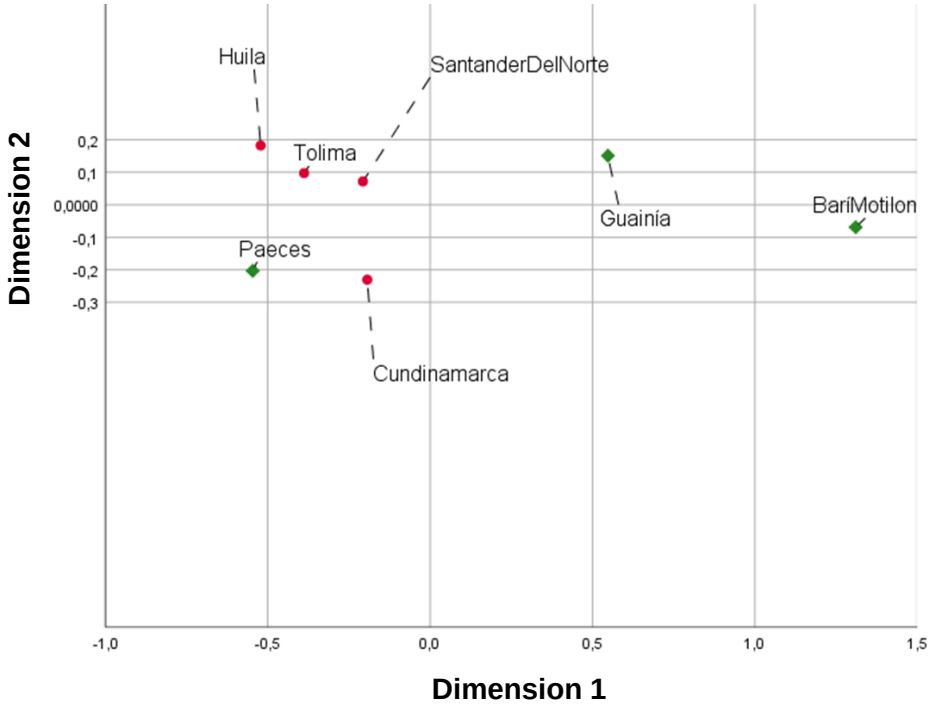


Figure 13: MDS plot of pairwise F_{ST} genetic distances based on 22 Y-haplogroups found in our Colombian populations. Green diamonds and red dots represent native and admixed populations, respectively (Stress = 0.00376)

This MDS is highly consistent with the previous one: the Barí-Motilon are again the group showing the sharper differentiation in the set of Colombian populations and, besides, the Paeces are the closest natives to the admixed Colombian populations. In this case, the relative positioning of the Barí-Motilon can be explained by the homogeneity of their male lineages, all falling in haplogroup Q, contrasting with the much more diverse repertoires of the other Colombian populations. The second highest frequency of Q lineages was detained by the Guainía, and hence their proximity with the Barí-Motilon. In the Paeces, non-Native lineages dominate over the Native ones, just as it happens in the admixed populations, a feature that likely accounts for the position of the Paeces in the plot, revealing more affinities with admixed populations than with other native groups.

In Figure 14 is displayed a network encompassing the entire set of samples that was constructed with the STR data. Overall, a reasonable agreement exists between the pattern of phylogenetic relationships between Y-STR haplotypes and the assigned haplogroups through Y-SNP typing.

The Y-haplogroups more well represented in the whole set of samples, were haplogroups Q-M346 and R-L23. However, as can be seen in the network, the level of molecular heterogeneity among Q-M346 lineages is considerably higher than among those falling in R-L23. Apart from these, the remaining haplogroups appeared at much lower or very rare frequencies. Of the detected haplogroups, only the Q is associated to a Native-American ancestry, as will be briefly described just below.

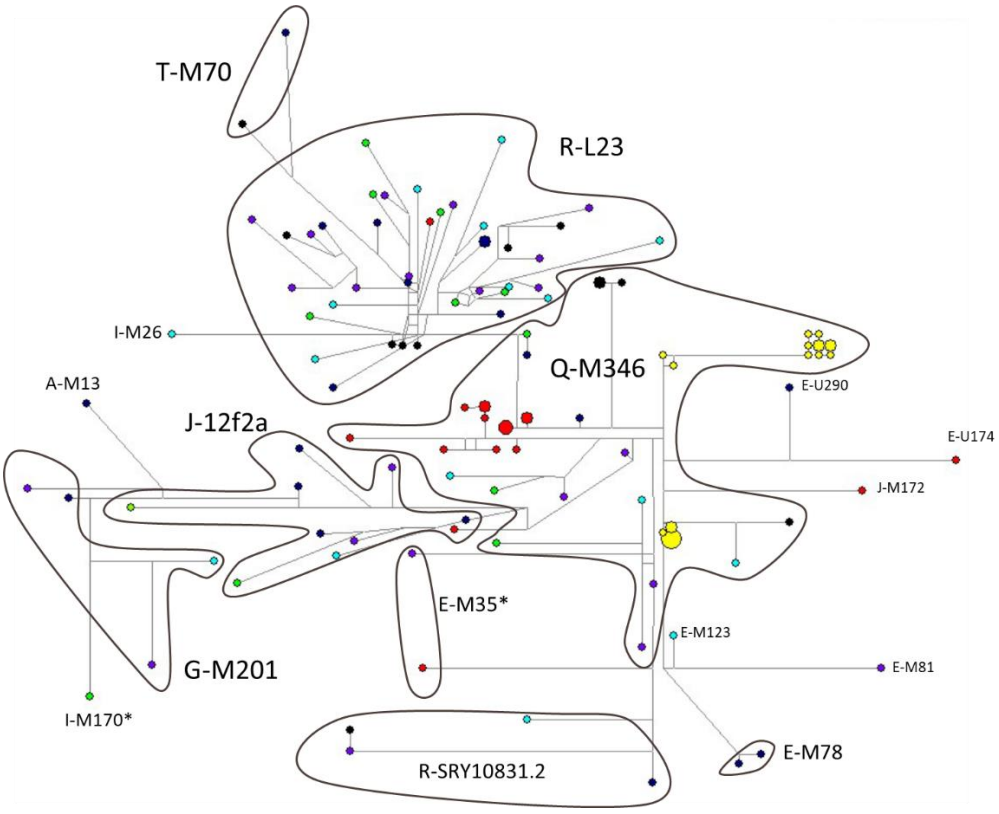


Figure 14: Network constructed with the 121 samples using 23 Y-STRs. The corresponding haplogroup is defined by borders, with the colour scheme representing: Red – Guainía, Yellow – Barí-Motilon, Black – Paeces, Dark blue – Cundinamarca, Green – Huila, Purple – Santander del Norte, Light blue – Tolima

4.2.1. Haplogroup A

Haplogroup A is one of the deep branches of the human Y chromosome phylogenetic tree. Usually, it is found in African populations occurring with considerable prevalence in populations from Eastern Africa, Northern Cameroon, for instance (Hassan, H.Y., et al. 2008, Scozzari, Rosaria, et al. 2012). Noteworthy, this lineage also was already found in Southern Europe and the Middle East, but at much lower frequencies and often interpreted as the result of recent migration from Africa (Karafet, T.M., et al. 2008). In this study, it was detected only once in Cundinamarca, representing 5.00% of that sample, and a mere 0.83% of the whole dataset. Given the pattern of distribution of the haplogroup, the unique sample bearing A-M13 was assumed to have an African origin.

4.2.2. Haplogroup E

The macrohaplogroup E is a major group of the Y-lineages that is present in different continents. It contains subclades that are particularly frequent in North, East and Sub-Saharan Africa, where the most ancestral lineages are found, however, also encompasses other subclades less common that are spread across Europe, Middle East and Central and South Asia (Karafet et al. 2008, Cruciani, Fulvio, et al. 2004).

These lineages were found in 6.61% of the present whole Y-chromosome dataset. E-M35*(xM78, M81, M123, V6, M293) and E-M78 were the most well represented sub-haplogroups (both 1.65%), but E-M81, E-M123, E-U174 and E-U290 were also detected (each with 0.83%). The E lineages were found in the Guainía (11.11%), in Cundinamarca (15.00%), in Santander del Norte (9.10%) and in Tolima (6.25%).

The E-M35 subclade is distributed between Africa, West Asia and Europe (Trombetta, Beniamino, et al. 2015, Cruciani, Fulvio, et al. 2004). Within E-M35, the E-M81 and E-M123 sub-lineages are commonly found in North Africa and more rarely in Europe and Middle East (Alvarez, Luis, et al. 2009, Cruciani, Fulvio, et al. 2004, Lancaster, A. 2009), while E-M78 is mainly found in Eastern and Northeastern Africa (Horn of Africa), appearing less in Southern Europe and Middle East (D'Atanasio, Eugenia, et al. 2018, Lancaster, A. 2009). The E-M35*(xM78, M81, M123, V6, M293) is an ancestral branch to the previously described, and like E-M78, its prevalence peaks in Eastern and Southern Africa, while being very low in South Europe (Cruciani, Fulvio, et al. 2004). Two E-M35*(xM78, M81, M123, V6, M293) chromosomes were found, one in the Guainía (5.56% of males) and another in Santander del Norte (4.54% of males); E-M81 had a unique occurrence in Santander del Norte (4.54%); E-M123 was also detected only once in Tolima (6.25%); and the E-M78 was bore by two men in Cundinamarca (10.00%).

Because E-M35*(xM78, M81, M123, V6, M293) and E-M78 are both present either Sub-Saharan Africa or Southern Europe, in order to clarify the likely origin of the lineages found in the Colombian samples, two networks were constructed using 10 Y-STRs haplotype information, one for E-M78 (Figure 15) and the other for E-M35* (Figure 16). The E-M78 network integrates samples from Sub-Saharan Africa (Berniell-Lee, Gemma, et al. 2009, Carvalho, M., et al. 2011, Gomes, V., et al. 2010, González, M., et al. 2013, Rosa, Alexandra, et al. 2007, Tofanelli, S., et al. 2009) and Europe (Adams, Susan M., 2008, Boattini, Alessio, et al. 2013, Nogueiro, I., et al. 2010, Rey-González D., et al. 2017, Solé-Morata N., et al.

2015), and that regarding E-M35* samples from Eurasia (Adams, Susan M., 2008, Haber, M., et al. 2011) and Sub-Saharan Africa (de Filippo, Cesare, et al. 2011).

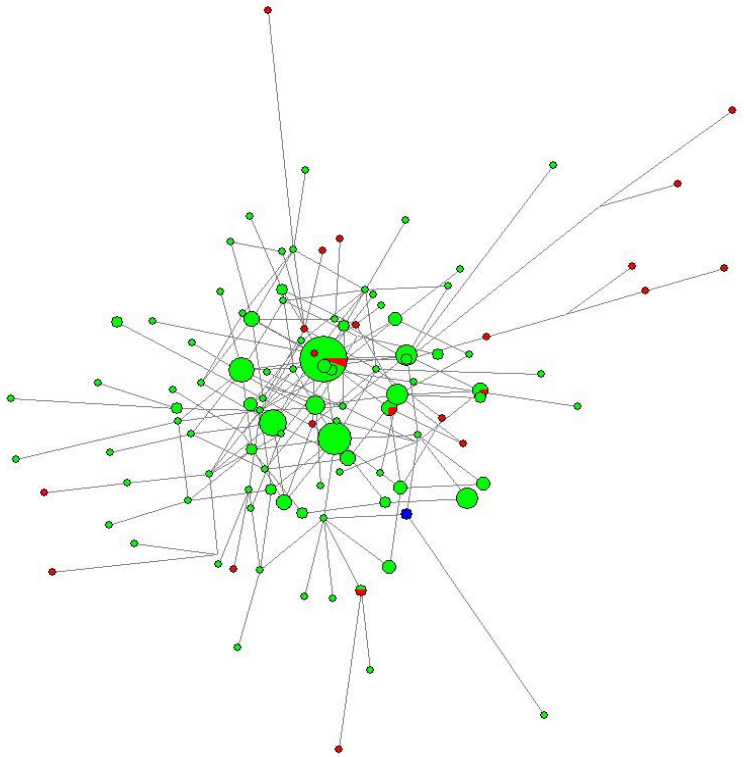


Figure 15: Network constructed with 224 South European (Spain, Portugal and Italy), 17 sub-Saharan Africa (Gabon, Guinea-Bissau, Equatorial Guinea, Madagascar and Uganda) and 2 Colombian samples from the Cundinamarca sub-set all samples belonging to the sub-haplogroup E-M78. Green samples are European, Red individuals are from sub-Saharan Africa and Blue are from Colombia.

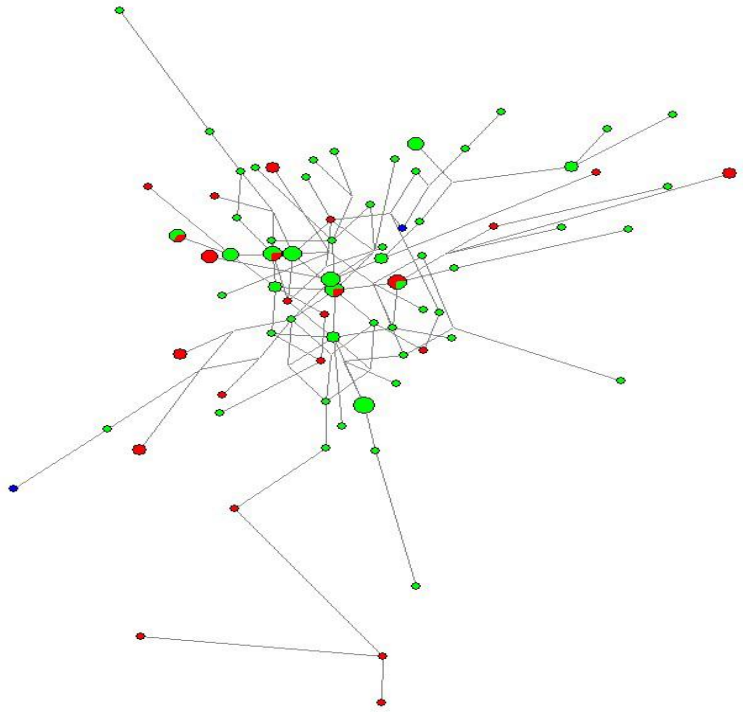


Figure 16: Haplogroup E-M35* phylogenetic network constructed with 77 Eurasian (Spain, Portugal, Iran and Lebanon), 32 sub-Saharan Africa (Gabon, Guinea-Bissau, Equatorial Guinea, Madagascar and Uganda) and 2 Colombian samples from the Guainía and Santander del Norte sub-sets. Green samples are Eurasian, Red individuals are from sub-Saharan Africa and Blue are from Colombia.

In both networks, it was not possible to discriminate clusters of European and Sub-Sahara African haplotypes. Furthermore, no sharing of Colombian haplotypes with those from others regions was observed. However, in both cases the Colombian haplotypes were closer to European/Eurasian than to sub-Saharan African haplotypes. Given this, together with the fact that in sub-Saharan Africa the E lineages are especially common around the Horn of Africa (eastern Africa), while not in Western Africa, which was the main source of African people during the Atlantic slave trade, rendered unlikely that the Colombian E samples have had a recent sub-Saharan origin. For these reasons, the samples were considered to be part of the repertoire of lineages introduced in Colombian by the Europeans.

Unlike E-M78 and E-M35*, haplogroups E-U174 and E-U290 are highly specific from Central and West Africa (Rosa, Alexandra, et al. 2007, Ansari Pour, Naser, et al. 2013). Beyond Africa, they only have been reported in Afro-American populations. Thus, the instances now identified in Colombian men were assumed to have Sub-Saharan African ancestry. A chromosome E-U174 was found once in Guainía (5.56%), while the E-U290 also appeared once in Cundinamarca (5.00%).

4.2.3. Haplogroups G, I and J

The haplogroup G, defined by M201, occurs mainly in the Middle East, Caucasus and Southern Europe (Rootsi, Siiri, et al. 2012). In present Columbian sample, it was found in 3.31%, and its distribution according to specific groups was: Cundinamarca, 5.00%; Tolima, 6.25%; and Santander del Norte, 9.10%.

Relatively to the haplogroup I, it is frequent in Europe, manly in the Balkan region, Eastern and Northern Europe (Karafet et al. 2008), having been rarely described outside this continent. In this study, this haplogroup was represented by I-M170*(xM26) lineages found once in Huila (8.33%), and by I-M26 detected in another sample from Tolima (6.25%). In the total of 121 samples the haplogroup accounted mere 1.65%.

The J clade has a distribution that spreads from Southern Europe, Northern Africa, Middle East, Levant, Central Asia, Pakistan and India (Di Giacomo, et al. 2004, Karafet et al. 2008). In the Colombian populations under study, 9.10% of all lineages belonged to haplogroup J, apportioned in two different subclades: J-12f2a*(xM62, M172) (4.96%), and J-M172 (4.13%). The first was present in Guainía, Huila, Santander del Norte, Tolima and Cundinamarca with relative frequencies of 5.56%, 8.33%, 4.54%, 6.25% and 10.00%, respectively. The second was detected in Guainía, Huila, Santander del Norte and Cundinamarca, representing 5.56%, 8.33%, 4.54% and 10.00%, respectively.

J-12f2a*(xM62, M172) is typically more common in North Africa, Middle East and Caucasus than in Southern Europe (Sahakyan, Hovhannes, et al. 2021, Di Giacomo, et al. 2004, Semino, Ornella, et al. 2004), while J-M172 tend to be more prevalent in the Levant and Southern Europe, than in Northern Africa and Southern Middle East (Sahakyan, Hovhannes, et al. 2021, Semino, Ornella, et al. 2004).

Considering the above distributions, and also the absence of significant direct migration influx into Colombia from regions where some of these haplogroups reach a peak in frequency (i,e, North Africa, Middle East, Caucasus, for instance, Weinstein, Asher, 2021), the introduction of G, I and J lineages in Colombia was assumed to be mainly mediated by the European men, who begun to enter the country in the early 16th century.

4.2.4. Haplogroup Q

The Q haplogroup, defined by the M242 mutation, and it's subclades are found with higher frequencies in Native American and Siberian populations (Dulik, M. C., et al. 2012, Zegura, S. L., et al. 2004, Singh Malhi, R., et al. 2008, Grugni, V., et al. 2019), as well as other that are present at lower frequencies in populations from Eastern Asia (Battaglia, V., et al. 2013, Kim, Soon-Hee, et al. 2011).

Of all the haplogroups identified in this study, the haplogroup Q was the one had the highest frequency (41.32%). Within this haplogroup, two subclades were present: Q-M346*(xM3) summing up only 2.48% of lineages, and Q-M3*(xM19, Z19319, Z19483, M557, SA05), which attained the substantial value of 38.84%.

The distribution of the Q-M346*(xM3) is globally identical to that described for the macrohaplogroup Q in which it is anchored (Malyarchuk, B., et al. 2011, Bailliet, G., et al. 2009, Dulik, M. C., et al. 2012). Within the Q-M346, a derived branch exists defined by M3 marker. The subclade Q-M3 is considered to be one of the main Native American founding sub-lineages, being practically American-specific (Malyarchuk, Boris, et al. 2011, Bortolini, M.C., et al. 2003, Grugni, V., et al., 2019). Throughout America, there are many lineages inside of Q-M3 haplogroup. In present study we investigated lineages defined by M19, Z19319, Z19483, M557 and SA05 Y-SNPs.

Nevertheless, none of them was identified in the Colombian Q-M3 samples, which were then classified as Q-M3*(xM19, Z19319, Z19483, M557, SA05). This haplogroup was found in all the populations analysed, with the following frequencies: 72.22% in Guainía, 100% in Barí-Motilon, 8.33% in Paeces, 10.00% in Cundinamarca, 17.00% in Huila, 16.67% in Tolima and 18.75% in Santander del Norte. Contrarily, Q-M346*(xM3) was only present in Paeces (25.00%).

Given the wide distribution of the haplogroup Q throughout the America, a network was constructed (Figure 17) using haplotype data based on 15 Y-STRs (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y GATA H4) of the Q lineages from different Native-American populations. Besides our samples, the network was constructed with samples from Argentina (Toscanini, U., et al. 2008), French Guyana (Mazières, et al. 2009), Bolivia (M. Gaya, Vidal et al. 2011), Ecuador (Geppert, et al. 2011), Colombia (Xavier, C., et al. 2015), Brazil (Roewer, L. et al. 2013), Peru (Roewer, L., et al. 2013, Leite, Gonçalo Aragão da Fonseca Pinto 2018), Mexico (Sandoval, et al. 2012), Venezuela (Roewer, L., et al. 2013) and Panama (Battaglia et al. 2013).

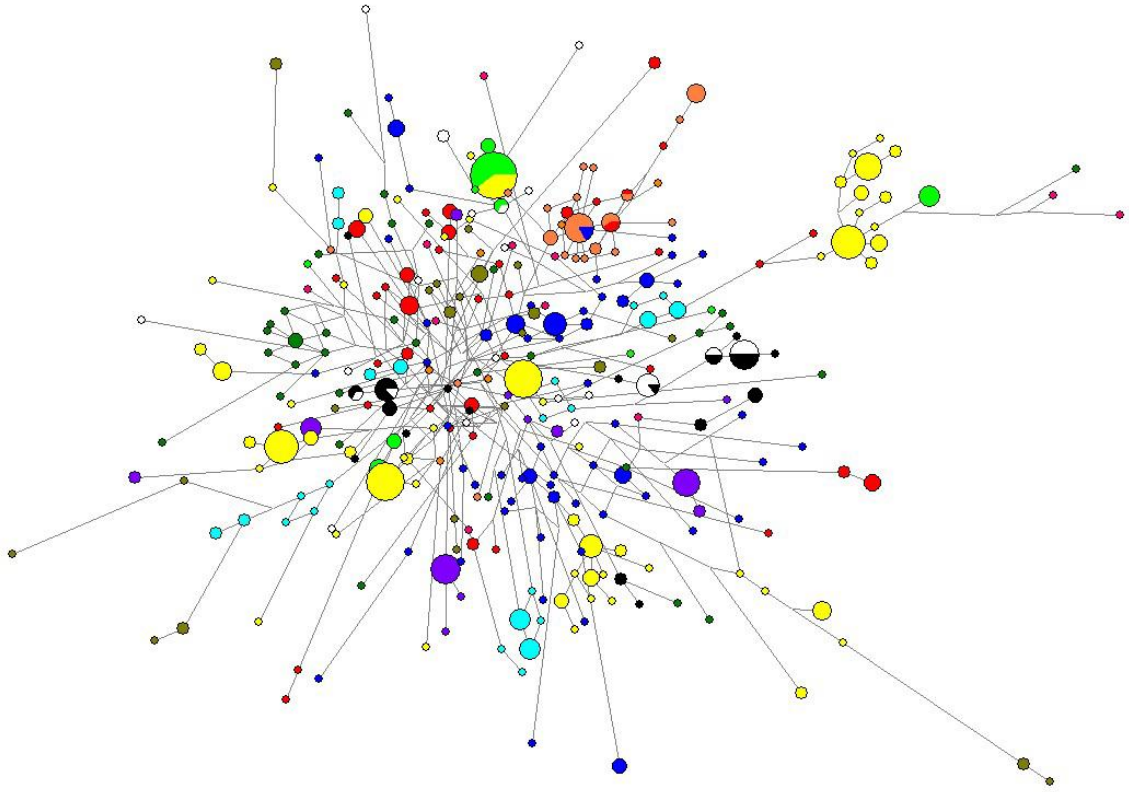


Figure 17: Network of samples belonging to the Q haplogroup based on 15 Y-STRs. Yellow - Brazil; Light green - French Guyana; Purple - Ecuador; Red: Argentina; Dark Blue – Peru; Light Blue – Colombia; Brown – Mexico; Dark Green – Panama; White – Venezuela; Orange – Bolivia; Black: Colombian Native groups (Guainía, Barí-Motilon and Paeces); Dark Pink – Colombian admixture populations (Cundinamarca, Huila, Santander del Norte and Tolima).

The heterogeneity of Q lineages is clearly reflected in the present network, which also reveals that haplotype sharing is quite reduced at the inter-population level, though common within specific populations. The Colombian samples of the present study are scattered through the network. Is also observe cases of haplotype sharing between Colombian and non-Colombian samples. Remarkably, they involved restrictively haplotypes present in the Barí-Motilon and in Venezuelan populations, which can be is understood given their proximity.

4.2.5. Haplogroup R

Haplogroup R is carried by >50% European men, likewise being rather frequent and is very wide spread across Europe, Asia and Africa (Kayser, Manfred, et al. 2003, Myres, Natalie M., et al. 2011, Underhill, Peter A., et al. 2015). It also contains a few members found in some parts of the Sahel region of Africa, which assumedly were brought to there as a consequence of early Holocene Eurasian backflow into Africa (Haber, M. et al, 2016).

In this study, the haplogroup R was the second most frequent macrohaplogroup, encompassing 35.54% of all male lineages, allocated into two main sub-clades, R1a and R1b.

Here R1a was identified by the presence of SRY10831.2, and represented 3.31% of the dataset. It was found in the Paeces, Cundinamarca, Santander del Norte and Tolima populations, with frequencies of 8.33%, 5.00%, 4.54% and 6.25% respectively.

R1b was ascertained by the presence of M269, and within it six downstream haplogroups were identified: R-L23*(xU106, S116), R-U106, R-S116*(xU152, M529, M153, M167), R-U152, R-M529 and R-M153, with frequencies of 0.83%, 1.65%, 20.66%, 4.13%, 4.13% and 0.83%, respectively, in the total dataset. R-L23*(xU106, S116), the precursor of the remaining 5 members, is mainly found in Europe, Anatolia and Caucasus (Herrera, Kristian J., et al. 2012, Underhill, Peter A., et al. 2015). Here, it was only detected in a man from Huila, accounting to 8.33% of the population. R-U106, which is quite common in Central Europe with frequency diminishing towards East Europe and Turkey (Myres, Natalie M., 2007), was identified once in Cundinamarca and also once in Santander del Norte (respectively at 5.00% and 4.54% relative frequencies).

R-S116, a branch parallel to R-U106, is mainly prevalent in Western Europe, especially in the Iberian Peninsula (Myres, Natalie M., 2007). The R-S116*(xU152, M529, M153, M167) sub-haplogroup was found with the second highest frequency here estimated: 20.66% of the Colombian men. It was present in all the populations analysed, although with variable frequencies: Guainía, 5.56%; Paeces, 25.00%; Cundinamarca, 15.00%; Huila, 50%; Santander del Norte, 31.82% and Tolima, 43.75%.

R-U152 is descendant of R-S116, mainly widespread in Western and Southern Europe (Cruciani, Fulvio, et al. 2011, Myres, Natalie M., 2007). In the Colombian populations, it was found in Santander del Norte (4.54%), Tolima (6.25%) and Cundinamarca (15.00%).

Also descendant of the R-S116 is R-M529, which has been primarily found in the British Isles (Myres, Natalie M., 2007). In this work, it was detected in the Paeces (25.00%) and in Santander del Norte (4.54%). Lastly, R-M153 lineage that has been almost exclusively detected in the Iberian Peninsula, usually at low frequencies except in the Basques (López-Parra, A. M., et al. 2009), was identified in a only man from Santander del Norte (5.00% internal frequency).

Taking into account the geographical distribution of the lineages belonging to the R haplogroup, and given their relative prevalence in Europe, all samples here identified as R were considered to be of European ancestrally.

4.2.6. Haplogroup T

Haplogroup T-M70 was identified in the present study in 1.65% of the entire dataset, having appeared only in the Paeces and in Cundinamarca (intra-population frequency of 8.33% and 5.00%, respectively). This haplogroup is distributed throughout Europe and the Middle East, and occasionally is detected in North Africa (Mendez, Fernando L., et al. 2011). Therefore, the presence of T lineages in our both groups was also attributed to admixture with Europeans.

4.2.7. Colombian Ancestry Origins

Mainly based on the Y-haplogroup results previously described, it was possible to infer the likely ancestry of the lineages detected in our 7 Colombian populations.

Haplogroups E (except E-U290 and E-U174), G, I, J, R and T found in our sample were assumed to be of Eurasian origin; haplogroup Q was considered to be of Native American ancestry; and haplogroup A plus the sub-clades E-U290 and E-U174 as being of African origin.

The different Colombian populations presented an uneven distribution of the 3 ancestry sources. Expectedly, however, while the Eurasian component prevailed in the admixed populations, the Native American component attained the highest proportions in the Native groups. The African ancestry was the minority, only detected in the Native Paeces and in the admixed population from Cundinamarca (Figure 18).

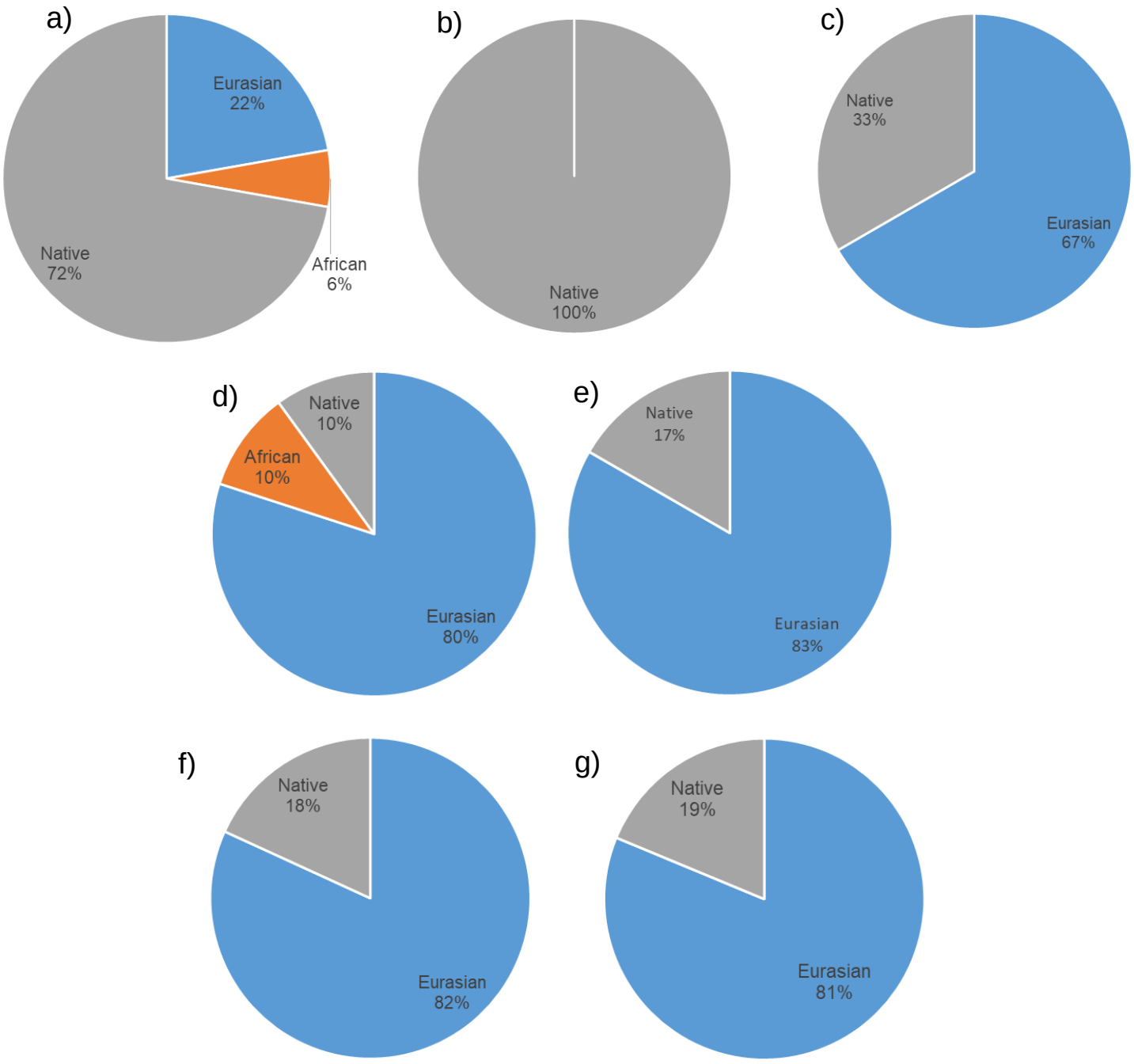


Figure 18: Inferred ancestries in seven Colombian populations. a) Guainía; b) Barí-Motilon; c) Paeces; d) Cundinamarca; e) Huila; f) Santander del Norte; g) Tolima

The pie charts in Figure 18 demonstrate the clear distinction between admixed and Native Colombian groups in terms of lineages' ancestry. In the admixed populations the Eurasian component is very similar, ranging between 80% and 83%, being complemented with an ancestry source that is uniquely Native American, except in Cundinamarca where an African ancestry is also present.

Unlike the admixed populations, in the native groups the ancestry components differ markedly between each other: In the Barí-Motilon all male lineages are of Native-American ancestry. In Guainía and Paeces the ancestry sources are much more heterogeneous: Guainía shows and ancestry pattern almost inverse to that of the admixed populations, especially from Cundinamarca: the native ancestry dominates over the Eurasian source, but also contains a small component of African origin. The Paeces are the natives with ancestry sources more similar to those of the Colombian admixed populations: the amount of Eurasian lineages is the highest comparatively to the other native groups, whereas in opposition the amount of Native-American lineages is the lowest.

4.3. Population Comparisons

The Y-STR and Y-SNP data obtained in the 7 Colombian populations studied was used to perform comparisons with other American populations whose data was retrieved from the literature.

4.3.1. Between American Native groups

Pairwise R_{ST} genetic distances were calculated based on 15 Y-STRs (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y GATA H4), which were those STRs shared in the batteries applied to characterize the Native American populations enrolled in the comparison that included, besides the Colombian groups addressed in this work, 28 native groups discriminated in Table 6 namely from Argentina (Toscanini, U., et al. 2008), French Guyana (Mazières, S., et al. 2009), Bolivia (Gayà-Vidal, M., et al. 2011), Panama (Grugni, V., et al. 2015, Battaglia, V., et al. 2013), Venezuela (Roewer, L., et al. 2013), Mexico (Sandoval, K., et al. 2012), Peru (Roewer, L., et al. 2013, Leite, G.A.D.F.P. 2018), Ecuador (Geppert, M., et al. 2011), Brazil (Roewer, L., et al. 2013) and Colombia (Roewer, L., et al. 2013, Xavier, C., et al. 2015). In addition, a population from Europe and another from Africa were also used as outgroup references (Table 6).

Table 6: Description of the Native, European and African populations used in the Y-STRs comparisons with the Guainía, Barí-Montilon and Paeces groups.

Reference	Country	Population	Code	N
Toscanini, U. et al. 2008	Argentina	Colla	COL	29
		Toba	TOBA	49
Mazières et al. 2009	French Guiana	Wayampi	WAY	29
		Palikur	PAL	28
		Emerillon	EM	13
	Brazil	Apalai	AP	27
M. Gaya, Vidal et al. 2011	Bolivia	Quechuas	QU	55
		Aymaras	AY	59
Geppert et al. 2011	Equador	Kichwa	KI	15
		Wao	WAO	40
Roewer et al. 2013	Brazil	Kayapó-Xikrin	KX	10
		Zoé	Z	25
		Urubu-Kaapor	UK	27
	Venezuela	Barí-Boxi	BB	16
		Wayuu	WAYUU	17
	Peru	Chuquibamba	CHU	16
		Huanca	HU	13
		Shipibo-Conibo	SC	21
	Colombia	Coconuco	COC	7
		Embera-Chamí	EC	24
		Guambiano	GUAM	16
Grugni et al. 2015	Panamá	Kuna Yala	KY	16
Gonçalo Aragão 2018	Peru	Asháninka	ASH	53
Sandoval et al. 2012	Mexico	Maya	MAYA	14
		Nahuas	NA	56
		Pima	PIMA	49
Xavier, C. 2015	Colombia	Antiquoa	ANT	23
		Cauca	CAU	48
Battaglia et al. 2013	Panama	Native Panamarian	PA	36
Purps et al. 2014	Europe	Spain	SP	499
	Africa	Yoruba	YO	77

Most of the R_{ST} distances between the three native Colombian groups and other native-American populations were statistically significant (P<0.05, Table 23 Appendix). The exceptions were i) the Paeces contrasted with the Guambiano (R_{ST}= 0.04532; P=0.06118) or the Coconuco (R_{ST}=0.05899; P=0.17503), both from Colombia, and with the Cola from Argentina (R_{ST}=0.05417, P=0.07742); ii) the Guainía relatively to the Kichwa from Equator (R_{ST}=0.05022, P=0.07524); and the Barí-Motilon compared to the Barí-Boxi from Venezuela (R_{ST}=0.09065, P=0.0690).

The R_{ST} pairwise genetic distances were used to obtain the two-dimensional MDS plot displayed in Figure 19. The majority of the Amerindian populations integrates a large cluster in the center of the plot, which is surrounded by a number of other native groups highly differentiated from each other. Both the Paeces and Guainía are positioned in the central cluster, whereas the Barí-Motilon stand more peripherally, approaching other well differentiated native groups like the Barí-Boxi (Venezuela) and the Toba (Argentina), which harbour a very elevated Native ancestry (~90%).

The Guainía and especially the Paeces are quite close to the reference European population used in the analysis, reflecting the substantial European ancestry in their male genepools.

The reference African population appears in the plot as an outlier, indicating that the African ancestry is in general a minor contributor for the male diversity of Native-American groups.

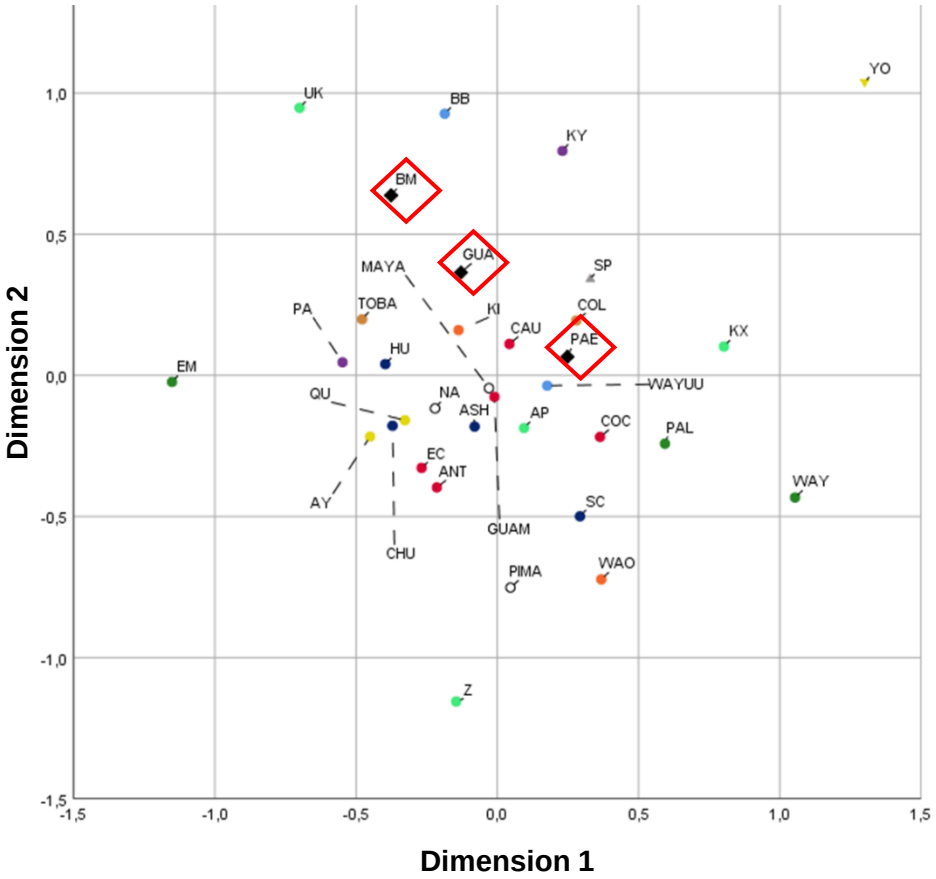


Figure 19: MDS plot based on the R_{ST} genetic distances calculated using the information of 15 Y-STR, between the 33 populations analysed (Stress=0.05810). Populations were grouped by country: Colombian (Red), Argentina (Brown), Ecuador (Orange), French Guiana (Dark Green), Bolivia (Yellow), Panama (Purple), Venezuela (Blue), Mexico (White), Peru (Dark Blue), Brazil (Light Green). The Guainía (GUA), Barí-Montilon (BM) and Paeces (PAE) are marked in black diamonds and highlighted in a red outline, the European group in a brown triangle and the African in a yellow triangle.

Using the Y-SNP data, the pairwise F_{ST} genetic distances were also calculated, based again on the data from the common SNPs examined in the different studies, which were: A-M13; B-M182; E-M96*(xM35; M2); E-M35; E-M2; G-M201; I-M170; J-12f2a*; KT-M9*; P-92R7; Q-M242*; Q-M346*; Q-M3*; R-M207*; R-SRY10831.2; R-P25. The native populations were the same used for the Y-STRs pairwise comparisons, except the Panamanian Kuna Yala and the Mexican Pima, which were not considered due to the lack of Y- SNPs data, plus a population from Argentina (Roewer, L. et al. (2013). The European and African references were now the populations: from Spain (Adam, Susan M. et al. 2008) and Mozambique (Rowold, D. J., et al. 2016), respectively (Table 7).

Table 7: Description of the Native, European and African populations used in the Y-SNPs comparisons with the Guainía, Barí-Montilon and Paeces groups.

Reference	Country	Population	Code	N
Mazières et al. 2009	French Guiana	Wayampi	WAY	29
		Palikur	PAL	28
		Emerillon	EM	13
	Brazil	Apalai	AP	27
M. Gaya, Vidal et al. 2011	Bolivia	Quechuas	QU	55
		Aymaras	AY	59
Geppert et al. 2011	Equador	Kichwa	KI	15
		Wao	WAO	40
Roewer et al. 2013	Brazil	Kayapó-Xikrin	KX	10
		Zoé	Z	25
		Urubu-Kaapor	UK	27
	Argentina	Colla	COL	14
		Toba	TOBA	136
	Venezuela	Barí-Boxi	BB	16
		Wayuu	WAYUU	17
	Peru	Chuquibamba	CHU	16
		Huanca	HU	13
		Shipibo-Conibo	SC	21
	Colombia	Coconuco	COC	7
		Embera-Chamí	EC	24
		Guambiano	GUAM	16
Gonçalo Aragão 2018	Peru	Asháninka	ASH	53
Sandoval et al. 2012	Mexico	Maya	MAYA	14
		Nahuas	NA	56
Xavier, C. 2015	Colombia	Antiquoa	ANT	23
		Cauca	CAU	48
Battaglia et al. 2013	Panama	Native Panamarian	PA	36
Rowold, D.J., et al. 2016	Europe	Spain	SP	499
	Africa	Mozambique	MOZ	77

The F_{ST} results (Appendix Table 24) were quite different from those obtained with Y-STRs, with most of the pairwise distances between native populations were not statically significant (P≥0.05). This can be explained by the scarce resolution assumed within the Q clade (including Q-M3), which was the most well represented across the native groups. Thus, without accounting for its internal diversity, it is virtually impossible to capture differences between populations.

Focusing the 3 Colombian groups, the Barí-Motilon only differed significantly from the following native groups: Guambiano (F_{ST}=0.57468, P=0.00010) and Cauca (F_{ST}=0.26886, P=0.00010) both from Colombia; and the Nahuas (F_{ST}=0.33896, P=0.00000) and the Maya (F_{ST}=0.25616, P=0.00119 with) from Mexico. Contrarily to the Barí-Motilon, which were all Q-M3*, non-Native lineages were rather frequent in these 4 native groups, explaining thus their significant differentiation from the Barí-Motilon. The Guainía presented more significant differences with other native groups: in addition to the natives that differed from Barí-Motilon, they also from two Venezuelan groups (F_{ST}=0.12358, P=0.04594 for the Wayuu; F_{ST}=0.11752, P=0.04703 with the Barí-Boxi), all the Peruvian (except the Huanca) and all the Brazilian (except the Kayapó-Xikrin). Contrarily, the Paeces differed significantly from the remaining native populations under comparison, and for that must have accounted the fact that in the Paeces the majority of lineages were non-native.

Given the ancestry profiles of the three native populations here studied, it was not surprising to have found out that comparatively to the European reference the Paeces presented the lowest distance (F_{ST}=0.0949; P=0.02426), followed by the Guainía (F_{ST}=0.50074, P=0.0000) and only then the Barí-Motilon (F_{ST}=0.64725, P=0.0000).

The F_{ST} genetic distances were plotted in a two-dimensional MDS (Figure 20). The majority of the native groups is concentrated in a well-defined cluster, where the Guaínia and Barí-Montilon are integrated. Far away from this cluster are Colombian Natives Guambiano and Paeces, which are also well separated from each other, even so more close to the European reference. Again, the African reference stands isolated in the plot. The proximity between Guambiano and Paeces seems consistent with the higher frequency of the Q-M346* relatively to Q-M3* in both native groups, which is not observed in the other Natives, where Q-M3* lineages dominates.

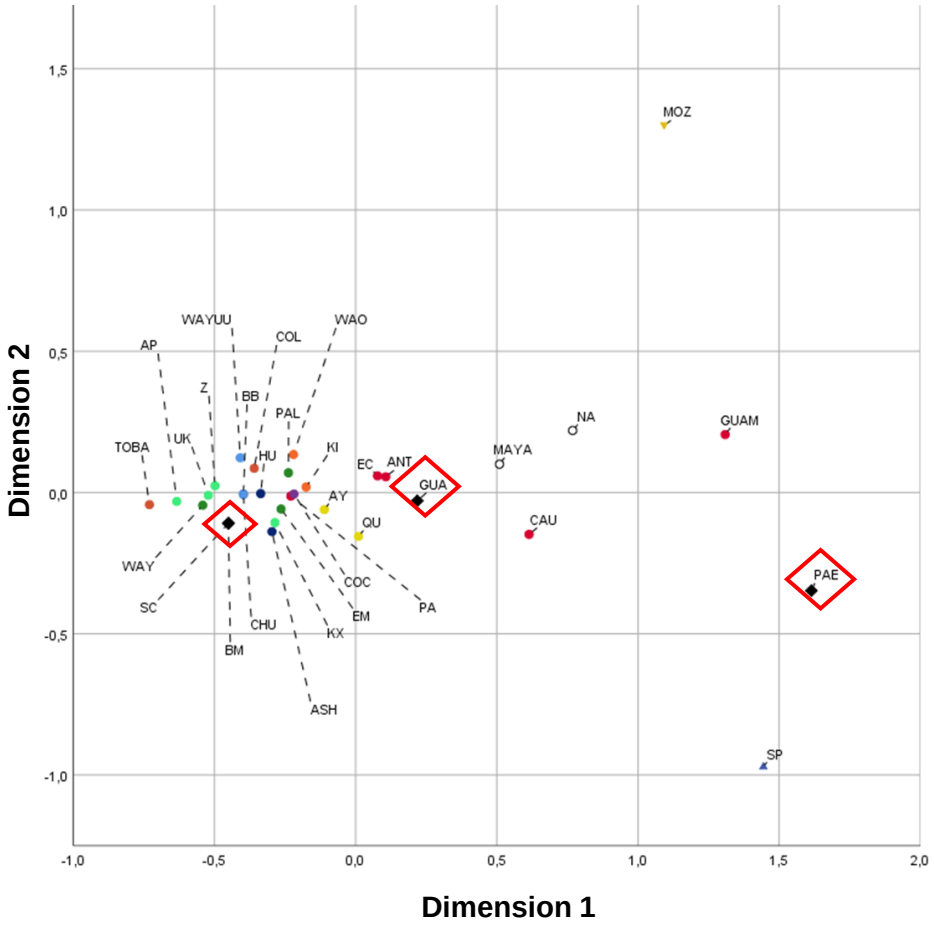


Figure 20: MDS plot based on F_{ST} genetic distances using 19 Y-SNPs (Stress=0.00840). Populations were coloured according to country: Colombian (Red), Argentina (Brown), Ecuador (Orange), French Guiana (Dark Green), Bolivia (Yellow), Panama (Purple), Venezuela (Blue), Mexico (White), Peru (Dark Blue), Brazil (Light Green). The Guainía (GUA), Barí-Motilon (BM) and Paeces (PAE) are marked with black diamonds and highlighted in a red outline, the European reference with a blue triangle and the African with a yellow triangle.

4.3.2. Comparisons between Admixed populations

Pairwise R_{ST} genetic distances between admixed populations were calculated with the same 15 Y-STRs that were used in the comparisons between native groups. In addition to the populations here studied, the following admixed populations from South and Central America were recruited for the analysis: Bolivia (Purps et al. 2014), Costa Rica (Purps et al. 2014), Argentina (Purps et al. 2014), Peru (Purps et al. 2014), Chile (Toscanini, U., et al. 2016), Ecuador (Toscanini, U., et al. 2018), Colombia (Ávila, S. J., et al. 2009), Puerto Rico (Vilar, M. G., et al. 2014), Mexico (López-Ramírez, Y. Let al. 2020) and 5 distinct regions of Brazil (Jannuzzi, J., et al. 2020). Two reference populations from Europe, one from Asia and another from sub-Saharan Africa (Purps et al. 2014) were also considered (Table 8), as well as the Barí-Motilon population as native reference population.

Table 8: Description of the Admixed, European, Asian and African populations used in the Y-STRs comparisons with the Cundinamarca, Huila, Santander del Norte and Tolima groups.

Reference	Country	Code	N
Purps et al. 2014	Bolivia	BOL	44
	Costa Rica	CR	165
	Argentina	ARG	50
	Panama	PAN	100
	Peru	PE	83
	Spain	SP	499
	Portugal	PT	248
	Yoruba	YOR	77
	China	CN	245
Toscanini et al. 2016	Chile	CH	196
Toscanini et al. 2018	Ecuador	EC	240
Ávila et al. 2009	Colombia	COL	287
Vilar et al. 2014	Puerto Rico	PR	121
López-Ramirez et al. 2020	Mexico	MEX	295
Jannuzzi et al. 2020	Northeast Brazil	NE BR	532
	North Brazil	NN BR	83
	Southeast Brazil	SE BR	244
	Central West Brazil	CW BR	303
	South Brazil	S BR	210

Most of the pairwise RST genetic distances involving our four admixture populations (Cundinamarca, Santander del Norte, Huila and Tolima) were non-significant (Appendix Table 25). Huila only differed from Bolivia (RST=0.05286, P= 0.04505) while Santander del Norte showed the highest number of significant distances.

The admixed populations from Colombia, Costa Rica, Argentine, Northern and Southeastern Brazil did not revealed any significant difference from Cundinamarca, Santander del Norte, Huila or Tolima.

The relationships between the populations considered in this analysis are summarized in the MDS plot shown in Figure 21. A major cluster is visible containing all admixed populations as well as European references. Some populations are more peripheral in the cluster, like Bolivia for instance. Tolima, Santander del Norte, and Huila are very close together, whereas Cundinamarca is a little more separated from them, tending towards the African reference in the plot. As stated before, Cundinamarca is the most diverse of the four in terms of genetic ancestrally.

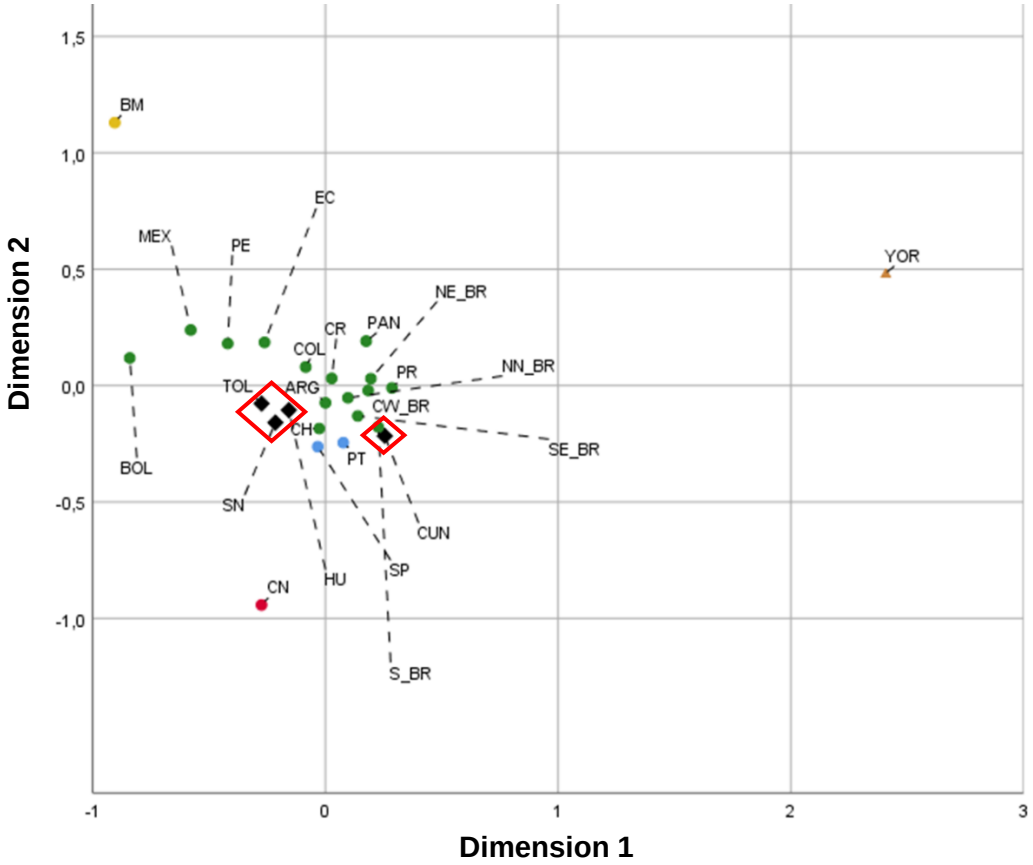


Figure 21: MDS plot based on the R_{ST} genetic distances calculated using the information of 15 Y-STR, between the 20 populations analysed (Stress=0.01233). Our populations are shown in a dark diamond and are highlighted in a red outline, other admixed groups are shown in green, the native one in yellow, the Asian in red and the African population in a brown triangle

Using Y-SNP data, the pairwise F_{ST} genetic distances were also calculated. This time, the resolution power was more reduced, due to the need to accommodate the lesser number of Y-SNPs shared across multiple studies, which totalized ninety: A-M13; A*(M13); B-M182; C-M130; E-M96*(xM35, M2); E-M35; E-M2; F-M213*(xM9); G-M201; I-M170; J-12f2a; KLNT-M9; P-92R7; Q-M242*(xM346); Q-M346*(xM3); Q-M3*; R-M207*(xSRY10831.2, P25); R-SRY10831.2 and R-P25. The populations were the following: Puerto Rico (Vilar et al. 2014), Colombia (Noguera, M. C., et al. 2014), Nicaragua (Núñez, C., et al. 2012), Bolivia (Cárdenas, J. M., et al. 2015), Chile (Flores-Alvarado, S., et al. 2022), Argentina (Corach, D., et al. 2010) and North, North East, Central West, South and South East Brazil (Resque et al. 2016). The Barí-Motilon as native reference, Spain as the reference from Europe (Adam, Susan M. et al. 2008) and Angola as the reference from Sub-Saharan Africa (Brito, P., et al. 2011) were also included (Table 9).

Table 9: Description of the Admixed, European and African populations used in the Y-SNPs comparisons with the Cundinamarca, Huila, Santander del Norte and Tolima groups.

Reference	Country	Code	N
Cárdenas, J. M., et al. 2015	Bolivia	BOL	103
Corach, D., et al. 2010	Argentina	ARG	250
Flores-Alvarado, S., et al. 2022	Chile	CH	59
Núñez, C., et al. 2012	Nicaragua	NI	55
Noguera, M. C., et al. 2014	Colombia	COL	175
Vilar et al. 2014	Puerto Rico	PR	121
Resque et al. 2016	Northeast Brazil	NEB	243
	North Brazil	NB	272
	Southeast Brazil	SEB	330
	Central West Brazil	CWB	135
	South Brazil	SB	237
Adam, S.M. et al. 2008	Spain	SP	140
Brito, P. et al. 2011	Angola	ANG	100

Significant F_{ST} values between our four admixed populations and other populations were only rarely observed (Appendix Table 26). Tolima revealed the larger number of non-significant distances, only differing from Central West, South and Southeast Brazil. By the way, most populations from Brazil showed significant distances with our admixed Colombian populations (except Northern Brazil in general, and Northeastern Brazil with Tolima). A closer look at these populations revealed the lack of haplogroup J-12f2a, which was present at low frequencies in all other populations.

Also of note were the non-significant distances between Bolivia and Huila or Tolima (F_{ST}=0.00333, P=0.66102 and F_{ST}=0.02172, P=0.17860, respectively), since the corresponding distances obtained with Y-STRs had been significant. Examining the Bolivian dataset, the high frequency of haplogroup R called attention, a feature that was also observed not only in Huila and Tolima but also in Cundinamarca and Santander del Norte. Thus, the differentiations now detected must be due to differences in the remaining haplogroups.

The two-dimensional MDS shown in Figure 22, which was constructed using the F_{ST} genetic distances values, is quite similar to that obtained with STR-based distances. A sharp cluster including the different admixed populations is well defined, with the European reference being very near it. In the plot Huila is very close European population, which is in agreement with the non-significant distance between them (F_{ST}=0.01213, P=0.26047). The four populations here studied are well integrated in the cluster, and within it the Brazilian populations tend to be the less central.

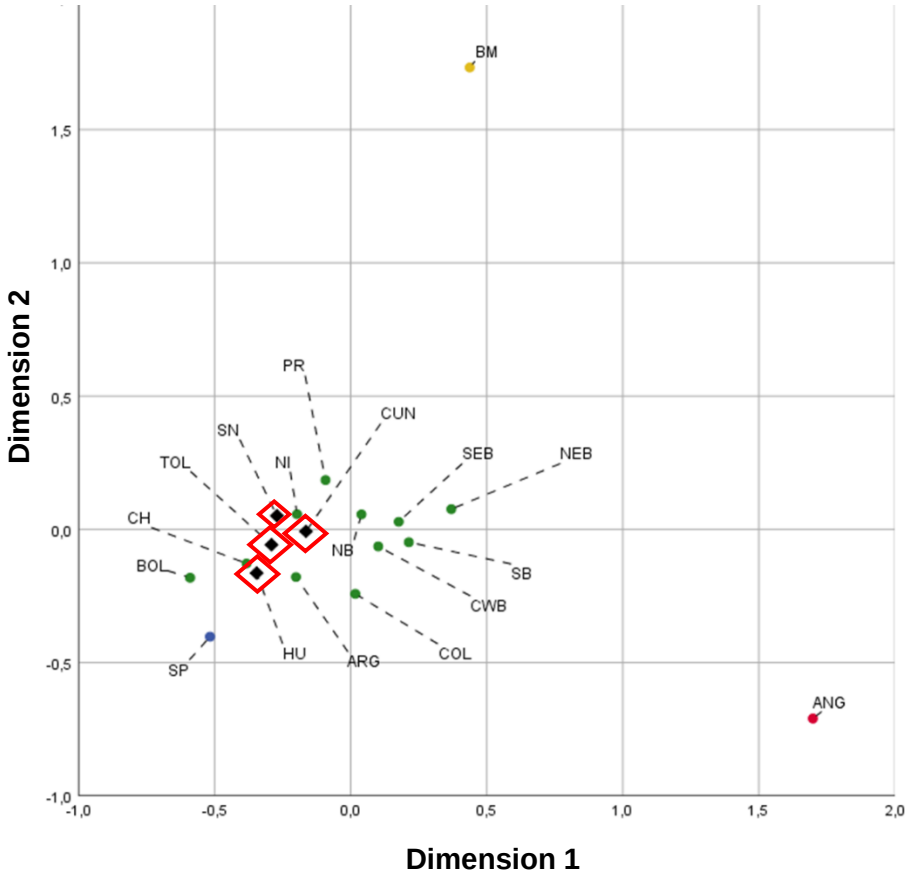


Figure 22: MDS plot based on the F_{ST} genetic distances calculated considering 19 Y-haplogroups, between the 19 populations analysed (Stress= 0.01723). Our populations are shown in a dark diamond and are highlighted by a red outline, other admixed groups are shown in green, the native one in yellow and the African in red.

4.3.3. Comparisons with Eurasian populations

The European ancestry is well impregnated in the male pools of all 7 Colombian populations, especially in the admixed ones. In fact, Colombia have had a large influx of European people in the last centuries, and also of Middle Eastern immigrants at least in the past decades. In order to obtain clues on the origin of the Eurasian lineages found in Colombia, a comprehensive analysis was undertaken with an enlarged series of European and Middle Eastern populations.

All these populations had available data for 19 Y-STRs (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y GATA H4, DYS576, DYS481, DYS533, DYS570), which used to calculate R_{ST} pairwise distances. The comparison involved 6 of the Colombian samples, those where Eurasian lineages were detected, twenty-one populations from Europe, one from the Middle East and one from Southeast Asia (Table 10). The European populations were from: Belgium, Croatia, Chezchia, England, Germany, Greece, Hungary, Ireland, the Netherlands, Poland, Portugal, Switzerland, Wales, Macedonia, Denmark, three regions from Spain (Asturias, Basque Country, and Madrid) and three from Italy (Northern, Southern, and Sicily) (Purps et al. 2014). The Middle East was represented by a population from Lebanon (Purps et al. 2014). The details on each population can be found at Table 10.

Table 10: Description of the Eurasian populations used in the Y-STRs comparisons with the Guainía, Paeces, Cundinamarca, Huila, Santander del Norte and Tolima groups.

Reference	Origin	Population	Code	N
Present Study	Colombia	Guainía	GUA	4
		Paeces	PAE	8
		Cundinamarca	CUN	16
		Huila	HUI	10
		Santander del Norte	SN	18
		Tolima	TOL	13
Purps et al. 2014	Europe	Belgium	BEL	206
		Croatia	CRO	124
		Czechia	CZE	72
		England	ENG	80
		Germany	GER	128
		Greece	GRE	213
		Hungary	HUN	141
		Ireland	IRE	30
		Brescia Italy (North)	BR_IT	124
		Puglia Italy (South)	PU_IT	159
		Sicily	SI_IT	157
		Netherlands	NET	94
		Poland	POL	150
		Portugal	POR	83
		Basque Country	AR_SP	199
		Asturias Spain	AS_SP	254
		Madrid	MA_SP	125
		Switzerland	SWI	149
		Wales	WAL	118
		Macedonia	MAC	101
		Denmark	DEN	185
	Middle East	Lebanon	LEB	503

The R_{ST} genetic distance results for the 6 Colombian population varied a lot (Appendix Table 27). The Native Guainía showed significant distances with Ireland (R_{ST}=0.19808, P=0.03148), Northern Italy (R_{ST}=0.16535, P=0.02307), Poland (R_{ST}=0.12443, P=0.03881), Spanish Madrid (R_{ST}=0.12857, P=0.04604) and Wales (R_{ST}=0.30822, P=0.00465). The R_{ST} values involving the Paeces were very different from those involving the Guainía: the only result that did not differed significantly was from Ireland (R_{ST}=0.09093, P=0.08029).

Moving on to admixed populations, Cundinamarca presented non-significant R_{ST} values with a variety of European populations: Belgium (R_{ST}=0.02095, P=0.11821), Czechia (R_{ST}=0.02118, P=0.12900), Germany (R_{ST}=0.01558, P=0.14979), Southern Italy (R_{ST}=-0.01071, P=0.72864), Sicily (R_{ST}=-0.0059, P=0.59192), the Netherlands (R_{ST}=0.01794, P=0.16721), Portugal (R_{ST}=0.00784, P=0.25740), Spanish Asturias (R_{ST}=0.01176, P=0.19275), Spanish Madrid (R_{ST}=0.03351 ,P=0.05148), Switzerland (R_{ST}=0.01536, P=0.19157) and Denmark (R_{ST}=0.01349, P=0.20731). Given that the capital of the country, Bogotá, is located in the Cundinamarca administrative region, the entire region is a melting point of people with very distinct origins, which can justify the absence of more distances statistically significant between Cundinamarca and European populations.

Also in Huila the few significant distances were with Croatia (R_{ST}=0.14779, P=0.00743), Poland (R_{ST}=0.13058, P=0.00208), Wales (R_{ST}=0.07834, P=0.01703), Macedonia (R_{ST}=0.09454, P=0.01782) and Lebanon (R_{ST}=0.05645, P=0.03812). Since this department neighbours the Cundinamarca region, it can have absorbed many of the demographic influences that modelled the capital.

Considering Santander del Norte, most of its distances from other populations were statistically significant, excepting from England (R_{ST}=0.03251, P=0.05950), Ireland (R_{ST}=0.02251, P=0.10445), the Netherlands (R_{ST}=0.03219, P=0.05554), Portugal (R_{ST}=0.02989, P=0.06910) and the three Spanish populations (R_{ST}=0.01841, P=0.13415 for the Basque Country; R_{ST}=0.02303, P=0.09227 for Asturias; R_{ST}=0.014, P=0.18454 for Madrid).

Results involving Tolima were similar to those involving Santander del Norte. The populations with non-statistically genetic distances different from Tolima were Belgium (R_{ST}=0.01721, P=0.16127), England (R_{ST}=0.02332, P=0.14405), Ireland (R_{ST}=0.02388, P=0.15939), Northern and Southern Italy (R_{ST}=-0.02737, P=0.96485 and R_{ST}=0.03422, P=0.05227 respectively), the Netherlands (R_{ST}=0.02614, P=0.15107), Portugal (R_{ST}=0.00577, P=0.30096), the three Spanish populations (R_{ST}=-0.00146, P=0.41441 for the Basque Country; R_{ST}=-0.00897, P=0.58489 for Asturias; R_{ST}=-0.00978, P=0.61073 for Madrid), Switzerland (R_{ST}=0.02363, P=0.16295) and Wales (R_{ST}=0.03342, P=0.06910).

Like in Santander del Norte, also in Tolima the component of European ancestry seems to have its main roots in the Iberian Peninsula.

With the exception of Guainía, all populations presented statistically significant distances with the Middle Eastern population, suggesting that the most of Eurasian component in the Colombian populations, most likely, has originated within European countries.

A two-dimensional MDS was constructed using the R_{ST} genetic distances values (Figure 23). The different European populations seem to be grouped mainly according to a geographical pattern. For instance, the Celtic speaking Wales and Ireland are close together; the Iberian Peninsula, Switzerland, Northwest Europe and Italy are positioned nearby, the Balkans (Greece, Macedonia and Croatia) are further from each other, but nonetheless forming a coherent group; Germany, Hungary, Denmark and Czechia are more peripheral; Poland is quite isolated, and then Lebanon relatively far from the remaining populations.

Concerning the Colombian populations, apart the Paeces that are quite isolated in the plot, the remaining are intermingled with various European populations, with some being rather near populations from the Iberian Peninsula, Italy and the British Isles. Cundinamarca and Huila are in the middle of populations from Central and East Europe. The native Colombian populations are spread across the plot, quite far from one another.

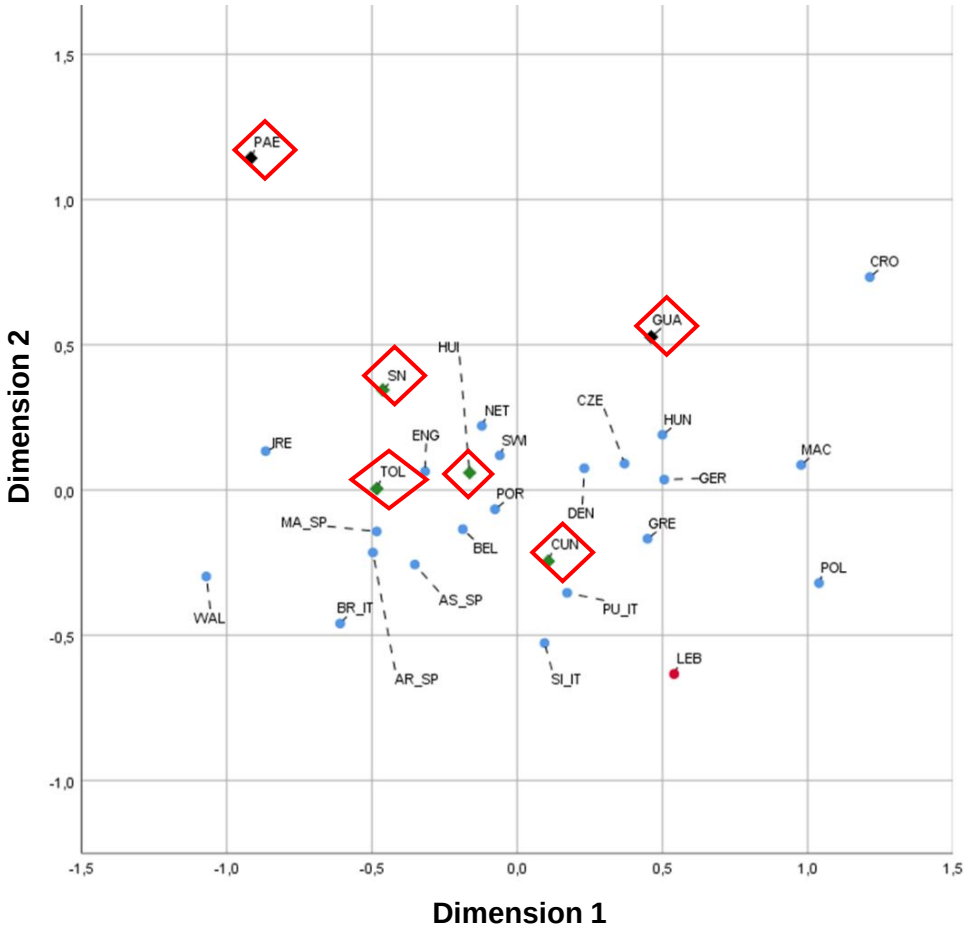


Figure 23: MDS plot based on the R_{ST} genetic distances calculated using the information of 19 Y-STR, between the 29 populations analysed (Stress=0.02648). Our populations are shown in diamond shapes and highlighted by red outlines, with the black ones representing native groups and green being the admixture; the European groups are shown in blue and the one from Middle East in red.

Identical analysis was conducted with the pairwise F_{ST} distances obtained with Y-SNP data. The haplogroups considered were: E-M35*(xM81, M78, M123); E-M78; E-M81; E-M123; F-M213*; G-M201; I-M170; J-12f2a; KLNT-M9; R-M207*(xSRY10831.2, M269); R-SRT10831.2; R-M269. The European populations compared were from: Portugal (Adams, S.M., et al. 2008), Germany (Rębała, K., et al. 2013), Poland (Rębała, K., et al. 2013), Croatia (Peričić, M., et al. 2005), Hungrary (Karachanak, S., et al. 2013), Denmark (Sanchez, J. J., et al., 2004), United Kingdom (Martínez-Cruz, B. et al., 2011), Greece (Martínez-Cruz, B. et al., 2011), France (Ramos-Luis, E., et al. 2014), three regions of Spain (Galicia, Basque Country and Asturias) (Adams, S.M. et al. 2008) and three regions of Italy (North, South and Sicily) (Boattini, A., et al. 2013). In addition, two populations from the Middle East were included: one from the Turkish region of Anatolia (Cinnioğlu, C., et al. 2004) and another from Lebanon (Zalloua, P. A., et al. 2008) (Table 11).

Table 11: Description of the Eurasian populations used in the Y-SNPs comparisons with the Guainía, Paeces, Cundinamarca, Huila, Santander del Norte and Tolima groups.

Study	Origin	Population	Code	N
Present Study	Colombia	Guainía	GUA	4
		Paeces	PAE	8
		Cundinamarca	CUN	16
		Huila	HU	10
		Santander del Norte	SN	18
		Tolima	TOL	13
Adams, S.M. et al. 2008	Europe	Portugal	PT	77
		Galicia (Spain)	GAL	88
		Basque Country (Spain)	BC	115
		Asturias (Spain)	AST	20
Rębała, K., et al. 2013		Germany	GER	123
		Poland	POL	202
Peričić, M., et al. 2005		Croatia	CRO	109
Karachanak, S., et al. 2013		Hungary	HUN	53
Sanchez, J. J., et al., 2004		Denmark	DEN	194
Martínez-Cruz, B. et al., 2011		United Kingdom	UK	19
		Greece	GRE	22
Ramos-Luis, E., et al. 2014		France	FR	86
Boattini, A., et al. 2013		North Italy	N_IT	161
		South Italy	S_IT	198
		Sicily (Italy)	SIC	140
Cinnioğlu, C., et al. 2004	Middle East	Turkey	TK	43
Zalloua, P. A., et al. 2008		Lebanon	LEB	83

Unlike the results based on Y-STRs distances, the rate of significant/non-significant F_{ST} distances was very similar in all six Colombian populations (Guainía, Paeces, Cundinamarca, Huila, Santander del Norte and Tolima) (Appendix Table 28). The statistically significant values were with Poland, Germany, Croatia and Hungary for all six Colombian populations. In all these populations the frequency of haplogroup R1a was high, especially when compared with haplogroup R1b. As already mentioned, R1a is quite common in East Europe, while the R1b is primarily found in countries from West Europe.

In opposition, Asturias, Galicia, Portugal, United Kingdom, France and Northern Italy did not revealed any F_{ST} distance statistically significant involving each of the six Colombian populations. In all these populations (except Guainía), R1b lineages prevail compared with R1a.

Overall, these results fit the existing historical records that document the important role played by West European colonizers and immigrates in reshaping the Colombian genetic background.

The pairwise F_{ST} genetic distances were plotted into a MDS (Figure 24). The admixed Colombian populations are closer together in the middle of Western European populations. The Guainía stands somehow closer to the populations from Middle East and Greece. The Paeces stand out, because unlike their isolated positioning in Y-STRs based MDS, now they are not so far from the rest of the European populations.

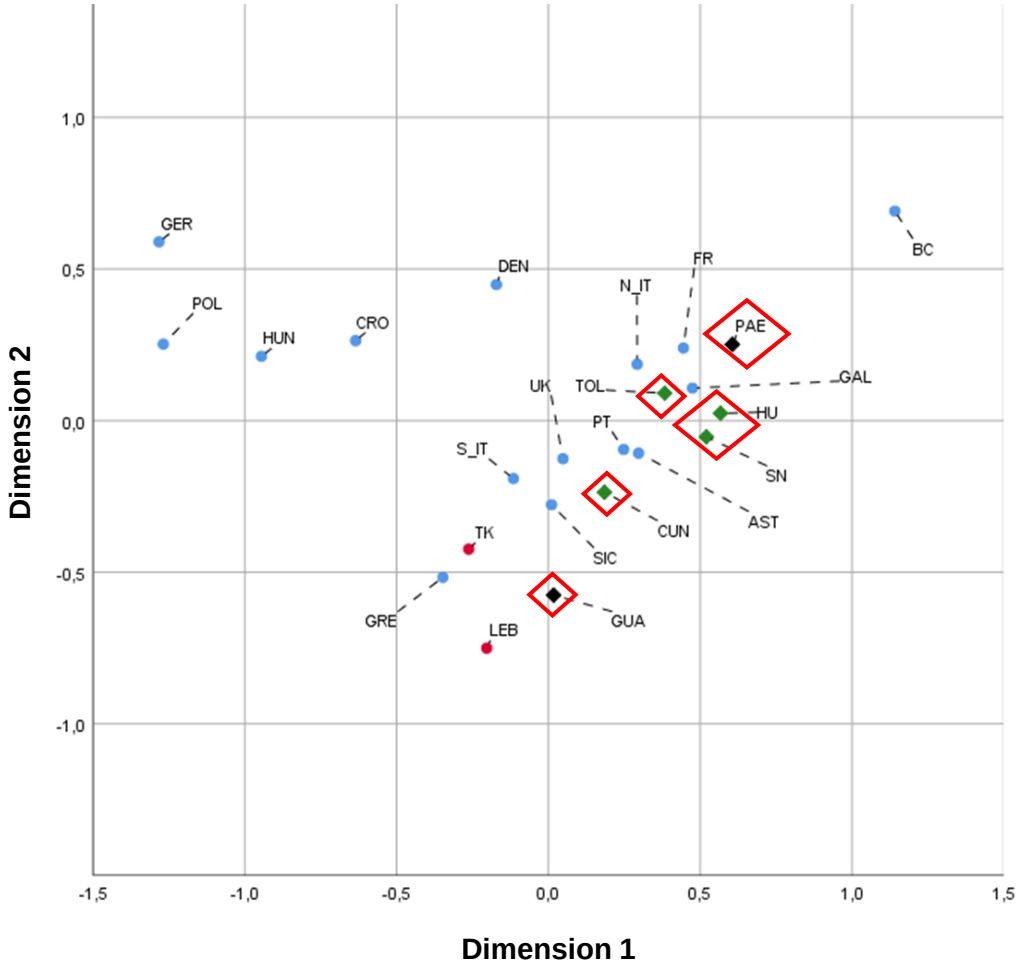


Figure 24: MDS plot based on the F_{ST} genetic distances calculated using the information of 12 Y-SNPs, between the 23 populations analysed (Stress= 0.01300). Our populations are shown in diamond shapes and are highlighted by red outlines, with the black ones representing native groups and green being the admixture; the European groups are shown in blue and the ones from Middle East in red.

5. Conclusion

The contrast between Native and Admixed Colombian Y-chromosomal lineages provides light on key demographic events that occurred in the region that corresponds nowadays to Colombia. After the first people stepped foot on the American continent and before the waves of colonization and migrations that started in the XVI century, Colombia was inhabited exclusively by Native American populations. With the beginning of the Exploration Age, European settlers and the African slaves they brought established contacts with different Native populations, some peaceful and other bloodier, but regardless, leaving traces that deeply impregnated in the genetic composition of people from the region. In the XIX and XX centuries, with the independence of the Colombian state, a large influx of Europeans and Asians from Middle East continued to alter the population landscape.

The male diversity in Native and Admixed populations varies amply. On average, 80% of the male lineages in the admixed groups' are of Eurasian origin, and the remaining of Native American ancestry. Exceptionally, in Cundinamarca 10% of lineages were African. These values are consistent with those reported by Ossa, H. et al (2016), who used ancestry-informative insertion/deletion markers (AIM-Indels) to study populations from the Andean region of the Colombia, among which the European ancestry was the majority, albeit, not as high as in our groups. The weak presence of Native American Y-lineages is likely related to the social hierarchy imposed by an European elite that spitefully considered the Native people in the bottom. Also the scarcity of the African ancestry suggests that the African people forced to migrate to the country during the transatlantic slave trade didn't contribute significantly to the admixture process from which arose the populations from urban areas of the country.

Recently, a study of mtDNA diversity addressing the same Colombian admixed populations demonstrated that the majority of the maternal lineages had Native American origin (Ribeiro, B. P. A., 2021). Crossing this finding with our results, the inferred genetic background of these groups gives overt clues on their underlying process of admixture: predominantly European males and Native woman were the main constructors of the current-day communities. This is a pattern found throughout admixture groups not only from Colombia (Rojas, W., et al. (2010), but also from South America in general, (Wang, S., et al. (2008).

Of the native groups, one, the Barí-Motilon, harboured exclusively Native Y-haplogroups. In the Guainía 72% of Y-lineages were Native, 22% Eurasian and the remaining 6% African. This indicates that besides the European also African men contributed to reconfigure the paternal

genetic background of the Guainía, where the majority of lineages still belong to Native groups. The Paeces is the Indigenous group with an ancestrally resembling more the pattern typically found in admixed populations: 67% of its paternal lineages belong to Eurasian haplogroups and 33% have Native ancestrally. Given that the Paeces habit the center of Andes, that may have contributed to intense admixture with Europeans. However, knowing that confronts between Paeces and colonists were particularly violent, we can also admit a scenario of strong devastation of the original Paeces accompanied by rapid expansion of the European colonists in their communities.

The great heterogeneity across the distinct native populations here studied, agrees with data reported in Xavier, C. et al. (2015) for other two Native groups from Colombian, one of which presented 100% of lineages with Native ancestry, contrarily to the other that contained lineages of distinct origins.

The genetic comparisons performed between our Native groups and other Latin American Native populations using Y-STR data, revealed that the Guainía and Paeces are well integrated in the context of the remaining Native populations. The Barí-Montilon showed to be closer to the Venezuelan Barí-Boxi group than to other groups, regardless of their countries. The Barí-Montilon and the Barí-Boxi live very close geographically, being only separated by the Colombia-Venezuela border, and even share the same language, which can explain the strong genetic ties between each other.

Y-SNP data further showed that whereas the Barí-Motilon are close to the great majority of other Latin American Native groups, the Guainía are more differentiated from non-Colombian groups, and the Paeces clearly depart from other Natives. In the case of the Paeces, the very high European ancestry justifies the deviation from other native groups and the trend to come close to European reference used in the analysis.

The admixed Colombian populations also fitted well in other Latin American admixed populations, when compared on the basis of Y-STR or Y-SNP data. In general, all admixed populations from America Latina revealed a predominant European ancestry, usually complemented only by a Native-American component. The African ancestry was in most cases residual, evidencing the modest influence that African people had in shaping diversity in those populations.

Overall, 56% of our samples were Eurasian lineages. Given this high proportion, to dissect better the Eurasian component a comprehensive analysis was conducted recruiting the

Colombian populations as well populations from other countries, mainly from Europe and the Middle East. In general, the Colombian populations were closest to those from West Europe, than from the Middle East and Eastern Europe. These findings are consistent with historical records documenting the remarkable influx into Colombia of Iberian people, especially from Spain.

In summary, from the Y-STR and Y-SNP data obtained in this study for Colombian populations (Native and admixed), it was possible to extract valuable inferences on the demographic events that shaped their patterns of male genetic diversity.

In the future, it would be important to analyse these populations for other kind of markers, and cross the data with that here obtained. More populations should also be studied. All populations and genetic tools available will be crucial to unravel the complex population history of America.

6. References

Adams, Susan M., et al. "The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula." The American Journal of Human Genetics 83.6 (2008): 725-736.

Adnan, A., Ralf, A., Rakha, A., Kousouri, N., & Kayser, M. (2016). Improving empirical evidence on differentiating closely related men with RM Y-STRs: a comprehensive pedigree study from Pakistan. Forensic Science International: Genetics, 25, 45-51.

Alghafri, R., Goodwin, W., Ralf, A., Kayser, M., & Hadi, S. (2015). A novel multiplex assay for simultaneously analysing 13 rapidly mutating Y-STRs. Forensic Science International: Genetics, 17, 91-98.

Alonso Morales, L. A., Casas-Vargas, A., Rojas Castro, M., Resque, R., Ribeiro-dos-Santos, Â. K., Santos, S., Gusmão, L. & Usaquén, W. (2018). Paternal portrait of populations of the middle Magdalena River region (Tolima and Huila, Colombia): New insights on the peopling of Central America and northernmost South America. PloS one, 13(11), e0207130.

Alvarez, L., Santos, C., Montiel, R., Caeiro, B., Baali, A., Dugoujon, J. M., & Aluja, M. P. (2009). Y-chromosome variation in South Iberia: insights into the North African contribution. American Journal of Human Biology: The Official Journal of the Human Biology Association, 21(3), 407-409.

Alves, Cíntia et al. 2003. " Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data." Forensic science international **134**(2-3): 126-133

Amos, William (2010). " Mutation Biases and Mutation Rate Variation Around Very Short Human Microsatellites Revealed by Human–Chimpanzee–Orangutan Genomic Sequence Alignments." Journal of Molecular Evolution **71**(3): 192-201

Ansari Pour, Naser, Christopher A. Plaster, and Neil Bradman. "Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people." European Journal of Human Genetics 21.4 (2013): 423-429.

Anzola, P. and Luis, F. (2004). The mining in the Departments of Guainía and Vaupes: socioeconomic condition and the population's dynamics. Small auriferous mining-Toward an environmental handling of mercury and cyanide.

Arango, R., and E. Sánchez. "Los pueblos indígenas de Colombia en el umbral del nuevo milenio. Departamento Nacional de Planeación." (2004).

Athey, T. W. (2006). Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. J Genet Geneal, 2(2), 34-9.

Ávila, S. J., Briceño, I., & Gómez, A. (2009). Genetic population analysis of 17 Y-chromosomal STRs in three states (Valle del Cauca, Cauca and Nariño) from Southwestern Colombia. Journal of Forensic and Legal Medicine, 16(4), 204-211.

Avvaru, A. K., Sharma, D., Verma, A., Mishra, R. K., & Sowpati, D. T. (2020). MSDB: a comprehensive, annotated database of microsatellites. Nucleic acids research, 48(D1), D155-D159.

Ayub, Q., Haber, M., & Tyler-Smith, C. (2016). Evolutionary and Population Genetics in Forensic Science. Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation, 2, 33.

Bachtrog, D. (2006). "A dynamic view of sex chromosome evolution." Current Opinion in Genetics & Development **16**(6): 578-585.

Bachtrog, D. (2013). "Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration." Nature Reviews Genetics **14**(2): 113-124.

Baeta, M., Núñez, C., Villaescusa, P., Ortueta, U., Ibarbia, N., Herrera, R. J., Blazquez-Caeiro, J. L., Builes, J. J., Jiménez-Moreno, S. & de Pancorbo, M. M. (2018). Assessment of a subset of Slowly Mutating Y-STRs for forensic and evolutionary studies. Forensic Science International: Genetics, 34, e7-e12.

Bailliet, Graciela, et al. "Brief communication: Restricted geographic distribution for Y-Q* paragroup in South America." American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists 140.3 (2009): 578-582.

Ballantyne, Kaye N., et al. "Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications." The American Journal of Human Genetics 87.3 (2010): 341-353.

Ballantyne, K. N., and Manfred, K. "Additional Y-STRs in Forensics: Why, Which, and When." Forensic science review 24.1 (2012): 63-78.

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

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Molecular biology and evolution, 16(1), 37-48.

Battaglia, Vincenza, et al. "The first peopling of South America: new evidence from Y-chromosome haplogroup Q." PLoS One 8.8 (2013): e71390.

Berniell-Lee, Gemma, et al. "Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages." Molecular biology and evolution 26.7 (2009): 1581-1589.

Bickmore, W. A. and Cooke, H. J. (1987). "Evolution of homologous sequences on the human X and Y chromosomes, outside of the meiotic pairing segment." Nucleic Acids Research **15**(15): 6261-6271.

Boattini, Alessio, et al. "Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata." PloS one 8.5 (2013): e65441.

Bortolini, M.C., et al. (2003). "Y-chromosome evidence for differing ancient demographic histories in the Americas." American journal of human genetics **73**(3): 524-539.

Brinkmann, B. et al. (1998) "Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat." American journal of human genetics **62**(6): 1408-1415

Brion, M., Sobrino, B., Blanco-Verea, A., Lareu, M. V., & Carracedo, A. (2005). Hierarchical analysis of 30 Y-chromosome SNPs in European populations. International Journal of Legal Medicine, 119(1), 10-15.

Brito, P., et al. "Y-SNP analysis in an Angola population." Forensic Science International: Genetics Supplement Series 3.1 (2011): e369-e370.

Brookes, Anthony J. (1999) "The essence of SNPs." Gene **234**(2): 177-186

Brown, Matthew (2012). "The struggle for power in post-independence Colombia and Venezuela" Springer

Bruckmayr, P. (2010). Syro-Lebanese migration to Colombia, Venezuela and Curacão: From mainly christian to predominantly Muslim phenomenon. European Journal of Economic and Political Studies, 3(Special Issue), 151-178.

Budowle, B., Chakraborty, R., Giusti, A. M., Eisenberg, A. J., & Allen, R. C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. American journal of human genetics, 48(1), 137.

Budowle, B., & Van Daal, A. (2008). Forensically relevant SNP classes. Biotechniques, 44(5), 603-610.

Butler, J. M. (2005). Forensic DNA typing: biology, technology, and genetics of STR markers. Elsevier.

Campos, A. C. (2018). Genetic characterization of the male lineages present in Nigeria population.

Canale, L. C., Parson, W., and Holland, M. M. (2022). The time is now for ubiquitous forensic mtMPS analysis. Wiley Interdisciplinary Reviews: Forensic Science, 4(1), e1431.

Capodiferro, Marco Rosario, et al. "Archaeogenomic distinctiveness of the Isthmo-Colombian area." Cell 184.7 (2021): 1706-1723.

Cárdenas, J. M., et al. "The multiethnic ancestry of Bolivians as revealed by the analysis of Y-chromosome markers." Forensic Science International: Genetics 14 (2015): 210-218.

Carvalho, M., et al. "Paternal and maternal lineages in Guinea-Bissau population." Forensic Science International: Genetics 5.2 (2011): 114-116.

Casas-Vargas, A., et al. (2011). "High genetic diversity on a sample of pre-Columbian bone remains from Guane territories in northwestern Colombia." American Journal of Physical Anthropology **146**(4): 637-649.

Castro e Silva, M. A., Ferraz, T., Couto-Silva, C. M., Lemes, R. B., Nunes, K., Comas, D., & Hünemeier, T. (2022). Population histories and genomic diversity of South American natives. Molecular biology and evolution, 39(1).

Castro, M. A., Ferraz, T., Bortolini, M. C., Comas, D., & Hünemeier, T. (2021). Deep genetic affinity between coastal Pacific and Amazonian natives evidenced by Australasian ancestry. *Proceedings of the National Academy of Sciences*, *118*(14).

Cetkovic Gentula, M., and A. Nevski. "NevGen Y-DNA haplogroup predictor." (2015).

Cinnioğlu, C., King, R., Kivisild, T., Kalfoğlu, E., Atasoy, S., Cavalleri, G. L., Lillie, A. S., Roseman, C. C., Lin, A. A., Prince, K., Oefner, P. J., Shen, P., Semino, O., Cavalli-Sforza, L. L., & Underhill, P. A. (2004). Excavating Y-chromosome haplotype strata in Anatolia. Human Genetics, 114(2), 127–148.

Class, B., & Dingemanse, N. J. (2022). A variance partitioning perspective of assortative mating: Proximate mechanisms and evolutionary implications. Journal of Evolutionary Biology, 35(4), 483-490.

Colaco, Stacy, and Deepak Modi. "Genetics of the human Y chromosome and its association with male infertility." Reproductive biology and endocrinology 16.1 (2018): 1-24.

Collard, B. C., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica, 142(1), 169-196.

Cook, N. D. (2002). "Sickness, starvation, and death in early Hispaniola." Journal of Interdisciplinary History **32**(3): 349-386.

Corach, Daniel, et al. "Inferring continental ancestry of Argentineans from autosomal, Y-chromosomal and mitochondrial DNA." Annals of human genetics 74.1 (2010): 65-76.

Cortez, D., et al. (2014). "Origins and functional evolution of Y chromosomes across mammals." Nature **508**(7497): 488-493.

Cruciani, Fulvio, et al. "Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa." The American Journal of Human Genetics 74.5 (2004): 1014-1022.

Cruciani, Fulvio, et al. "Strong intra-and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152." Forensic Science International: Genetics 5.3 (2011): e49-e52.

D'Atanasio, Eugenia, et al. "The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages." Genome biology 19.1 (2018): 1-15.

D'Atanasio, E., Iacovacci, G., Pistillo, R., Bonito, M., Dugoujon, J. M., Moral, P., El-Chennawi, F., Melhaoui, M., Baali, A. & Cruciani, F. (2019). Rapidly mutating Y-STRs in rapidly expanding populations: Discrimination power of the Yfiler Plus multiplex in northern Africa. Forensic Science International: Genetics, 38, 185-194.

Dauber, E. M., Kratzer, A., Neuhuber, F., Parson, W., Klintschar, M., Bär, W., & Mayr, W. R. (2012). Germline mutations of STR-alleles include multi-step mutations as defined by sequencing of repeat and flanking regions. Forensic Science International: Genetics, 6(3), 381-386.

de Filippo, Cesare, et al. "Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups." Molecular biology and evolution 28.3 (2011): 1255-1269.

de Pérez, M. S. G. (2010). Las lenguas indígenas de Colombia a vuelo de pájaro. Revista de Lenguas Indígenas y Universos Culturales Año 2010, núm. 7, 2010(7), 9-25.

de Knijff, P. (2022). On the Forensic Use of Y-Chromosome Polymorphisms. Genes, 13(5), 898.

Demy, T. J., & Shaw, J. M. (Eds.). (2019). Religion and Contemporary Politics: A Global Encyclopedia [2 volumes]. ABC-CLIO.

Di Giacomo, Fabio, et al. "Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe." Human genetics 115.5 (2004): 357-371.

Dixon, E. J. (2001). "Human colonization of the Americas: timing, technology and process." Quaternary Science Reviews **20**(1-3): 277-299.

Dulik, Matthew C., et al. "Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians." The American Journal of Human Genetics 90.2 (2012): 229-246.

Edelmann, Jeanett et al. 2002. "Sequence variation and allele nomenclature for the X-linked STRs DXS9895, DXS8378, DXS7132, DXS6800, DXS7133, GATA172D05, DXS7423 and DXS8377." Forensic Science International **129**(2): 99-103

Escobar, C. (2015). Report on citizenship law: Colombia.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources, 10(3), 564-567.

Fawcett, L., & Posada-Carbo, E. (2013). Arabs and Jews in the development of the Colombian Caribbean 1850–1950. Arab and Jewish Immigrants in Latin America, 57-79.

Flores-Alvarado, S., Orellana-Soto, M., & Moraga, M. (2022). Ancestry and admixture of a southernmost Chilean population: The reflection of a migratory history. American Journal of Human Biology, 34(1), e23598.

Forster, P., Hohoff, C., Dunkelmann, B., Schürenkamp, M., Pfeiffer, H., Neuhuber, F., & Brinkmann, B. (2015). Elevated germline mutation rate in teenage fathers. Proceedings. Biological sciences, 282(1803), 20142898. https://doi.org/10.1098/rspb.2014.2898

Freier, Luisa Feline et al. (2019). "La emigración venezolana: respuestas latinoamericanas." Documentos de trabajo (Fundación Carolina): Segunda época **3**

García, William and Gravel, François. "Native Recognition and Public Education in Colombia in the Context of Latin American Countries: a Historical Approach"

Gayà-Vidal, Magdalena, et al. "mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations." American Journal of Physical Anthropology 145.2 (2011): 215-230.

Gedicks, A. (2003). "Resource wars against Native peoples in Colombia." Capitalism Nature Socialism **14**(2): 85-111.

Geppert, Maria, et al. "Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America." Forensic Science International: Genetics 5.2 (2011): 100-104.

Gitelman, HM (1978). "Amerikanuak: Basques in the New World. By William A. Douglass and Jon Bilbao. University of Nevada Press: Reno, Nevada, 1975. Pp. 519. - By the Sweat of Their Brow: Mexican Immigrant Labor in the United States, 1900–1940. By Mark Reisler. Westport, Conn.: Greenwood Press, 1976. Pp. xi, 298. $14.50." The Journal of Economic History **38**(2): 533-535.

Gomes, V., Sánchez-Diz, P., Amorim, A., Carracedo, Á., & Gusmão, L. (2010). Digging deeper into East African human Y chromosome lineages. Human genetics, 127(5), 603-613.

Gómez-Carballa, Alberto, et al. "The peopling of South America and the trans-Andean gene flow of the first settlers." Genome research 28.6 (2018): 767-779.

González, M., et al. "The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88." European Journal of Human Genetics 21.3 (2013): 324-331.

Goodwin, W., Linacre, A., & Hadi, S. (2011). An introduction to forensic genetics. John Wiley & Sons.

Greenberg, J. H. (1987). Language in the Americas. Stanford University Press.

Grugni, Viola, et al. "Exploring the Y chromosomal ancestry of modern Panamanians." PLoS One 10.12 (2015): e0144223.

Grugni, Viola, et al. "Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas." BMC biology 17.1 (2019): 1-14.

Guo, Xihan et al. (2020). "Mosaic loss of human Y chromosome: what, how and why." Human Genetics **139**(4): pp. 421-446.

Gusmão, L., et al. "International Society of Forensic Genetics. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis." Int. J. Legal Med 120.4 (2006): 191-200.

Haber, M., et al. "Influences of history, geography, and religion on genetic structure: the Maronites in Lebanon." European Journal of Human Genetics 19.3 (2011): 334-340.

Haber, M., et al. "Chad genetic diversity reveals an African history marked by multiple Holocene Eurasian migrations." The American Journal of Human Genetics 99.6 (2016): 1316-1324.

Hassan, H. Y., Underhill, P. A., Cavalli-Sforza, L. L., & Ibrahim, M. E. (2008). Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography, and history. American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 137(3), 316-323.

Hernández, M. G. (2019). Motilones: de indios «mansos» o «bravos» a Yukpas y Barís (1910-1960). Boletín americanista, 71-90.

Herrera, Kristian J., et al. "Neolithic patrilineal signals indicate that the Armenian plateau was repopulated by agriculturalists." European journal of human genetics 20.3 (2012): 313-320.

Hinch, A. G., Altemose, N., Noor, N., Donnelly, P., & Myers, S. R. (2014). Recombination in the human pseudoautosomal region PAR1. PLoS genetics, 10(7), e1004503.

Ibarra, Adriana, et al. "Comparison of the genetic background of different Colombian populations using the SNPforID 52plex identification panel." International journal of legal medicine 128.1 (2014): 19-25.

Jannuzzi, J., et al. "Male lineages in Brazilian populations and performance of haplogroup prediction tools." Forensic Science International: Genetics 44 (2020): 102163.

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

Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. Nature Reviews Genetics, 5(10), 739-751.

Karachanak S, Grugni V, Fornarino S, Nesheva D, Al-Zahery N, et al. 2013. Y-chromosome diversity in modern Bulgarians: new clues about their ancestry. PLoS One 8:e56779

Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., & Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome research, 18(5), 830-838.

Kayser, Manfred, et al. "Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea." The American Journal of Human Genetics 72.2 (2003): 281-302.

Kim, Soon-Hee, et al. "High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea." Investigative genetics 2.1 (2011): 1-11.

Kimura, M., & Ohta, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. Proceedings of the National Academy of Sciences, 75(6), 2868-2872.

King, R and Zontini, E (2000). "The role of gender in the South European immigration model." Papers: revista de sociología, P 35-52.

Krausz, Csilla and Degl'Innocenti, Selene (2006). " Y chromosome and male infertility: update, 2006." Front Biosci **11**(1): pp 3049 - 3061.

Lancaster, A. (2009). Y haplogroups, archaeological cultures and language families: a review of the possibility of multidisciplinary comparisons using the case of haplogroup E-M35. J Genet Geneal, 5, 35-65.

Lander, Eric S et al. (2001) "Initial sequencing and analysis of the human genome."

Lang, Min, et al. "Forensic characteristics and genetic analysis of both 27 Y-STRs and 143 Y-SNPs in Eastern Han Chinese population." Forensic Science International: Genetics 42 (2019): e13-e20.

Larmuseau, M. H., Van Geystelen, A., Kayser, M., van Oven, M., & Decorte, R. (2015). Towards a consensus Y-chromosomal phylogeny and Y-SNP set in forensics in the next-generation sequencing era. Forensic Science International: Genetics, 15, 39-42.

LaRosa, M. J., & Mejía, G. R. (2017). Colombia: A concise contemporary history. Rowman & Littlefield.

Layrisse, Z., et al. (2001). "Extended HLA haplotypes in a Carib Amerindian population: the Yucpa of the Perija Range." Human immunology **62**(9): 992-1000.

Lebow, Jeremy, et al. (2021). "Immigration and Trust: The Case of Venezuelans in Colombia."

Lee, Eun Young, et al. (2014). "Analysis of 22 Y chromosomal STR haplotypes and Y haplogroup distribution in Pathans of Pakistan." Forensic Science International: Genetics **11**: 111-116.

Leigh, E. G., et al. (2014). "Historical biogeography of the I sthmus of P anama." Biological Reviews **89**(1): 148-172.

Leite, Gonçalo Aragão da Fonseca Pinto. "Characterization of male lineages in the Ashaninka from Peru." (2018).

Li, Gongcheng, et al. "Regulatory variants and disease: the e-cadherin− 160C/A SNP as an example." Molecular biology international 2014 (2014).

Lizarralde, M., & Lizarralde, R. (2015). Los Barí. Los aborígenes de Venezuela, 5, 725-886.

López-Parra, A. M., et al. "In search of the pre-and post-neolithic genetic substrates in Iberia: evidence from Y-chromosome in Pyrenean populations." Annals of human genetics 73.1 (2009): 42-53.

Lu, Yi-Fan et al. “IFNL3 mRNA structure is remodeled by a functional non-coding polymorphism associated with hepatitis C virus clearance.” Scientific reports vol. 5 16037. 4 Nov. 2015

Maan, Akhlaq A., et al. "The Y chromosome: a blueprint for men’s health?." European Journal of Human Genetics 25.11 (2017): 1181-1188.

Malaspinas, A.S., et al. (2014). "Two ancient human genomes reveal Polynesian ancestry among the indigenous Botocudos of Brazil." Current Biology **24**(21): R1035-R1037.

Malyarchuk, Boris, et al. "Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a." Journal of human genetics 56.8 (2011): 583-588.

Marshall, L. G., et al. (1982). "Mammalian evolution and the great American interchange." science **215**(4538): 1351-1357.

Martínez-Cruz, B., Ziegle, J., Sanz, P., Sotelo, G., Anglada, R., Plaza, S., and Comas, D. (2011). Multiplex single-nucleotide polymorphism typing of the human Y chromosome using TaqMan probes. Investigative Genetics, 2(1), 1-7.

Martinez-Laso, J., et al. (1995). "A new HLA-B15 allele (B* 1522) found in Barí-Motilones Amerindians in Venezuela: comparison of its intron 2 sequence with those of B* 1501 and B* 3504." Immunogenetics **43**(1): 108-109.

Martinez-Martin, Abel Fernando et al. (2017). " Trópico y raza. Miguel Jiménez López y la inmigración japonesa en Colombia, 1920-1929." Historia y Sociedad: 103-138.

Matisoo-Smith, E. and J.M. Ramirez (2010). "Human skeletal evidence of Polynesian presence in South America? Metric analyses of six crania from Mocha Island, Chile." Journal of Pacific Archaeology–Vol **1**(1).

Mazières, S., Sevin, A., Callegari-Jacques, S. M., Crubézy, E., Larrouy, G., Dugoujon, J. M., & Salzano, F. M. (2009). Population genetic dynamics in the French Guiana region. American Journal of Human Biology: The Official Journal of the Human Biology Association, 21(1), 113-117.

Meltzer, D. J. (2013). "8 The human colonization of the Americas: archaeology." The encyclopedia of global human migration.

Mendez, Fernando L., et al. "Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa." Human biology 83.1 (2011): 39-53.

Merbs, C. F. (1992). "A new world of infectious disease." American Journal of Physical Anthropology **35**(S15): 3-42.

Wierdl, M., Dominska, M., & Petes, T. D. (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics, 146(3), 769-779.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986, January). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In Cold Spring Harbor symposia on quantitative biology (Vol. 51, pp. 263-273). Cold Spring Harbor Laboratory Press.

Muzzio, Marina et al. (2011). " Software for Y-haplogroup predictions: a word of caution." International journal of legal medicine **125**(1): 143-147.

Myres, Natalie M., et al. "Y-chromosome short tandem repeat DYS458. 2 non-consensus alleles occur independently in both binary haplogroups J1-M267 and R1b3-M405." Croatian Medical Journal 48.4 (2007): 450.

Myres, Natalie M., et al. "A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe." European Journal of Human Genetics 19.1 (2011): 95-101.

Nilsson, M., Andréasson-Jansson, H., Ingman, M., & Allen, M. (2008). Evaluation of mitochondrial DNA coding region assays for increased discrimination in forensic analysis. Forensic Science International: Genetics, 2(1), 1-8.

Nogueiro, I., Manco, L., Gomes, V., Amorim, A., & Gusmão, L. (2010). Phylogeographic analysis of paternal lineages in NE Portuguese Jewish communities. American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists, 141(3), 373-381.

Noguera, M. C., et al. "Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar." Annals of human biology 41.5 (2014): 453-459.

Núñez, Carolina, et al. "Y chromosome haplogroup diversity in a Mestizo population of Nicaragua." Forensic Science International: Genetics 6.6 (2012): e192-e195.

O'Dea, A., et al. (2016). "Formation of the Isthmus of Panama." Science advances **2**(8): e1600883.

Ossa, Humberto, et al. (2015). "Analysis of admixture in Native American populations from Colombia." Forensic Science International: Genetics Supplement Series **5**: e332-e334.

Ossa, Humberto, et al. (2016). "Outlining the ancestry landscape of Colombian admixed populations." PLoS One **11**(10): e0164414.

Ossa, Humberto, et al. "Patterns of genetic diversity in Colombia for 38 indels used in human identification." Forensic Science International: Genetics 53 (2021): 102495.

Pastor, J.M.A. (2004). "Possible Paradises: Basque Emigration to Latin America."

Pearson, Christopher E. et al. (2005) " Repeat instability: mechanisms of dynamic mutations." Nature Reviews Genetics **6**(10): 729-742.

Pereira, V., & Gusmão, L. (2016). Types of Genomes, Sequences and Genetic Markers (Repeats, SNPs, Indels, Haplotypes). Handbook Of Forensic Genetics: Biodiversity And Heredity In Civil And Criminal Investigation, 2, 163.

Petit, E., Balloux, F., & Excoffier, L. (2002). Mammalian population genetics: why not Y?. Trends in Ecology & Evolution, 17(1), 28-33.

Puerto, H. (2018). A New Hope: Perspectives on Implementing the Pilot of a New Comprehensive Health Care Model in Guainía, Colombia.

Purps, J., Siegert, S., Willuweit, S., Nagy, M., Alves, C., Salazar, R., Angustia, S. M., Santos, L.H., Anslinger, K., Bayer, B., Ayub, Q., Wei, W., Xue, Y., Tyler-Smith, C., Bafalluy, M. B., Martínez-Jarreta, B., Egyed, B., Balitzki, B., Tschumi, S., Ballard, D., Court, D. S., Barrantes, X., Bäßler, G., Wiest, T., Berger, B., Niederstätter, H., Parson, W., Davis, C., Budowle, B., Burri, H., Borer, U., Koller, C., Carvalho, E. F., Domingues, P. M., Chamoun, W. T., Coble, M.

D., Hill, C. R., Corach, D., Caputo, M., D'Amato, M. E., Davison, S., Decorte, R., Larmuseau, M. H., Ottoni, C., Rickards, O., Lu, D., Jiang, C., Dobosz, T., Jonkisz, A., Frank, W. E., Furac, I., Gehrig, C., Castella, V., Grskovic, B., Haas, C., Wobst, J., Hadzic, G., Drobnic, K., Honda, K., Hou, Y., Zhou, D., Li, Y., Hu, S., Chen, S., Immel, U. D., Lessig, R., Jakovski, Z., Ilievska, T., Klann, A. E., García, C. C., de Knijff, P., Kraaijenbrink, T., Kondili, A., Miniati, P., Vouropoulou, M., Kovacevic, L., Marjanovic, D., Lindner, I., Mansour, I., Al-Azem, M., Andari, A. E., Marino, M., Furfuro, S., Locarno, L., Martín, P., Luque, G. M., Alonso, A., Miranda, L. S., Moreira, H., Mizuno, N., Iwashima, Y., Neto, R. S., Nogueira, T. L., Silva, R., Nastainczyk-Wulf, M., Edelmann, J., Kohl, M., Nie, S., Wang, X., Cheng, B., Núñez, C., Pancorbo, M. M., Olofsson, J. K., Morling, N., Onofri, V., Tagliabracci, A., Pamjav, H., Volgyi, A., Barany, G., Pawlowski, R., Maciejewska, A., Pelotti, S., Pepinski, W., Abreu-Glowacka, M., Phillips, C., Cárdenas, J., Rey-Gonzalez, D., Salas, A., Brisighelli, F., Capelli, C., Toscanini, U., Piccinini, A., Piglionica, M., Baldassarra, S. L., Ploski, R., Konarzewska, M., Jastrzebska, E., Robino, C., Sajantila, A., Palo, J. U., Guevara, E., Salvador, J., Ungria, M. C., Rodriguez, J. J., Schmidt, U., Schlauderer, N., Saukko, P., Schneider, P. M., Sirker, M., Shin, K. J., Oh, Y. N., Skitsa, I., Ampati, A., Smith, T. G., Calvit, L. S., Stenzl, V., Capal, T., Tillmar, A., Nilsson, H., Turrina, S., De Leo, D., Verzeletti, A., Cortellini, V., Wetton, J. H., Gwynne, G. M., Jobling, M. A., Whittle, M. R., Sumita, D. R., Wolańska-Nowak, P., Yong, R. Y., Krawczak, M., Nothnagel, M., & Roewer, L. (2014). A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. Forensic Science International: Genetics, 12(100), 12–23.

Qamar, R., Ayub, Q., Mohyuddin, A., Helgason, A., Mazhar, K., Mansoor, A., Zerjal, T., Tyler-Smith, C. & Mehdi, S. Q. (2002). Y-chromosomal DNA variation in Pakistan. The American Journal of Human Genetics, 70(5), 1107-1124.

Quesada, J. D. (2004). The Barí language of Venezuela: a glimpse at the Chibchan periphery. STUF-Language Typology and Universals, 57(4), 362-376.

Raghavan, Maanasa, et al. "Genomic evidence for the Pleistocene and recent population history of Native Americans." Science 349.6250 (2015).

Rappaport, J. (1990). The politics of memory: Native historical interpretation in the Colombian Andes (Vol. 70). CUP Archive.

Rapone, C., D'Atanasio, E., Agostino, A., Mariano, M., Papaluca, M. T., Cruciani, F., & Berti, A. (2016). Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler® Plus kit) in an Italian population sample. Forensic Science International: Genetics, 21, e1-e5.

Raudsepp, Terje and Chowdhary, Bhanu P (2015). "The eutherian pseudoautosomal region." Cytogenetic and Genome Research. **147**(2-3): pp. 81-94.

Rausch, Jane M. (2021). " Germans in the History of Colombia from Colonial Times to the Present." Xlibris Corporation.

Rębała, K., Martínez-Cruz, B., Tönjes, A., Kovacs, P., Stumvoll, M., Lindner, I., Büttner, A., Wichmann, H. E., Siváková, D., Soták, M., Quintana-Murci, L., Szczerkowska, Z., Comas, D., & Genographic Consortium (2013). Contemporary paternal genetic landscape of Polish and German populations: from early medieval Slavic expansion to post-World War II resettlements. European Journal of Human Genetics: EJHG, 21(4), 415–422.

Redd, A. J. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups.

Resque, R., Gusmão, L., Geppert, M., Roewer, L., Palha, T., Alvarez, L., Ribeiro-dos-Santos, Â., & Santos, S. (2016). Male lineages in Brazil: intercontinental admixture and stratification of the European background. PLoS ONE, 11(4), e0152573.

Rey-González D, Gelabert-Besada M, Cruz R, Brisighelli F, Lopez-Soto M, et al. 2017. Micro and macro geographical analysis of Y-chromosome lineages in South Iberia. Forensic Science International: Genetics 29:e9-e15

Ribeiro, B. P. A. (2021). Genetic characterization of the maternal lineages in Andean Col ombian populations.

Rice, W. R. (1996). "Evolution of the Y Sex Chromosome in Animals." BioScience **46**(5): 331-343.

Robert, F., & Pelletier, J. (2018). Exploring the impact of single-nucleotide polymorphisms on translation. Frontiers in genetics, 507.

Robino, Carlo, et al. Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. International Journal of Legal Medicine, 2008, 122.3: 251-255.

Roca-Rada, X., et al. (2021). "Ancient mitochondrial genomes from the Argentinian Pampas inform the early peopling of the Southern Cone of South America." Iscience **24**(6): 102553.

Roewer, Lutz, et al. "Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans." PLoS genetics 9.4 (2013): e1003460.

Roewer, L. (2019). Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA. Wiley Interdisciplinary Reviews: Forensic Science, 1(4), e1336.

Roewer, Lutz, et al. "DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations on the interpretation of Y-STR results in forensic analysis." Forensic Science International: Genetics 48 (2020): 102308.

Rojas, W., et al. (2010). "Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers." American Journal of Physical Anthropology **143**(1): 13-20.

Rootsi, Siiri, et al. "Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus." European Journal of Human Genetics 20.12 (2012): 1275-1282.

Rosa, Alexandra, et al. "Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective." BMC Evolutionary Biology 7.1 (2007): 1-11.

Ross, M. T., et al. (2005). "The DNA sequence of the human X chromosome." Nature **434**(7031): 325-337.

Rowold, D. J., Perez-Benedico, D., Stojkovic, O., Garcia-Bertrand, R., & Herrera, R. J. (2016). On the Bantu expansion. Gene, 593(1), 48-57.

Sahakyan, Hovhannes, et al. "Origin and diffusion of human Y chromosome haplogroup J1-M267." Scientific reports 11.1 (2021): 1-14.

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: The Official Journal of the Human Biology Association, 20(5), 584-591.

Salzano, F. M., & Sans, M. (2014). Interethnic admixture and the evolution of Latin American populations. Genetics and molecular biology, 37, 151-170.

Sandoval, Karla, et al. "Y-chromosome diversity in Native Mexicans reveals continental transition of genetic structure in the Americas." American journal of physical anthropology 148.3 (2012): 395-405.

Schwartz, A., et al. (1998). "Reconstructing hominid Y evolution: X-homologous block, created by X–Y transposition, was disrupted by Yp inversion through LINE—LINE recombination." Human Molecular Genetics **7**(1): 1-11.

Scozzari, Rosaria, et al. "Molecular dissection of the basal clades in the human Y chromosome phylogenetic tree." PLoS One 7.11 (2012): e49170.

Semino, Ornella, et al. "Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area." The American Journal of Human Genetics 74.5 (2004): 1023-1034.

Shabir, Ghulam, et al. "Rice molecular markers and genetic mapping: Current status and prospects." Journal of integrative agriculture 16.9 (2017): 1879-1891.

Singh Malhi, Ripan, et al. "Distribution of Y chromosomes among native North Americans: a study of Athapaskan population history." American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists 137.4 (2008): 412-424.

Solé-Morata N, Bertranpetit J, Comas D, Calafell F. 2015. Y-chromosome diversity in Catalan surname samples: insights into surname origin and frequency. European Journal of Human Genetics 23:1549-57

Staub, N. L. (2002). Teaching evolutionary mechanisms: genetic drift and M&M's®. BioScience, 52(4), 373-377.

Sutter, R. C. (2021). The pre-Columbian peopling and population dispersals of South America. Journal of Archaeological Research, 29(1), 93-151.

Syndercombe Court, D. (2021). The Y chromosome and its use in forensic DNA analysis. Emerging topics in life sciences, 5(3), 427-441.

Teama, Salwa. "DNA polymorphisms: DNA-based molecular markers and their application in medicine." Genetic Diversity and Disease Susceptibility 25 (2018).

Templeton, A. R. (2021). Population genetics and microevolutionary theory. John Wiley & Sons.

Tofanelli, S., Bertoncini, S., Castrì, L., Luiselli, D., Calafell, F., Donati, G., & Paoli, G. (2009). On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. Molecular biology and evolution, 26(9), 2109-2124.

Toscanini, U., Gusmao, L., Berardi, G., Amorim, A., Carracedo, A., Salas, A., & Raimondi, E. (2008). Y chromosome microsatellite genetic variation in two Native American populations from Argentina: population stratification and mutation data. Forensic Science International: Genetics, 2(4), 274-280.

Toscanini, U., et al. "Analysis of Y-chromosome STRs in Chile confirms an extensive introgression of European male lineages in urban populations." Forensic Science International: Genetics 21 (2016): 76-80.

Toscanini, U., et al. "The geographic mosaic of Ecuadorian Y-chromosome ancestry." Forensic Science International: Genetics 33 (2018): 59-65.

Trombetta, Beniamino, et al. "Phylogeographic refinement and large scale genotyping of human Y chromosome haplogroup E provide new insights into the dispersal of early pastoralists in the African continent." Genome biology and evolution 7.7 (2015): 1940-1950.

Turrina, S., Caratti, S., Ferrian, M., & De Leo, D. (2015). Haplotype data and mutation rates for the 23 Y-STR loci of PowerPlex® Y 23 System in a Northeast Italian population sample. International journal of legal medicine, 129(4), 725-728.

Tvedebrink, Torben. "Review of the Forensic Applicability of Biostatistical Methods for Inferring Ancestry from Autosomal Genetic Markers." Genes 13.1 (2022): 141.

Underhill, Peter A., et al. "The phylogenetic and geographic structure of Y-chromosome haplogroup R1a." European Journal of Human Genetics 23.1 (2015): 124-131.

Van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R., & Larmuseau, M. H. (2014). Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. Human mutation, 35(2), 187-191.

Vélez-Henao, Johan-Andrés et al. (2020). "Environmental rebound effect of energy efficiency improvements in Colombian households." Energy Policy **145**: pp 111697

Vilar, M. G., et al. "Genetic diversity in Puerto Rico and its implications for the peopling of the Island and the West Indies." American Journal of Physical Anthropology 155.3 (2014): 352-368.

Wang, S., et al. (2007). "Genetic variation and population structure in Native Americans." PLoS genetics **3**(11): e185.

Wang, S., Ray, N., Rojas, W., Parra, M. V., Bedoya, G., Gallo, C., Poletti, G., Mazzotti, G., Hill, K., and Ruiz-Linares, A. (2008). Geographic patterns of genome admixture in Latin American Mestizos. PLoS genetics, 4(3), e1000037.

Weinstein, Asher (2021). "Developing Density: Diverging Approaches to Urbanization between South Africa and Colombia." Global Majority E-Journal: p. 60

Xavier, C., et al. (2015). "Admixture and genetic diversity distribution patterns of non-recombining lineages of Native American ancestry in Colombian populations." PLoS One **10**(3): e0120155.

Xue, Yali, et al. "Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree." Current Biology 19.17 (2009): 1453-1457.

Y Chromosome Consortium. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome research, 12(2), 339-348.

Yue, P., & Moult, J. (2006). Identification and analysis of deleterious human SNPs. Journal of molecular biology, 356(5), 1263-1274.

Zalloua, P. A., Xue, Y., Khalife, J., Makhoul, N., Debiane, L., Platt, D. E., Royyuru, A. K., Herrera, R. J., Hernanz, D. F., Blue-Smith, J., Wells, R. S., Comas, D., Bertranpetit, J., Tyler-Smith, C., & Genographic Consortium (2008). Y-chromosomal diversity in Lebanon is structured by recent historical events. American Journal of Human Genetics, 82(4), 873–882.

Zegura, Stephen L., et al. "High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas." Molecular biology and evolution 21.1 (2004): 164-175.

Zwisler, J. J. (2018). "The Pijao of Natagaima: Post-Linguicide Indigenous Identity and Language." Cadernos de Ethnolinguistica **5**(1): 52-82.

7. Appendix

Table 12: Y-STR haplotypes and their corresponding haplogroups (Haplogroup Genotyped) from the Guainía Indigenous. The haplogroup assignment obtained by NEVGEN and Haplogroup Predictor (HP) are also represented.

Sample ID	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385 a/b	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	HP	NEVGEN	Haplogroup Genotyped
844	18	14	20	31	19	11	17	17	10	21	10	15	21	11	11	38	17	14	18,18	31	15	12	27	38,38	11	E1b1a	E1b1a *V38>> L485*	E-U174
848	20	12	20	28	22	11	16.2	14	12	20	10	15	25	10	11	38	16	14	14,18	27	12	11	26	36,38	11	J1	J2a1 *Z6046*	J-12f2a*
850	19	13	21	30	21	12	15	14	13	21	10	12	22	10	11	39	22	14	13,13	35	14	11	26	35,37	12	E1b1b	E1b1b *V257> PF2431*	E-M35*
851	17	13	22	30	21	11	19	13	11	21	10	16	24	11	13	38	17	14	14,19	29	13	13	25	35,38	12	Q	Q *M346>> M3> M902*	Q-M3*
855	18	12	22	30	21	10	17	13	12	19	10	17	24	12	15	39	19	14	14,16	28	13	11	22	36,39	11	Q	Q *M346>> M3> M902*	Q-M3*
856	18	12	22	30	22	10	16	13	12	19	10	17	23	12	15	39	19	14	14,15	28	13	11	22	36,39	11	Q	Q *M346>> Y4800> F835> L932 (possibly BZ4800(439=11,389ii=18))*	Q-M3*
857	17	13	22	30	20	10	19	13	11	21	10	15	24	11	14	37	17	14	14,18	29	13	12	25	35,38	12	Q	Q *M346>> M3> M902*	Q-M3*
860	19	13	22	30	21	11	19	13	11	21	10	15	24	11	14	39	17	14	14,18	30	13	13	26	35,38	12	Q	Q *M346>> M3> M902*	Q-M3*
861	18	12	22	30	22	10	16	13	12	19	10	17	23	12	15	39	19	14	14,15	28	13	11	22	36,39	11	Q	Q *M346>> Y4800> F835> L932*	Q-M3*
874	18	13	22	32	22	10	16	15	11	20	10	15	24	11	14	39	17	14	14,14	29	13	13	24	35,39	12	Q	Q *M346>> M3> M902*	Q-M3*
887	17	13	22	30	21	11	19	13	11	21	10	16	24	11	13	38	17	14	14,19	30	13	13	25	35,38	12	Q	Q *M346>> M3> M902*	Q-M3*
889	18	13	23	29	22	12	19	14	12	19	11	15	24	12	13	38	17	14	11,14	30	13	12	22	35,36	11	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	-
893	17	13	22	31	21	11	20	13	11	21	10	16	24	11	13	38	17	14	15,19	30	13	13	25	35,38	12	Q	Q M346>> M3	-
902	18	12	22	30	21	10	17	13	12	19	10	17	24	12	15	39	19	14	14,16	28	13	11	22	36,39	11	Q	Q *M346>> M3> M902*	Q-M3*
909	17	13	22	30	21	11	19	13	11	21	10	16	24	11	13	38	17	14	14,19	30	13	13	25	35,38	12	Q	Q *M346>> M3> M902*	-
916	17	13	23	30	21	11	20	13	11	21	10	15	24	11	14	39	17	14	14,18	30	13	13	25	35,38	12	Q	Q *M346>> M3> M902*	Q-M3*
925	16	13	21	28	20	10	20	13	12	20	10	14	25	9	11	39	18	16	13,18	28.2	13	11	24	37,37	11	J1	J2a1 *Z6046*	J-M172
927	18	12	22	30	21	10	17	13	12	19	10	17	24	12	15	39	19	14	14,16	28	13	11	22	36,39	11	Q	Q *M346>> M3> M902*	Q-M3*

Table 13: Y-STR haplotypes and their corresponding haplogroups (Haplogroup Genotyped) from the Barí-Motion Community. The haplogroup assignment obtained by NEVGEN and Haplogroup Predictor (HP) are also represented.

Sample ID	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385 a/b	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	HP	NEVGEN	Haplogroup Genotyped
952	16	12	22	29	19	10	17	13	12	20	10	15	24	11	14	39	18	14	15,16	29	12	12	24	36,39	11	Q	Q M346>> M3> Y4276	Q-M3*
972	19	14	22	32	23	9	18	13	11	20	10	15	24	11	13	39	19	14	13,19	31	14	13	24	36,38	12	E1b1b	Q M346>> M3	-
1006	19	14	22	33	23	9	18	13	11	20	10	15	24	11	13	39	19	14	13,19	31	14	13	24	36,38	12	E1b1b	Q M346>> M3> M902	-
1007	16	12	22	29	19	10	17	13	12	20	10	15	24	11	14	39	18	14	15,16	29	12	12	24	36,39	11	Q	Q M346>> M3> Y4276	Q-M3*
1010	18	13	22	32	21	10	16	13	11	19	10	15	24	11	14	39	19	14	14,16	30	14	12	26	36,37	11	Q	Q M346>> Z780	Q-M3*
1011	16	12	22	29	19	10	17	13	12	20	10	15	24	11	14	39	18	14	14,16	29	12	12	24	36,39	11	Q	Q M346>> M3> M902	-
1013	19	14	22	33	23	9	18	14	11	20	10	15	24	11	13	39	19	14	13,19	30	14	13	24	36,38	12	Q	Q M346>> M3> M902	-
1022	16	12	22	29	19	10	17	13	12	20	10	15	24	11	14	39	18	14	15,16	29	12	12	24	36,39	11	Q	Q M346>> M3> Y4276	Q-M3*
1027	18	13	22	32	21	10	17	13	11	19	10	15	24	11	14	38	19	14	14,16	30	14	12	26	36,37	11	Q	Q M346>> Z780	-
1031	16	12	22	29	19	10	17	13	12	20	10	15	24	11	14	39	18	14	14,16	29	12	12	24	36,39	11	Q	Q M346>> M3> M902	-
1037	19	14	22	33	23	9	18	13	11	20	10	15	24	11	13	39	19	14	13,19	31	14	13	24	36,38	12	E1b1b	Q M346>> M3> M902	Q-M3*
1040	19	14	22	32	22	9	18	13	11	20	10	15	24	11	13	39	19	14	13,19	31	14	13	24	36,38	12	E1b1b	Q M346>> M3> M902	
1043	19	14	22	33	24	9	18	13	11	20	10	15	24	11	13	39	19	14	13,19	31	14	13	24	36,38	12	E1b1b	Q M346>> M3> M902	-
1044	16	12	22	30	19	10	17	13	12	20	10	15	24	11	14	39	18	14	14,16	29	12	12	24	36,39	11	Q	Q M346>> M3> M902	-
1046	19	14	22	33	24	9	18	13	11	20	10	15	24	11	13	39	19	14	13,19	31	14	13	24	36,36	12	Q	Q M346>> M3> M902	-
1049	19	14	22	33	22	9	18	13	11	20	10	15	24	11	13	39	19	14	13,19	31	14	13	24	36,38	12	E1b1b	Q M346>> M3> M902	-
1050	19	14	22	33	23	9	18	13	11	20	10	15	24	11	13	39	19	14	13,19	30	14	13	24	36,38	12	Q	Q M346>> M3> M902	Q-M3*
1051	16	12	22	30	19	10	17	13	12	20	10	15	24	11	14	39	18	14	14,16	29	12	12	24	37,39	11	Q	Q M346>> M3> M902	-
1052	17	12	22	29	19	10	17	13	12	20	10	15	24	11	14	39	18	14	15,16	29	12	12	24	36,39	11	Q	Q M346>> M3> M902	-
1056	19	13	22	31	23	9	18	13	11	20	10	15	24	11	13	39	19	14	13,19	31	14	13	24	36,38	12	E1b1b	Q M346>> M3	-
1075	16	12	22	29	19	10	17	13	12	20	10	15	24	11	14	39	18	14	15,16	29	12	12	24	36,39	11	Q	Q M346>> M3> Y4276	-

Table 14: Y-STR haplotypes and their corresponding haplogroups (Haplogroup Genotyped) from the Paeces Indigenous. The haplogroup assignment obtained by NEVGEN and Haplogroup Predictor (HP) are also represented.

Sample ID	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385 a/b	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	HP	NEVGEN	Haplogroup Genotyped
1083	19	13	24	29	22	11	17	14	12	18	11	16	24	12	13	41	17	15	11,14	32	13	11	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-M529
1086	18	13	22	29	22	12	15	13	12	19	6	16	23	11	14	37	15	14	15,17	28	13	11	24	35,39	10	Q	Q *M346>> Z780*	Q-M346*
1087	18	13	23	31	18	11	17	16	13	0	11	15	25	11	11	43	19	15	11,14	29	13	11	23	36,39	12	R1a	R1a *(for 67+ markers, try level for R1a-s, 70+ subclades)*	R-SRY10831.2
1089	18	13	23	31	19	11	17	13	13	22	11	15	23	11	14	39	19	15	14,19	30	12	12	24	35,40	11	Q	Q *M346>> M3> Y4276*	Q-M3*
1090	18	13	23	31	21	11	17	14	12	19	11	15	24	12	13	38	17	16	11,14	29	13	12	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	-
1112	19	14	23	30	19	10	17	14	11	18	10	16	24	12	13	38	17	14	11,15	30	13	13	22	36,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
1114	18	13	23	30	21	11	16	14	13	19	10	15	24	12	13	38	17	15	11,14	29	13	12	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-M529
1116	18	13	22	29	22	12	15	13	12	19	6	16	23	11	14	37	15	14	15,17	28	13	11	24	35,39	10	Q	Q *M346>> Z780*	-
1137	18	13	23	30	23	11	17	14	13	19	11	16	24	12	13	41	17	15	11,14	32	13	12	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
1140	18	14	22	30	22	12	15	13	12	19	6	16	23	11	14	37	16	14	14,17	28	13	11	24	35,39	10	Q	Q *M346>> Z780*	Q-M346*
1148	19	13	23	30	21	11	16	14	13	19	10	15	24	12	13	38	17	15	11,14	29	13	12	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-M529
1149	15	13	20	29	19	11	15	14	11	19	10	15	23	9	13	37	18	14	14,16	34	13	10	22	37,38	12	E1b1b	T >> *CTS11451> Y4119>> Y3781*	-

Table 15: Y-STR haplotypes and their corresponding haplogroups (Haplogroup Genotyped) from the Cundinamarca administrative region. The haplogroup assignment obtained by NEVGEN and Haplogroup Predictor (HP) are also represented.

Sample ID	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385 a/b	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	HP	NEVGEN	Haplogroup Genotyped
97	15	13	21	31	20	10	16	15	12	21	10	15	21	12	11	37	18	14	16,17	28	13	12	29	36,39	11	E1b1a	E1b1a *V38>> M4231*	E-U290
105	18	13	21	29	22	11	18.2	14	11	20	10	15	23	10	11	37	17	14	14,17	25	12	12	25	36,36	11	J1	J1a2a1a2 *P58*	J-12f2a*
108	17	13	23	29	21	11	16	14	12	19	11	15	24	12	13	38	18	14	11,12	29	13	12	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
110	15	13	23	30	21	11	16	14	11	19	10	15	24	12	13	38	19	14	11,14	28	13	12	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-U152
116	18	13	23	29	21	11	17	14	10	19	10	16	23	12	13	37	17	15	11,14	29	13	11	23	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-U106
117	18	13	23	29	22	11	16	15	12	20	11	16	24	12	13	36	17	14	11,14	30	13	11	21	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-U152
119	18	14	23	30	23	10	17	14	11	18	10	17	24	12	13	38	17	14	11,14	31	13	12	23	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
127	14	13	23	30	21	11	17	15	10	20	10	15	23	9	13	35	16	14	12,16	35	14	11	23	35,38	13	E1b1a	T *>> CTS11451> Y4119>> Y3781*	T-M70
128	18	12	22	29	22	10	14	13	11	20	10	15	25	10	11	41	20	14	18,19	32	13	11	23	36,37	12	E1b1b	E1b1b *L67*	E-M78
130	17	13	21	30	18	11	18.2	14	11	20	10	16	23	10	11	35	20	14	13,18	26	12	11	25	39,39	11	J1	J1a2a1a2 *P58*	J-12f2a*
138	18	13	23	29	23	12	17	14	13	19	10	15	24	12	13	38	17	15	12,14	30	14	8	21	34,36	11	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
139	17	13	21	29	22	10	14	15	12	21	9	16	23	9	11	39	18	14	13,16	30	12	12	22	37,39	12	H	J2a1 *Z387*	J-M172
141	18	13	23	31	16	11	15	15	13	20	10	14	24	11	11	42	20	14	11,14	31	13	10	23	36,38	12	R1a	R1a *(for 67+ markers, try level for R1a-s, 70+ subclades)*	R-SRY10831.2
145	18	13	22	31	20	10	16	13	12	19	10	16	23	12	15	39	18	14	15,15	28	13	11	24	35,39	11	Q	Q *M346>> M3> M902*	Q-M3*
147	18	12	23	28	20	10	17	13	13	19	10	15	23	13	15	39	18	14	15,15	29	13	11	23	35,41	12	Q	Q *M346>> M3> M902*	Q-M3*
148	17	12	21	29	20	10	18	15	11	21	10	14	22	10	10	38	17	16	15,15	27	14	12	23	37,38	10	G2a	G2a1 *Z6552*	G-M201
151	17	13	19	31	22	10	18	15	11	22	10	14	21	10	11	39	16	16	11,13	33	14	13	26	36,36	10	G2a	A1b1b2b *M13*	A-M13
157	18	12	22	29	21	10	14	13	11	20	10	15	25	10	11	41	21	14	18,19	33	13	11	23	36,37	12	E1b1b	E1b1b *L67*	E-M78
158	18	13	23	29	22	11	16	15	12	20	11	16	24	12	13	36	17	14	11,14	30	13	11	21	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-U152
159	17	13	21	30	21	10	15	14	11	20	10	15	22	9	11	38	16	15	14,15	32	12	11	23	37,40	11	L	J2a1 *L26>Z500*	J-M172

Table 15: Y- STR haplotypes and their corresponding haplogroups (Haplogroup Genotyped) from the Huila administrative region. The haplogroup assignment obtained by NEVGEN and Haplogroup Predictor (HP) are also represented.

Sample ID	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385 a/b	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	HP	NevGen	Haplogroup Genotyped
185	17	14	21	31	23	10	18	15	11	19	10	15	21	9	11	36	18	15	16,19	29	13	11	24	36,38	11	E1b1a	J2a2 *PF5008*	J-M172
194	17	14	21	32	19	11	18.2	15	11	21	10	15	23	9.2	11	39	19	14	13,17	29	12	11	26	36,40	12	J1	J1a3 *Z1828*	J-12f2a*
197	18	12	23	28	21	10	17	13	13	19	10	15	23	12	15	39	18	14	15,15	29	13	11	23	35,41	12	Q	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	-
207	17	12	23	28	24	11	16	14	12	19	11	15	24	12	13	38	16	15	10,14	28	13	12	22	36,37	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
211	17	13	23	29	23	11	16	14	12	19	11	16	24	12	13	38	17	15	11,14	28	13	11	22	36,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
214	18	14	23	30	22	11	17	14	11	19	11	15	23	12	13	40	19	14	11,17	29	12	11	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
215	18	12	23	28	23	11	17	14	11	18	11	17	23	12	13	38	18	14	11,14	28	14	13	23	36,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
220	17	13	22	29	22	11	19	14	12	19	10	15	24	12	13	38	16	15	10,14	31	13	12	23	34,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
227	17	13	23	28	22	10	16	14	13	19	10	15	24	12	14	38	18	15	11,15	28	12	13	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-L23*
228	16	12	23	28	20	10	15	14	11	20	10	14	22	10	11	40	21	16	14,14	28	13	11	25	37,38	10	I1	I1 *(for 67+ markers, try level for I-s, 40+ subclades)*	I-M170*
231	17	14	22	32	18	11	17	13	12	19	10	10	23	11	14	38	16	14	14,14	30	13	11	23	33,39	11	Q	Q *M346>> Z780*	Q-M3*
232	21	13	22	31	22	10	15	13	12	20	11	15	24	11	14	36	15	14	14,17	31	13	12	24	36,37	11	Q	Q *M346>> M3*	Q-M3*

Table 16: Y- STR haplotypes and their corresponding haplogroups (Haplogroup Genotyped) from the Santander del Norte administrative region. The haplogroup assignment obtained by NEVGEN and Haplogroup Predictor (HP) are also represented (Continues in the next page).

Sample ID	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385 a/b	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	HP	NevGen	Haplogroup Genotyped
312	20	13	22	30	23	10	16	13	12	21	10	15	25	11	14	37	16	14	14,17	31	13	12	24	35,37	11	Q	Q *M346>> M3*	Q-M3*
315	16	13	23	29	22	10	16	16	11	18	12	16	23	12	13	38	15	14	12,14	29	13	12	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
318	17	13	23	29	21	11	17	14	11	19	11	15	24	12	13	38	17	14	11,12	29	13	11	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-U152
320	19	12	22	31	17	11	18	13	12	20	10	14	24	10	11	38	14	15	15,17	31	14	10	25	34,38	11	E1b1b	E1b1b *PF1975*	E-M35*
321	14	12	20	28	19	10	15	15	12	20	10	15	22	10	11	36	18	16	14,14	32	14	11	22	37,40	9	G2a	G2a*2b2a1b - L497 (for 67+ markers, try level for G-s)*	G-M201
324	19	13	23	31	19	11	16	13	12	21	10	14	24	11	15	38	20	14	14,17	30	13	11	23	36,38	12	Q	Q *M346>> M3> M902*	Q-M3*
327	18	14	23	30	22	10	17	14	11	18	10	16	24	12	13	39	17	14	11,14	30	13	11	22	34,36	13	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-M153
328	18	13	23	29	24	10	17	15	11	18	11	15	23	12	13	37	18	14	11.2,12	31	13	12	22	36,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
332	20	12	21	29	22	11	18	15	12	20	11	14	22	10	11	39	16	15	14,14	32	13	10	21	37,39	10	G2a	G2a *>> PF3359*	G-M201
333	18	14	23	30	20	11	18	15	12	20	10	15	25	12	13	37	18	15	11,14	30	13	13	22	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	-
337	18	12	23	27	22	10	17	14	12	19	11	16	24	12	13	42	16	15	11,14	31	14	11	22	35,35	13	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
342	17	14	23	30	23	10	17	14	11	18	11	15	24	12	13	38	17	14	11,14	29	13	12	22	35,35	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-S116*
343	18	13	23	31	18	11	17	16	13	0	11	15	25	11	11	42	19	15	11,14	30	13	11	23	36,39	12	R1a	R1a *(for 67+ markers, try level for R1a-s, 70+ subclades)*	R-SRY10831.2
345	18	13	23	29	19	11	17	13	11	21	10	14	24	11	15	36	18	14	14,17	30	13	12	23	34,39	11	Q	Q *M346>> M3> M902*	Q-M3*
349	16	14	21	30	22	11	16	14	12	21	10	15	23	9	11	40	17	15	14,16	33	12	12	22	37,39	11	G2a	J2a1 *Z7700*	J-M172
353	18	13	23	29	23	11	18	14	12	19	11	16	24	12	13	36	17	15	11,14	31	13	12	21	35,36	12	R1b	R1b *(for 67+ markers, try level for R1b-s, 300+ subclades)*	R-M529

(Continues next page)

382	19	13	23	29	23	11	18	14	11	18	11	15	24	12	13	37	18	14	11,14	29	13	11	23	35,36	12	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-S116*
385	18	13	23	28	22	10	16	14	12	20	11	16	25	12	13	38	19	15	11,14	28	13	12	22	35,37	12	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-U106
386	18	13	20	30	21	11	18.2	14	11	20	10	18	23	11	11	37	19	14	13,18	25	12	11	24	35,38	11	J1	J1a2a1a2 *P58*	J-12f2a*
430	20	13	22	31	21	10	16	13	12	20	10	14	23	10	14	38	17	14	14,15	29	13	12	25	35,39	11	Q	Q *M346>> M3> M902*	Q-M3*
432	16	14	22	30	19	11	20	13	12	20	9	16	24	10	11	41	22	14	13,14	32	13	10	26	36,38	11	E1b1b	E1b1b *V257>M81*	E-M81
433	19	13	23	29	24	11	16	14	12	19	11	16	24	12	13	38	17	15	11,14	28	13	11	22	36,36	12	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-S116*

Table 17: Y- STR haplotypes and their corresponding haplogroups (Haplogroup Genotyped) from the Tolima administrative region. The haplogroup assignment obtained by NEVGEN and Haplogroup Predictor (HP) are also represented.

Sample ID	DYS576	DYS389I	DYS635	DYS389II	DYS627	DYS460	DYS458	DYS19	YGATAH4	DYS448	DYS391	DYS456	DYS390	DYS438	DYS392	DYS518	DYS570	DYS437	DYS385 a/b	DYS449	DYS393	DYS439	DYS481	DYF387S1	DYS533	HP	NevGen	Haplogroup Genotyped
621	14	13	23	29	21	11	17	14	12	18	10	16	23	11	13	37	17	15	11,14	30	13	14	22	35,36	13	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-S116*
624	20	13	21	30	20	10	18.2	14	11	21	10	16	23	10	11	40	17	14	13,15	26	12	12	24	34,34	11	J1	J1a2a1a2 *P58*	J-12f2a*
629	19	13	23	29	16	10	15	14	11	18	11	16	25	12	13	39	17	15	11,14	27	13	14	22	35,36	12	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-S116*
633	17	13	23	29	24	11	17	14	12	19	11	15	25	13	13	36	18	14	11,14	28	13	13	22	35,37	12	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-U152
636	18	13	22	30	22	10	16	13	12	22	10	15	24	11	15	35	15	14	14,18	32	14	12	22	34,40	11	Q	Q *M346>> M3> M902*	Q-M3*
642	19	13	23	31	18	12	18	15	13	20	12	15	24	11	11	40	18	14	11,14	34	13	10	22	38,38	12	R1a	R1a *M198 (for 67+ markers, try level for R1a-s, 100+ subclades)*	R-SRY10831.2
643	17	12	21	29	19	10	18	13	11	21	10	15	25	10	11	40	20	14	16,16	33	12	12	24	39,41	10	E1b1b	E1b1b *M123>M34> M84*	E-M123
648	16	13	23	29	24	11	17	14	12	19	11	18	25	12	13	37	18	15	11,15	28	13	12	22	35,35	12	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-S116*
653	18	13	23	29	22	11	15	14	12	19	10	16	23	12	13	40	16	15	11,11	30	13	11	22	36,37	13	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-S116*
699	16	12	22	28	18	11	15	15	12	21	10	15	24	10	11	37	19	16	16,16	28	14	12	22	38,39	9	G2a	G2a2b2a1c *CTS342*	G-M201
701	18	13	24	29	23	11	16	14	12	19	10	16	24	12	13	37	18	15	11,15	28	13	12	22	36,36	12	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-S116*
711	17	12	23	27	20	11	16	14	12	19	11	15	24	12	13	40	17	15	12,14	30	13	13	22	36,36	12	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-S116*
715	17	12	23	30	18	11	17	13	13	22	10	15	24	11	14	40	18	15	14,18	31	12	13	24	36,37	11	Q	Q *M346>> Y4800> F835> L932*	Q-M3*
716	19	13	20	28	23	11	18	15	13	20	10	15	23	10	11	36	19	15	12,12	29	13	12	20	35,37	12	I2a1	I2a1a *Sardinian M26*	I-M26
717	18	13	23	32	23	10	20	14	12	19	11	15	24	12	13	37	17	15	12,14	30	13	13	22	36,36	12	R1b	R1b *M269 (for 67+ markers, try level for R1b-s, 400+ subclades)*	R-S116*
719	19	13	23	31	19	11	17	13	12	20	10	13	24	13	15	40	19	14	15,15	30	13	11	23	34,38	11	Q	Q *M346>> M3> M902*	Q-M3*

Table 18: Pairwise R_{ST} genetic distances (below the diagonal) and the corresponding P-Values (above the diagonal) using the Y-STR data of the 7 populations studied. Values non-statistically significant are highlighted in bold (P≥0.05).

	Guaínia	Barí-Motilon	Paeces	Cundinamarca	Huila	Santander del Norte	Tolima
Guainía	-	0.02891	0.00356	0.01752	**0.06356**	0.01119	0.01693
Barí-Montilon	0.06911	-	0.00000	0.00000	0.00099	0.00000	0.00040
Paeces	0.10981	0.27613	-	**0.15583**	**0.25750**	**0.35937**	**0.44679**
Cundinamarca	0.06552	0.17993	**0.02814**	-	**0.59994**	**0.55787**	**0.31779**
Huila	**0.05354**	0.21705	**0.01852**	**- 0.01220**	-	**0.75428**	**0.45015**
Santander del Norte	0.06135	0.18878	**0.00520**	**-0.00608**	**-0.02004**	-	**0.87140**
Tolima	0.06352	0.18880	**-0.00145**	**0.00791**	**-0.00249**	**-0.02376**	**-**

Table 19: Pairwise F_{ST} genetic distances (below the diagonal) and the corresponding P-Values (above the diagonal) using the Y-SNP data of the 7 populations studied. Values non-statistically significant are highlighted in bold (P≥0.05).

	Guainía	Barí-Montilon	Paeces	Cundinamarca	Huila	Santander del Norte	Tolima
Guainía	-	0.01465	0.00079	0.0001	0.00188	0.0002	0.00069
Barí-Montilon	0.13343	-	0	0	0	0	0
Paeces	0.28005	0.64568	-	**0.18711**	**0.24681**	**0.35056**	**0.29938**
Cundinamarca	0.20468	0.48084	**0.02733**	-	**0.10108**	**0.52262**	**0.18305**
Huila	0.27214	0.63542	**0.02674**	**0.03932**	-	**0.73488**	**0.9999**
Santander del Norte	0.18208	0.45634	**0.00478**	**-0.00422**	**-0.02137**	-	**0.97228**
Tolima	0.2328	0.54975	**0.01401**	**0.02121**	**-0.04609**	**-0.03371**	-

Table 20: Haplogroup results for each of the populations studied, as well as, the total frequency across all the 121 samples. Every haplogroup has their absolute value (N) and the relative frequency (Freq). The Haplogroups are divided in their major clade, as well in their sub-clade.

Haplogroup		**Populations**														**Total**	
Clade	Sub-Clade	Guainía		Barí-Motilon		Paeces		Cundinamarca		Huila		Santander del Norte		Tolima			
		N	Freq.	N	Freq.	N	Freq.	N	Freq.	N	Freq.	N	Freq.	N	Freq.	N	Freq.
A	A-M13							1	5.00							1	0.83
E	E-M35*	1	5.56									1	4.54			2	1.65
	E-M78							2	10.00							2	1.65
	E-M81											1	4.54			1	0.83
	E-M123													1	6.25	1	0.83
	E-U174	1	5.56													1	0.83
	E-U290							1	5.00							1	0.83
G	G-M201							1	5.00			2	9.10	1	6.25	4	3.31
I	I-M170*									1	8.33					1	0.83
	I-M26													1	6.25	1	0.83
J	J-12f2a*	1	5.56					2	10.00	1	8.33	1	4.54	1	6.25	6	4.96
	J-M172	1	5.56					2	10.00	1	8.33	1	4.54			5	4.13
Q	Q-M346*					3	25									3	2.48
	Q-M3*	13	72.22	21	100	1	8.33	2	10.00	2	16.67	4	18.18	3	18.75	47	38.84
R	R-SRY10831.2					1	8.33	1	5.00			1	4.54	1	6.25	4	3.31
	R-L23*									1	8.33					1	0.83
	R-U106							1	5.00			1	4.54			2	1.65
	R-S116*	1	5.56			3	25	3	15.00	6	50.00	7	31.82	7	43.75	25	20.66
	R-U152							3	15.00			1	4.54	1	6.25	5	4.13
	R-M529					3	25					1	4.54			5	4.13
	R-M153											1	4.54			1	0.83
T	T-M70					1	8.33	1	5.00							2	1.65
		18		21		12		20		12		22		16		121	

Table 21: R_{ST} genetic distances results (bellow the diagonal) and its P-Values (above the diagonal) with 15 Y-STRs between our three Native populations (Guainía, Barí-Motilon and Paeces) and other Latin American Native groups; a European population and an African group. Non-statistically significant values are in bold (Continues in the next pages).

	GUA	BM	PAE	GUAM	COC	EC	ANT	CAU	COL	TOBA	KI	WAO	WAY	PAL	EM	AY	QU	KY	PA
GUA	*	0.03614	0.00069	0.01386	0.01683	0.00000	0.00000	0.00436	0.00000	0.00000	**0.07524**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000
BM	0.06379	*	0.00000	0.00030	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00792	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PAE	0.14369	0.2209	*	**0.06138**	**0.16860**	0.00030	0.00000	0.00069	**0.07821**	0.00000	0.00010	0.00000	0.00000	0.00040	0.00000	0.00000	0.00000	0.00000	0.00000
GUAM	0.096	0.22027	**0.04532**	*	**0.12434**	**0.11593**	0.03227	**0.08148**	0.00436	0.00050	**0.07554**	0.00000	0.00010	0.00119	0.00000	0.00010	0.00743	0.00069	0.00000
COC	0.15346	0.46449	**0.05899**	**0.07443**	*	0.00396	0.00040	0.02158	0.00386	0.00030	0.00277	0.01000	0.00000	0.00347	0.00000	0.00000	0.00535	0.00020	0.00000
EC	0.21478	0.33101	0.16098	**0.03813**	0.18791	*	**0.14682**	0.00000	0.00000	0.00000	0.00149	0.00000	0.00000	0.00000	0.00000	0.00010	0.00752	0.00000	0.00000
ANT	0.21476	0.33768	0.18587	0.07446	0.23479	**0.02744**	*	0.00010	0.00000	0.00000	0.00337	0.00000	0.00000	0.00000	0.00000	0.00000	0.00099	0.00000	0.00000
CAU	0.07737	0.15641	0.13001	**0.03843**	0.11816	0.16097	0.10935	*	0.00000	0.00000	0.02208	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00129	0.00000
COL	0.1816	0.29333	**0.05417**	0.0871	0.16882	0.18558	0.2647	0.15171	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
TOBA	0.15134	0.20274	0.2402	0.11224	0.24491	0.14247	0.26691	0.22372	0.14663	*	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
KI	**0.05022**	0.14318	0.14914	**0.05701**	0.22276	0.15482	0.12533	0.05915	0.2017	0.16025	*	0.00000	0.00000	0.00000	0.00000	0.00010	0.00554	0.00020	0.00010
WAO	0.2951	0.4927	0.2478	0.21752	0.2064	0.26667	0.29717	0.21348	0.24073	0.34381	0.33023	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
WAY	0.43392	0.6024	0.22202	0.31409	0.4453	0.45895	0.43919	0.25205	0.3645	0.48942	0.38485	0.55697	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PAL	0.34507	0.46717	0.15758	0.15779	0.22258	0.27316	0.27703	0.18816	0.23664	0.354	0.28219	0.30339	0.2622	*	0.00000	0.00000	0.00000	0.00000	0.00000
EM	0.27484	0.48068	0.21095	0.26749	0.495	0.28189	0.33562	0.23941	0.32571	0.3373	0.41439	0.51038	0.7156	0.54513	*	0.00000	0.00000	0.00000	0.00000
AY	0.2538	0.2649	0.29967	0.15582	0.25385	0.10876	0.14622	0.22088	0.25434	0.15197	0.14445	0.3163	0.42289	0.28153	0.35878	*	**0.56390**	0.00000	0.00000
QU	0.19794	0.22904	0.23641	0.09738	0.19849	0.06503	0.10301	0.17442	0.19573	0.11777	0.11233	0.2732	0.38028	0.25725	0.30242	**-0.00403**	*	0.00000	0.00000
KY	0.18718	0.33745	0.19546	0.28463	0.44197	0.40693	0.33132	0.12743	0.34734	0.43582	0.24257	0.42132	0.56391	0.401	0.6056	0.34794	0.30803	*	0.00000
PA	0.19825	0.18848	0.24242	0.14702	0.28962	0.12847	0.15607	0.18634	0.24931	0.17782	0.14622	0.31408	0.51256	0.35373	0.2546	0.1402	0.13108	0.25673	*
WAYUU	0.19256	0.25928	0.08978	**0.06007**	0.17083	0.18362	0.14864	0.06023	0.17514	0.26395	0.08296	0.331	0.1609	0.09741	0.37893	0.17665	0.15425	0.23635	0.20281
BB	0.17917	**0.09065**	0.20974	0.30856	0.54161	0.42667	0.41845	0.20444	0.34907	0.32674	0.26642	0.57226	0.60103	0.47691	0.6217	0.30925	0.28172	0.35647	0.26572

(Continues in the next 3 pages)

	PA	WAYUU	BB	NA	MAYA	PIMA	CHU	HU	SC	ASH	AP	KX	Z	UK	SP	YO
GUA	0.00000	0.00069	0.00050	0.00000	0.02525	0.00000	0.00020	0.00475	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
BM	0.00000	0.00010	**0.06574**	0.00000	0.00069	0.00000	0.00000	0.00089	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PAE	0.00000	0.00465	0.00000	0.00000	0.01198	0.00000	0.00000	0.00010	0.00000	0.00000	0.02475	0.00099	0.00000	0.00000	0.00079	0.00000
GUAM	0.00000	**0.06871**	0.00020	**0.06752**	**0.74923**	0.00089	0.00772	0.00921	0.00426	**0.07187**	**0.20661**	0.00257	0.00000	0.00000	0.00000	0.00000
COC	0.00000	0.00990	0.00000	0.00277	**0.12623**	0.00079	0.00168	0.00020	0.02237	0.00446	0.00931	0.00010	0.00000	0.00000	0.00663	0.00000
EC	0.00000	0.00000	0.00000	0.01643	0.04881	0.00000	0.00713	0.00119	0.00000	0.00000	0.00158	0.00010	0.00000	0.00000	0.00000	0.00000
ANT	0.00000	0.00109	0.00000	0.00911	0.01832	0.00030	0.00079	0.00010	0.00000	0.00000	0.00347	0.00000	0.00000	0.00000	0.00000	0.00000
CAU	0.00000	0.01535	0.00000	0.00000	0.04990	0.00000	0.00000	0.00010	0.00000	0.00000	0.00129	0.00050	0.00010	0.00000	0.00000	0.00000
COL	0.00000	0.00000	0.00000	0.00000	0.00238	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
TOBA	0.00000	0.00000	0.00000	0.00000	0.00198	0.00000	0.00020	0.00139	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
KI	0.00010	0.03218	0.00000	0.00317	**0.23374**	0.00000	0.00485	0.01297	0.00000	0.00050	0.00178	0.00030	0.00000	0.00000	0.00000	0.00000
WAO	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
WAY	0.00000	0.00317	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PAL	0.00000	0.00545	0.00000	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00347	0.00525	0.00000	0.00000	0.00000	0.00000
EM	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
AY	0.00000	0.00000	0.00000	0.00000	0.00030	0.00000	**0.99030**	**0.33838**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
QU	0.00000	0.00040	0.00000	0.00010	0.01792	0.00000	**0.74240**	**0.23671**	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
KY	0.00000	0.00079	0.00000	0.00000	0.00040	0.00000	0.00010	0.00099	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PA	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00158	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
WAYUU	0.20281	*	0.00040	0.00000	**0.08593**	0.00000	0.00198	0.00257	0.00000	0.01356	**0.05633**	0.00079	0.00000	0.00000	0.00000	0.00000
BB	0.26572	0.24463	*	0.00000	0.00000	0.00000	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

	GUA	BM	PAE	GUAM	COC	EC	ANT	CAU	COL	TOBA	KI	WAO	WAY	PAL	EM	AY	QU	KY	PA
NA	0.13013	0.21904	0.21431	**0.03488**	0.14739	0.04247	0.0517	0.10887	0.20488	0.15788	0.08922	0.17818	0.40541	0.26959	0.23432	0.13113	0.0878	0.2565	0.08718
MAYA	0.08501	0.21284	0.06837	**-0.02958**	**0.05806**	0.05887	0.08083	0.04604	0.10857	0.10894	**0.01772**	0.26155	0.31793	0.18336	0.32522	0.13045	0.08407	0.30862	0.13825
PIMA	0.39696	0.52237	0.23053	0.16849	0.30562	0.22158	0.17045	0.21333	0.35498	0.41222	0.29748	0.38282	0.36584	0.20509	0.56564	0.29053	0.24127	0.49474	0.37435
CHU	0.20094	0.30501	0.17125	0.12353	0.23938	0.10126	0.13527	0.17999	0.22964	0.15377	0.13174	0.31034	0.47506	0.24559	0.46212	**-0.0312**	**-0.01768**	0.33885	0.12967
HU	0.1334	0.20502	0.17077	0.14395	0.27527	0.15348	0.18128	0.1759	0.23529	0.1239	0.10531	0.35162	0.52165	0.30403	0.46966	**0.00407**	**0.01242**	0.28016	0.10366
SC	0.25651	0.47021	0.16515	0.13601	0.12684	0.21623	0.18176	0.15144	0.25918	0.34594	0.20982	0.28627	0.31581	0.19886	0.534	0.26821	0.21233	0.41188	0.36608
ASH	0.2244	0.2537	0.21192	**0.03598**	0.14302	0.10038	0.13335	0.12127	0.18031	0.14438	0.11606	0.3007	0.29452	0.14445	0.32499	0.08754	0.07108	0.33572	0.17054
AP	0.22114	0.2829	0.09435	**0.02228**	0.1822	0.11076	0.09893	0.07659	0.16942	0.24324	0.15195	0.24822	0.32448	0.10338	0.30854	0.19144	0.15237	0.27013	0.18976
KX	0.29435	0.53984	0.14299	0.22435	0.3522	0.32662	0.34086	0.20934	0.25188	0.3703	0.30682	0.32889	0.56437	0.16508	0.71902	0.30596	0.27961	0.50626	0.34244
Z	0.26719	0.6412	0.2599	0.275	0.60981	0.3256	0.25287	0.13781	0.36695	0.41965	0.3349	0.33082	0.70368	0.52029	0.88105	0.32718	0.25869	0.56684	0.34033
UK	0.21156	0.44887	0.31809	0.30741	0.6055	0.42028	0.37457	0.18693	0.42141	0.43412	0.33122	0.53007	0.69506	0.54201	0.68377	0.42231	0.36404	0.51359	0.33181
SP	0.16529	0.25405	0.17411	0.15437	0.18412	0.2603	0.26624	0.10227	0.10142	0.24409	0.20943	0.23165	0.29837	0.28546	0.28196	0.32229	0.2828	0.28982	0.30238
YO	0.52049	0.63457	0.53215	0.61131	0.60253	0.6907	0.68982	0.51915	0.53635	0.63737	0.58943	0.63653	0.64255	0.59963	0.73617	0.65645	0.63395	0.61846	0.68342

	WAYUU	BB	NA	MAYA	PIMA	CHU	HU	SC	ASH	AP	KX	Z	UK	SP	YO
NA	0.16998	0.31201	*	0.03039	0.00000	0.00208	0.00158	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
MAYA	**0.04891**	0.29928	0.0536	*	0.00119	0.01000	0.01030	0.00545	**0.10435**	0.02841	0.00069	0.00000	0.00000	0.00010	0.00000
PIMA	0.20422	0.55707	0.20708	0.1664	*	0.00000	0.00000	0.00218	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CHU	0.14394	0.35995	0.10018	0.1093	0.2834	*	**0.69845**	0.00000	0.00891	0.00158	0.00000	0.00000	0.00000	0.00000	0.00000
HU	0.17515	0.269	0.11268	0.12117	0.36533	**-0.02996**	*	0.00000	0.00218	0.00099	0.00000	0.00000	0.00000	0.00000	0.00000
SC	0.19429	0.51577	0.1986	0.10538	0.1115	0.24776	0.29192	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ASH	0.05664	0.28688	0.12064	**0.03049**	0.21061	0.06595	0.10562	0.19679	*	0.00079	0.00000	0.00000	0.00000	0.00000	0.00000
AP	**0.05711**	0.32499	0.12361	0.07988	0.16124	0.14932	0.19027	0.20332	0.07724	*	0.00733	0.00000	0.00000	0.00000	0.00000
KX	0.19603	0.61422	0.26201	0.2187	0.33441	0.30123	0.33715	0.32455	0.25215	0.17493	*	0.00000	0.00000	0.00000	0.00000
Z	0.40043	0.79429	0.14026	0.33496	0.44194	0.49057	0.54012	0.34539	0.34015	0.32997	0.80968	*	0.00000	0.00000	0.00000
UK	0.40006	0.58716	0.20546	0.32408	0.49662	0.5275	0.49344	0.49339	0.36878	0.36108	0.67276	0.79559	*	0.00000	0.00000
SP	0.20996	0.31117	0.24332	0.16817	0.32703	0.31501	0.31813	0.25877	0.2366	0.21988	0.30171	0.21358	0.30029	*	0.00000
YO	0.58044	0.64154	0.64827	0.59676	0.69986	0.64695	0.63062	0.62018	0.63069	0.62549	0.57209	0.71706	0.70235	0.42053	*

Table 22: F_{ST} genetic distances results (bellow the diagonal) and its P-Values (above the diagonal) with 16 Y-SNPs between our three Native populations (Guainía, Barí-Motilon and Paeces) and other Latin American Native groups; a European population and an African group. Non-statistically significant values are in bold (Continues in the next pages).

	GUA	BM	PAE	GUAM	COC	EC	ANT	CAU	COL	TOBA	WAY	PAL	EM	AY	QU	PA
GUA	*	0.01515	0.00020	0.00188	**0.30848**	**0.17751**	**0.15375**	0.03049	**0.05950**	0.00000	0.00495	0.02723	**0.19751**	**0.07643**	**0.45718**	0.02653
BM	0.14576	*	0.00000	0.00010	**0.99990**	**0.11108**	**0.11345**	0.00010	**0.99990**	**0.99990**	**0.99990**	**0.37828**	**0.99990**	**0.12573**	0.03812	**0.37115**
PAE	0.34862	0.68701	*	0.00416	0.00020	0.00000	0.00000	0.03515	0.00000	0.00000	0.00000	0.00000	0.00020	0.00000	0.00000	0.00000
GUAM	0.26336	0.57468	0.24578	*	0.01911	0.01554	0.01733	0.02534	0.00079	0.00000	0.00000	0.00000	0.00881	0.00050	0.00000	0.00000
COC	**0.03747**	**0**	0.52974	0.42564	*	**0.54905**	**0.54579**	0.01554	**0.99990**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	**0.68211**	**0.36175**	**0.99990**
EC	**0.03317**	**0.11857**	0.47463	0.26328	**0.02778**	*	**0.99990**	0.00297	**0.26849**	0.00178	0.03406	**0.18097**	**0.31571**	**0.63598**	**0.07237**	**0.31363**
ANT	**0.0318**	**0.12791**	0.4636	0.24967	**0.0338**	**-0.04427**	*	0.00297	**0.14543**	0.00109	0.03386	**0.15880**	**0.30344**	**0.63132**	**0.07376**	**0.31462**
CAU	0.07126	0.26886	0.08963	0.08653	0.18925	0.13008	0.12365	*	0.00020	0.00000	0.00000	0.00000	0.00436	0.00000	0.00020	0.00000
COL	**0.10449**	**0**	0.62584	0.5141	**0**	**0.08547**	**0.09316**	0.23809	*	**0.99990**	**0.99990**	**0.54618**	**0.99990**	**0.24017**	**0.11306**	**0.54816**
TOBA	0.40273	**-0.02382**	0.89315	0.82312	**-0.07673**	0.30507	0.32496	0.50209	**-0.03666**	*	**0.99990**	0.02772	**0.99990**	0.00030	0.00000	0.02831
WAY	0.18384	**0**	0.73579	0.62675	**0**	0.14852	0.15956	0.29766	**0**	**-0.01675**	*	**0.24522**	**0.99990**	**0.05653**	0.01564	**0.24691**
PAL	0.06836	**0.01672**	0.61446	0.47147	**-0.04701**	**0.03101**	**0.03638**	0.22607	**-0.00204**	**0.08411**	**0.03076**	*	**0.72448**	**0.28779**	**0.06108**	**0.70696**
EM	**0.06263**	**0**	0.56297	0.45572	**0**	**0.05013**	**0.05661**	0.20741	**0**	**-0.05858**	**0**	**-0.02773**	*	**0.48411**	**0.31819**	**0.99990**
AY	**0.04628**	**0.04561**	0.56755	0.37948	**-0.0144**	**-0.0185**	**-0.01658**	0.1938	**0.0279**	0.12413	**0.05889**	**0.00425**	**0.00371**	*	0.04039	**0.48866**
QU	**-0.00681**	0.0779	0.44803	0.35576	**0.01412**	**0.04004**	**0.04119**	0.1303	**0.05777**	0.19853	0.09415	**0.03712**	**0.03233**	0.0317	*	**0.05099**
PA	0.07171	**0.01502**	0.61963	0.46188	**-0.04805**	**0.01251**	**0.01766**	0.22494	**-0.00341**	0.0755	**0.02867**	**-0.01368**	**-0.02886**	**-0.00791**	**0.04222**	*
WAYUU	0.12358	**0**	0.6549	0.54231	**0**	**0.10093**	**0.10934**	0.25226	**0**	**-0.02984**	**0**	**0.00727**	**0**	**0.03669**	**0.06755**	**0.00576**
BB	0.11752	**0**	0.64576	0.53333	**0**	**0.09605**	**0.10422**	0.24774	**0**	**-0.03182**	**0**	**0.00444**	**0**	**0.03402**	**0.06454**	**0.00298**

(Continues in the next 3 pages)

	WAYUU	BB	KI	WAO	NA	MAYA	CHU	HU	SC	ASH	AP	KX	Z	UK	SP	MOZ
GUA	0.04594	0.04703	**0.15484**	**0.05603**	0.00158	**0.13256**	**0.05000**	**0.10454**	0.01445	0.01158	0.00099	**0.12543**	0.01020	0.00624	0.00000	0.00000
BM	**0.99990**	**0.99990**	**0.42134**	**0.28601**	0.00000	0.00119	**0.99990**	**0.99990**	**0.99990**	**0.44649**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
PAE	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.02426	0.00000
GUAM	0.00020	0.00079	0.00109	0.00000	0.00000	0.00337	0.00109	0.00109	0.00000	0.00000	0.00000	0.00782	0.00000	0.00000	0.00000	0.00000
COC	**0.99990**	**0.99990**	**0.99990**	**0.70805**	0.00545	**0.10197**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
EC	**0.12415**	**0.13405**	**0.19493**	**0.05485**	0.00010	0.01762	**0.14177**	**0.28067**	**0.11157**	**0.19275**	0.00950	**0.29344**	**0.05059**	0.04257	0.00000	0.00000
ANT	**0.12207**	**0.12949**	**0.19226**	**0.05118**	0.00000	0.02119	**0.13197**	**0.27770**	**0.10949**	**0.18681**	0.00832	**0.29255**	0.04227	0.04029	0.00000	0.00000
CAU	0.00000	0.00020	0.00059	0.00000	0.00000	0.01485	0.00010	0.00089	0.00000	0.00000	0.00000	0.00257	0.00000	0.00000	0.00000	0.00000
COL	**0.99990**	**0.99990**	**0.99990**	**0.44976**	0.00010	0.00842	**0.99990**	**0.99990**	**0.99990**	**0.56905**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
TOBA	**0.99990**	**0.99990**	**0.18018**	0.00356	0.00000	0.00000	**0.99990**	**0.99990**	**0.99990**	0.02109	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
WAY	**0.99990**	**0.99990**	**0.34601**	**0.17573**	0.00000	0.00040	**0.99990**	**0.99990**	**0.99990**	**0.28116**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
PAL	**0.53371**	**0.54430**	**0.79279**	**0.35610**	0.00000	0.00069	**0.53985**	**0.55014**	**0.37590**	**0.54856**	**0.07376**	**0.72518**	**0.25384**	**0.24720**	0.00000	0.00000
EM	**0.99990**	**0.99990**	**0.99990**	**0.68736**	0.00198	0.04257	**0.99990**	**0.99990**	**0.99990**	**0.71033**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
AY	**0.18780**	**0.24354**	**0.34244**	**0.12593**	0.00000	0.00327	**0.24631**	**0.33472**	**0.12563**	**0.40224**	0.01139	**0.36620**	**0.09752**	**0.08207**	0.00000	0.00000
QU	**0.07979**	**0.08583**	**0.16979**	0.04010	0.00000	0.00812	**0.08999**	**0.16038**	0.03950	0.01129	0.00079	**0.22542**	**0.02713**	0.02089	0.00000	0.00000
PA	**0.54163**	**0.53678**	**0.79052**	**0.35323**	0.00000	0.00109	**0.54470**	**0.54985**	**0.37957**	**0.99990**	**0.07356**	**0.71577**	**0.26354**	**0.24869**	0.00000	0.00000
WAYUU	*	**0.99990**	**0.46976**	**0.29928**	0.00000	0.00277	**0.99990**	**0.99990**	**0.99990**	**0.56876**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
BB	**0**	*	**0.47886**	**0.44659**	0.00000	0.00406	**0.99990**	**0.99990**	**0.99990**	**0.55905**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000

	GUA	BM	PAE	GUAM	COC	EC	ANT	CAU	COL	TOBA	WAY	PAL	EM	AY	QU	PA
KI	**0.04524**	**0.02326**	0.55742	0.43786	**-0.06061**	**0.03602**	**0.04125**	0.19714	**-0.00478**	**0.10558**	**0.0471**	**-0.01942**	**-0.03789**	**0.00305**	**0.02517**	**-0.01948**
WAO	**0.06247**	**0.0281**	0.61147	0.47583	**-0.03539**	**0.04921**	**0.05432**	0.23184	**0.00918**	0.10919	**0.04244**	**0.0029**	**-0.01636**	**0.01997**	0.03982	**0.00325**
NA	0.16864	0.33896	0.28499	0.25862	0.26496	0.24237	0.23758	0.14732	0.30964	0.56455	0.36675	0.30676	0.28124	0.29923	0.24387	0.301
MAYA	**0.03613**	0.25616	0.2874	0.21954	**0.1315**	0.11209	0.10743	0.07801	0.20622	0.56322	0.30265	0.18677	0.15828	0.16076	0.10209	0.18274
CHU	**0.11752**	**0**	**0.64576**	0.53333	**0**	**0.09605**	**0.10422**	0.24774	**0**	**-0.03182**	**0**	0.00444	**0**	**0.03402**	**0.06454**	**0.00298**
HU	0.09742	**0**	**0.61493**	0.50375	**0**	**0.07965**	**0.0871**	0.23286	**0**	**-0.03965**	**0**	-0.00584	**0**	**0.02432**	**0.05388**	**-0.00716**
SC	0.14576	**0**	0.68701	0.57468	**0**	**0.11857**	**0.12791**	0.26886	**0**	**-0.02382**	**0**	**0.01672**	**0**	**0.04561**	0.0779	**0.01502**
ASH	0.09725	**0.01101**	0.66652	0.50833	**-0.04725**	**0.02412**	**0.03028**	0.26038	**-0.00512**	0.05658	**0.02213**	**-0.009**	**-0.02882**	**-0.0013**	0.05575	**-0.02339**
AP	0.25716	**0**	0.80684	0.71009	**0**	0.20661	0.22096	0.35412	**0**	**-0.00922**	**0**	**0.05478**	**0**	0.08164	0.12385	**0.0518**
KX	**0.07268**	**0**	0.57746	0.46894	**0**	**0.05882**	**0.06554**	0.21472	**0**	**-0.05236**	**0**	**-0.02083**	**0**	**0.0102**	**0.039**	**-0.02201**
Z	0.1656	**0**	0.7135	0.6025	**0**	**0.1342**	0.14441	0.28382	**0**	**-0.01973**	**0**	**0.02429**	**0**	**0.05277**	**0.08654**	**0.02239**
UK	0.17489	**0**	0.7251	0.61502	**0**	0.14149	0.15213	0.29086	**0**	**-0.01813**	**0**	**0.02763**	**0**	**0.05593**	0.09044	**0.02564**
SP	0.50074	0.64725	0.0949	0.51412	0.6188	0.58079	0.5774	0.30784	0.63517	0.75477	0.6593	0.62276	0.62443	0.60835	0.52731	0.62521
MOZ	0.47683	0.62216	0.41925	0.48191	0.5899	0.55272	0.54904	0.387	0.60852	0.74149	0.63569	0.59764	0.59632	0.58956	0.54693	0.6006

	WAYUU	BB	KI	WAO	NA	MAYA	CHU	HU	SC	ASH	AP	KX	Z	UK	SP	MOZ
KI	**0.00861**	**0.00444**	*	**0.99990**	0.00000	0.01475	**0.48371**	**0.99990**	**0.41293**	**0.74349**	**0.23047**	**0.99990**	**0.36689**	**0.35719**	0.00000	0.00000
WAO	**0.01854**	**0.01569**	**-0.04236**	*	0.00000	0.00218	**0.43907**	**0.56084**	**0.28631**	**0.20642**	0.04069	**0.68191**	**0.18632**	**0.18434**	0.00000	0.00000
NA	0.32307	0.31877	0.27021	0.30452	*	**0.18790**	0.00010	0.00000	0.00000	0.00000	0.00000	0.00099	0.00000	0.00000	0.00000	0.00000
MAYA	0.22913	0.2218	0.14486	0.17424	**0.02552**	*	0.00455	0.01139	0.00089	0.00020	0.00000	0.03435	0.00040	0.00040	0.00000	0.00000
CHU	**0**	**0**	**0.00444**	**0.01569**	0.31877	0.2218	*	**0.99990**	**0.99990**	**0.57242**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
HU	**0**	**0**	**-0.00999**	**0.00539**	0.30473	0.19786	**0**	*	**0.99990**	**0.68072**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
SC	**0**	**0**	**0.02326**	**0.0281**	0.33896	0.25616	**0**	**0**	*	**0.44708**	**0.99990**	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
ASH	**0.00305**	**0.00061**	**-0.01453**	**0.00985**	0.34017	0.22623	**0.00061**	**-0.00853**	**0.01101**	*	**0.11840**	**0.70627**	**0.29651**	**0.29759**	0.00000	0.00000
AP	**0**	**0**	**0.09214**	0.0673	0.42131	0.38952	**0**	**0**	**0**	**0.03953**	*	**0.99990**	**0.99990**	**0.99990**	0.00000	0.00000
KX	**0**	**0**	**-0.02941**	**-0.00953**	0.28791	0.16942	**0**	**0**	**0**	**-0.02233**	**0**	*	**0.99990**	**0.99990**	0.00000	0.00000
Z	**0**	**0**	**0.03581**	**0.03581**	0.35337	0.28042	**0**	**0**	**0**	**0.01711**	**0**	**0**	*	**0.99990**	0.00000	0.00000
UK	**0**	**0**	**0.04158**	**0.03923**	0.36017	0.29176	**0**	**0**	**0**	**0.01972**	**0**	**0**	**0**	*	0.00000	0.00000
SP	0.6406	0.63884	0.60948	0.61971	0.44981	0.47805	0.63884	0.63324	0.64725	0.64386	0.684	0.62685	0.65345	0.65641	*	0.00000
MOZ	0.61466	0.61268	0.5821	0.59572	0.422	0.44565	0.61268	0.60633	0.62216	0.62152	0.66326	0.59908	0.62912	0.63245	0.50728	*

Table 23: R_{ST} genetic distances results (bellow the diagonal) and its P-Values (above the diagonal) with 15 Y-STRs between our four admixed populations (Cundinamarca, Huila, Santander del Norte and Tolima) and other Latin American admixed groups; two European population; a Native American group (Barí-Motilon); a Southeast Asian population and an African group. Non-statistically significant values are in bold (Continues in the next pages).

	CUN	**HU**	**SN**	**TOL**	**BOL**	**EC**	**COL**	**CR**	**CH**	**MEX**	**ARG**	**PAN**	**PE**	**PR**
CUN	*	**0.28393**	**0.12524**	0.04316	0.00000	0.00545	**0.24641**	**0.05049**	0.04029	0.00000	**0.41303**	**0.35244**	0.00129	**0.13464**
HU	**0.01112**	*	**0.93575**	**0.37264**	0.04505	**0.22235**	**0.60331**	**0.29086**	**0.32106**	**0.07861**	**0.70775**	**0.27631**	**0.22721**	**0.12464**
SN	**0.02169**	**-0.03957**	*	**0.36739**	0.00099	0.01950	**0.10365**	**0.04643**	**0.18622**	0.00475	**0.42214**	0.01099	0.00277	0.02594
TOL	0.05742	**0.00462**	**0.00338**	*	0.00921	**0.05633**	**0.18780**	**0.37907**	**0.14444**	0.04910	**0.20750**	0.01960	**0.12236**	0.03396
BOL	0.1563	0.05286	0.07667	0.06782	*	0.00000	0.00000	0.00000	0.00000	0.04346	0.00000	0.00000	0.00485	0.00000
EC	0.04461	**0.01193**	0.03878	**0.02607**	0.05703	*	0.00000	0.00000	0.00000	0.00000	0.00178	0.00000	0.00475	0.00000
COL	**0.00665**	**-0.00907**	**0.01686**	**0.01248**	0.0905	0.01264	*	0.00059	0.00386	0.00000	**0.43144**	0.00000	0.00000	0.00020
CR	**0.02911**	**0.00665**	**0.03222**	**0.00047**	0.11887	0.03521	0.01307	*	0.03871	0.00000	**0.25661**	0.00050	0.00000	0.02881
CH	0.03026	**0.00431**	**0.01084**	**0.01657**	0.13487	0.03799	0.0092	0.0066	*	0.00000	**0.76695**	0.00000	0.00000	0.00297
MEX	0.10091	**0.03446**	0.06637	0.03687	0.01283	0.02578	0.04961	0.08145	0.08824	*	0.00000	0.00000	0.00010	0.00000
ARG	**-0.00117**	**-0.01917**	**-0.00038**	**0.01376**	0.11405	0.02876	**-0.00034**	**0.00274**	**-0.00495**	0.08061	*	**0.35996**	0.00010	**0.30611**
PAN	**0.00196**	**0.00977**	0.04489	0.04601	0.11973	0.04614	0.02802	0.02919	0.03438	0.10169	**0.00079**	*	0.00000	0.01683
PE	0.07771	**0.01317**	0.05935	**0.02036**	0.03064	0.01402	0.03037	0.0481	0.06512	0.02108	0.05277	0.06317	*	0.00000
PR	**0.01583**	**0.02765**	0.03906	0.0422	0.16338	0.05859	0.02278	0.00989	0.01873	0.12057	**0.00178**	0.014	0.08058	*

(Continues in the next 3 pages)

	CW_BR	N_BR	NE_BR	S_BR	SE_BR	PT	SP	CN	BM	YOR
CUN	**0.13820**	**0.18622**	**0.27284**	**0.13108**	**0.18582**	0.02624	0.01703	0.00000	0.00000	0.00000
HU	**0.23602**	**0.32987**	**0.25958**	**0.12276**	**0.38442**	**0.16969**	**0.24611**	0.00040	0.00000	0.00000
SN	0.01891	**0.05712**	0.04227	0.02237	**0.07772**	**0.05079**	**0.16464**	0.00000	0.00000	0.00000
TOL	**0.09197**	**0.13989**	**0.08969**	0.04623	**0.06296**	**0.11553**	**0.13801**	0.00000	0.00000	0.00000
BOL	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
EC	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
COL	0.00000	0.00436	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
CR	**0.08316**	**0.70161**	0.00881	0.01495	0.03020	0.02376	0.00673	0.00000	0.00000	0.00000
CH	0.01158	0.01980	0.00564	0.01931	**0.07880**	**0.18959**	**0.67815**	0.00000	0.00000	0.00000
MEX	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ARG	**0.62974**	**0.25235**	**0.56103**	**0.23661**	**0.64637**	**0.43164**	**0.25948**	0.00000	0.00000	0.00000
PAN	0.00198	0.00604	0.00069	0.00010	0.00069	0.00010	0.00000	0.00000	0.00000	0.00000
PE	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000
PR	**0.32898**	**0.12949**	**0.49599**	0.00495	0.03188	0.01525	0.00000	0.00000	0.00000	0.00000
CW_BR	*	**0.18602**	**0.89605**	0.01564	**0.25285**	**0.05693**	0.00000	0.00000	0.00000	0.00000
N_BR	**0.00291**	*	**0.12840**	**0.10276**	**0.16988**	0.00693	0.00842	0.00000	0.00000	0.00000
NE_BR	**-0.00134**	**0.00394**	*	0.01000	**0.32403**	0.01505	0.00000	0.00000	0.00000	0.00000
S_BR	0.00663	**0.00556**	0.00631	*	**0.12098**	0.00347	0.00119	0.00000	0.00000	0.00000
SE_BR	**0.00087**	**0.0033**	**0.00028**	**0.00254**	*	0.04613	0.00446	0.00000	0.00000	0.00000
PT	**0.0035**	0.01657	0.00494	0.01104	0.00435	*	0.01535	0.00000	0.00000	0.00000
SP	0.01261	0.01509	0.01339	0.01131	0.00734	0.005	*	0.00000	0.00000	0.00000
CN	0.14137	0.12841	0.12935	0.14788	0.12826	0.15098	0.15072	*	0.00000	0.00000
BM	0.23586	0.25288	0.21408	0.27422	0.24535	0.25986	0.25405	0.25487	*	0.00000
YOR	0.34092	0.36694	0.29631	0.373	0.3574	0.40693	0.42053	0.42839	0.63457	*

	CUN	HU	SN	TOL	BOL	EC	COL	CR	CH	MEX	ARG	PAN	PE	PR
CW_BR	**0.01362**	**0.01098**	0.03855	**0.02306**	0.14649	0.05241	0.01696	**0.00365**	0.00764	0.10727	**-0.00299**	0.01878	0.06767	**0.00058**
N_BR	**0.01182**	**0.00494**	**0.02902**	**0.01935**	0.12723	0.04117	0.01621	**-0.00305**	0.01325	0.09259	**0.00341**	0.0214	0.05397	**0.00581**
NE_BR	**0.00479**	**0.00827**	0.03061	**0.02666**	0.14107	0.05076	0.01723	0.00734	0.0083	0.10631	**-0.00212**	0.01919	0.06756	**-0.0006**
S_BR	**0.01479**	**0.02696**	0.04048	0.03385	0.1642	0.0609	0.02583	0.00948	0.00741	0.12167	**0.00299**	0.02871	0.08782	0.01533
SE_BR	**0.01054**	**0.00104**	**0.02234**	**0.02998**	0.14317	0.05	0.01637	0.00673	**0.00353**	0.10527	**-0.00321**	0.02406	0.07104	0.00806
PT	0.03674	**0.01891**	**0.0286**	**0.0209**	0.14978	0.05313	0.0178	0.00754	**0.00171**	0.10291	**-0.00067**	0.03306	0.08	0.01098
SP	0.04539	**0.00964**	**0.0117**	**0.01785**	0.13812	0.04492	0.01904	0.00854	**-0.00107**	0.09781	**0.00236**	0.04566	0.07258	0.02316
CN	0.14086	0.10451	0.09574	0.14795	0.19599	0.15915	0.1301	0.14445	0.13671	0.15642	0.12953	0.15514	0.1435	0.15065
BM	0.30935	0.28731	0.16858	0.22647	0.16833	0.16464	0.1717	0.23016	0.24784	0.12267	0.23147	0.21135	0.12908	0.25333
YOR	0.36861	0.49232	0.44394	0.51583	0.56799	0.45572	0.3828	0.38978	0.42859	0.48348	0.40395	0.32802	0.48081	0.33347

	CW_BR	N_BR	NE_BR	S_BR	SE_BR	PT	SP	CN	BM	YOR
CW_BR	*	**0.18602**	**0.89605**	0.01564	**0.25285**	**0.05693**	0.00000	0.00000	0.00000	0.00000
N_BR	**0.00291**	*	**0.12840**	**0.10276**	**0.16988**	0.00693	0.00842	0.00000	0.00000	0.00000
NE_BR	**-0.00134**	**0.00394**	*	0.01000	**0.32403**	0.01505	0.00000	0.00000	0.00000	0.00000
S_BR	0.00663	**0.00556**	0.00631	*	**0.12098**	0.00347	0.00119	0.00000	0.00000	0.00000
SE_BR	**0.00087**	**0.0033**	**0.00028**	**0.00254**	*	0.04613	0.00446	0.00000	0.00000	0.00000
PT	**0.0035**	0.01657	0.00494	0.01104	0.00435	*	0.01535	0.00000	0.00000	0.00000
SP	0.01261	0.01509	0.01339	0.01131	0.00734	0.005	*	0.00000	0.00000	0.00000
CN	0.14137	0.12841	0.12935	0.14788	0.12826	0.15098	0.15072	*	0.00000	0.00000
BM	0.23586	0.25288	0.21408	0.27422	0.24535	0.25986	0.25405	0.25487	*	0.00000
YOR	0.34092	0.36694	0.29631	0.373	0.3574	0.40693	0.42053	0.42839	0.63457	*

Table 24: F_{ST} genetic distances results (bellow the diagonal) and its P-Values (above the diagonal) with 19 Y-SNPs between our four admixed populations (Cundinamarca, Huila, Santander del Norte and Tolima) and other Latin American admixed groups; a European population; a Native American group (Barí-Motilon) and an African group. Non-statistically significant values are in bold.

	CUN	HU	SN	TOL	COL	ARG	PR	BOL	CH	NI	NN_B	NE_B	CW_B	S_B	SE_B	BM	SP	ANG
CUN	*	**0.48659**	**0.84487**	**0.67300**	**0.24453**	**0.39026**	**0.29948**	0.00119	0.03505	**0.85378**	**0.05910**	0.00059	0.02188	0.01089	0.01356	0.00000	0.00089	0.00000
HU	**-0.00908**	*	**0.62964**	**0.93189**	**0.11345**	**0.71755**	**0.17662**	**0.30690**	**0.66102**	**0.72993**	**0.08484**	0.00020	0.03831	0.01495	0.01337	0.00000	**0.26047**	0.00000
SN	**-0.02961**	**-0.02503**	*	**0.98564**	**0.10702**	**0.27344**	**0.56004**	0.02178	**0.09375**	**0.78646**	**0.08227**	0.00040	0.01931	0.00752	0.00871	0.00000	0.00277	0.00000
TOL	**-0.01773**	**-0.0541**	**-0.04925**	*	**0.16424**	**0.62043**	**0.50361**	**0.17860**	**0.58806**	**0.86882**	**0.17701**	0.00030	**0.05257**	0.02287	0.02198	0.00000	0.04376	0.00000
COL	**0.00781**	**0.03222**	**0.02252**	**0.01787**	*	0.00000	0.00653	0.00000	0.00000	0.00158	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ARG	**0.00009**	**-0.01946**	**0.00629**	**-0.01038**	0.04023	*	0.00000	0.00000	0.01960	0.04673	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000
PR	**0.00479**	**0.02139**	**-0.00657**	**-0.00554**	0.01507	0.0421	*	0.00000	0.00000	0.01030	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
BOL	0.12418	**0.00333**	0.07293	**0.02172**	0.10395	0.06128	0.10208	*	0.03178	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00218	0.00000
CH	0.04768	**-0.01794**	**0.02878**	**-0.01065**	0.05625	0.02053	0.05402	0.02412	*	0.00941	0.00040	0.00000	0.00030	0.00000	0.00000	0.00000	0.01841	0.00000
NI	**-0.01549**	**-0.01838**	**-0.01439**	**-0.02006**	0.01895	0.0064	0.01374	0.06478	0.02678	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
N_B	**0.02533**	**0.03357**	**0.02305**	**0.01365**	0.03573	0.04551	0.02858	0.09647	0.03899	0.03369	*	0.00000	**0.10682**	0.00198	0.03336	0.00000	0.00000	0.00000
NE_B	0.08052	0.13479	0.09792	**0.10351**	0.09579	0.12424	0.08908	0.21602	0.13205	0.10855	0.03157	*	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000
CW_B	0.04451	0.05834	0.04934	0.03985	0.06	0.06001	0.05924	0.11946	0.05863	0.05452	**0.00408**	0.02972	*	**0.10751**	**0.15820**	0.00000	0.00000	0.00000
S_B	0.05352	0.08085	0.05865	0.05533	0.06882	0.07472	0.06419	0.14581	0.08272	0.06812	0.01143	0.0159	**0.0045**	*	0.01970	0.00000	0.00000	0.00000
SE_B	0.04536	0.06943	0.05442	0.04833	0.05675	0.06809	0.05567	0.13166	0.06412	0.05882	0.00459	0.01535	**0.00277**	0.00625	*	0.00000	0.00000	0.00000
BM	0.53625	0.69207	0.51848	0.58461	0.43537	0.46848	0.36016	0.60282	0.54258	0.414	0.391	0.40959	0.45314	0.43417	0.41469	*	0.00000	0.00000
SP	0.13341	**0.01213**	0.11772	0.05599	0.1117	0.04415	0.12975	0.0337	0.02897	0.07554	0.10943	0.221	0.12937	0.15025	0.1369	0.64725	*	0.00000
ANG	0.58296	0.69664	0.63207	0.65655	0.34609	0.47847	0.40248	0.65141	0.60076	0.45116	0.38637	0.36856	0.45987	0.41833	0.38032	0.83881	0.64184	*

Table 25: R_{ST} genetic distances results (bellow the diagonal) and its P-Values (above the diagonal) with 19 Y-STRs between our six populations that presented Eurasian haplogroups (Guainía, Paeces, Cundinamarca, Huila, Santander del Norte and Tolima) and other European populations and an group from Middle East. Non-statistically significant values are in bold (Continues in the next pages).

	GUA	PAE	CUN	HUI	SN	TOL	BEL	CRO	CZE	ENG	GER	GRE	HUN	IRE	BR_IT
GUA	*	**0.18978**	**0.09633**	**0.32937**	**0.24819**	**0.09801**	**0.07801**	**0.17523**	**0.16097**	**0.05920**	**0.15216**	**0.35798**	**0.22265**	0.03148	0.02307
PAE	**0.11204**	*	0.00228	0.04099	**0.56618**	0.03960	0.00168	0.00000	0.00020	0.00396	0.00010	0.00030	0.00149	**0.08029**	0.00069
CUN	**0.10029**	0.15651	*	**0.36046**	**0.10692**	**0.10959**	**0.11821**	0.00020	**0.12900**	0.04168	0.04643	**0.14979**	0.04524	0.00149	0.02604
HUI	**0.01795**	0.06834	**0.00443**	*	**0.78250**	**0.39481**	**0.92892**	0.00743	**0.15404**	**0.80012**	**0.13910**	**0.12108**	**0.08791**	**0.18563**	**0.17137**
SN	**0.0184**	**-0.02463**	**0.02637**	**-0.02936**	*	**0.61994**	0.02772	0.00000	0.00139	**0.05950**	0.00040	0.00050	0.00119	**0.10445**	0.03317
TOL	**0.09905**	0.07683	**0.03926**	**0.00034**	**-0.01097**	*	**0.16127**	0.00010	0.02099	**0.14405**	0.00218	0.00931	0.00782	**0.15939**	**0.96485**
BEL	**0.09395**	0.18999	**0.02095**	**-0.03144**	0.04179	**0.01721**	*	0.00000	0.00030	**0.58390**	0.00000	0.00000	0.00000	0.00376	0.00000
CRO	**0.0698**	0.29576	0.17385	0.14779	0.20796	0.21673	0.18443	*	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CZE	**0.05757**	0.2096	**0.02118**	**0.02732**	0.08594	0.06548	0.03227	0.09054	*	0.00099	**0.65875**	0.04653	**0.39026**	0.00000	0.00000
ENG	**0.12002**	0.17388	0.04503	**-0.02612**	**0.03251**	**0.02332**	**-0.00201**	0.18094	0.04481	*	0.00020	0.00000	0.00000	0.04554	0.00000
GER	**0.05797**	0.24841	0.03643	**0.02976**	0.10164	0.09975	0.04212	0.08839	**-0.00328**	0.0509	*	0.00030	**0.08514**	0.00000	0.00000
GRE	**0.00757**	0.2329	**0.01558**	**0.02994**	0.09915	0.06989	0.04809	0.09611	0.00973	0.06672	0.02225	*	0.01089	0.00000	0.00000
HUN	**0.03897**	0.21233	0.0394	**0.04023**	0.09487	0.08724	0.05258	0.0609	**0.00003**	0.0611	**0.006**	0.0117	*	0.00000	0.00000
IRE	0.19808	**0.09093**	0.13047	**0.0255**	**0.02251**	**0.02388**	0.04981	0.23662	0.12284	0.02657	0.14107	0.14016	0.12687	*	0.00129
BR_IT	0.16535	0.19729	0.04594	**0.02154**	0.03984	**-0.02737**	0.04435	0.2868	0.10874	0.04924	0.13147	0.10819	0.13259	0.05525	*

(Continues in the next 3 pages)

	PU_IT	SI_IT	NET	POL	POR	AR_SP	AS_SP	MA_SP	SWI	WAL	MAC	DEN	LEB
GUA	**0.15563**	**0.11831**	**0.10326**	0.03881	**0.36135**	**0.05594**	**0.06980**	0.04604	**0.20939**	0.00465	**0.23909**	**0.12474**	**0.38115**
PAE	0.00020	0.00010	0.01168	0.00000	0.00277	0.00594	0.00356	0.00634	0.00594	0.00703	0.00010	0.00228	0.00030
CUN	**0.72864**	**0.59192**	**0.16721**	0.00208	**0.25740**	0.02950	**0.19275**	**0.05148**	**0.19157**	0.00000	0.00901	**0.20731**	0.02188
HUI	**0.27235**	**0.13177**	**0.85190**	0.00208	**0.78467**	**0.62053**	**0.75428**	**0.68666**	**0.94931**	0.01703	0.01782	**0.33224**	0.03812
SN	0.00040	0.00099	**0.05554**	0.00000	**0.06910**	**0.13415**	**0.09227**	**0.18454**	0.04415	0.00347	0.00000	0.00495	0.00000
TOL	**0.05227**	0.04653	**0.15107**	0.00010	**0.30096**	**0.41441**	**0.58489**	**0.61073**	**0.16295**	**0.06910**	0.00178	0.02366	0.00040
BEL	0.00000	0.00000	**0.35303**	0.00000	0.01376	0.00168	0.00317	0.00604	**0.31997**	0.00000	0.00000	0.00000	0.00000
CRO	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CZE	0.00050	0.00000	0.00455	0.00010	0.00198	0.00000	0.00000	0.00000	0.00713	0.00000	0.00000	**0.09197**	0.00000
ENG	0.00000	0.00000	**0.23027**	0.00000	0.01832	0.04841	0.02158	0.02515	**0.17879**	0.00020	0.00000	0.00040	0.00000
GER	0.00000	0.00000	0.00050	0.00000	0.00020	0.00000	0.00000	0.00000	0.00050	0.00000	0.00000	0.01703	0.00000
GRE	0.00366	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
HUN	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
IRE	0.00000	0.00000	0.00604	0.00000	0.00327	0.02990	0.00287	0.01861	0.00277	**0.55123**	0.00000	0.00000	0.00000
BR_IT	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00119	0.01040	0.00000	0.00000	0.00000	0.00000	0.00000

	GUA	PAE	CUN	HUI	SN	TOL	BEL	CRO	CZE	ENG	GER	GRE	HUN	IRE	BR_IT
PU_IT	**0.04806**	0.21705	**-0.01071**	**0.0103**	0.07533	**0.03422**	0.03335	0.17028	0.02865	0.05496	0.04878	0.01226	0.04401	0.11992	0.06214
SI_IT	**0.06381**	0.22458	**-0.0059**	**0.02558**	0.07068	0.03537	0.04862	0.1953	0.05067	0.06936	0.07048	0.02491	0.05992	0.12724	0.05957
NET	**0.0769**	0.13186	**0.01794**	**-0.03062**	**0.03219**	**0.02614**	**0.00032**	0.15242	0.02542	**0.0028**	0.03377	0.04853	0.04552	0.04857	0.05855
POL	0.12443	0.29463	0.07819	0.13058	0.16216	0.15482	0.13024	0.10082	0.03915	0.13317	0.041	0.06553	0.03177	0.20724	0.19913
POR	**0.00603**	0.17421	**0.00784**	**-0.02569**	**0.02989**	**0.00577**	0.01466	0.17009	0.03521	0.02033	0.04521	0.0373	0.04747	0.06029	0.03589
AR_SP	**0.11859**	0.17889	0.04452	**-0.01367**	**0.01841**	**-0.00146**	0.01524	0.23451	0.07527	0.00959	0.08688	0.08032	0.08859	0.02461	0.02488
AS_SP	**0.1044**	0.17408	**0.01176**	**-0.02041**	**0.02303**	**-0.00897**	0.01058	0.23714	0.06115	0.01218	0.07699	0.06637	0.08349	0.04198	0.01774
MA_SP	0.12857	0.17191	**0.03351**	**-0.01662**	**0.014**	**-0.00978**	0.01459	0.23596	0.07677	0.01464	0.09137	0.07671	0.09065	0.03115	0.01382
SWI	**0.03564**	0.16688	**0.01536**	**-0.03702**	0.03876	**0.02363**	**0.00051**	0.15515	0.02312	**0.00364**	0.02691	0.04052	0.04599	0.05813	0.05193
WAL	0.30822	0.18366	0.16726	0.07834	0.06459	**0.03342**	0.06389	0.29981	0.15498	0.04404	0.17042	0.16331	0.16	**-0.00324**	0.063
MAC	**0.03372**	0.26821	0.07638	0.09454	0.16354	0.13192	0.11779	0.11016	0.05884	0.13733	0.08067	0.03732	0.05222	0.20905	0.19267
DEN	**0.06723**	0.21044	**0.01349**	**0.00133**	0.08529	0.07074	0.02343	0.1305	**0.00709**	0.03296	0.01059	0.02745	0.02927	0.11384	0.09902
LEB	**0.0013**	0.30084	0.03916	0.05645	0.13331	0.11686	0.09316	0.16074	0.07099	0.11974	0.07499	0.02424	0.06627	0.20089	0.13653

	PU_IT	SI_IT	NET	POL	POR	AR_SP	AS_SP	MA_SP	SWI	WAL	MAC	DEN	LEB
PU_IT	*	**0.16573**	0.00010	0.00000	0.00099	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000
SI_IT	**0.00254**	*	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
NET	0.0364	0.05659	*	0.00000	0.00495	0.00020	0.00040	0.00069	**0.42560**	0.00000	0.00000	0.02515	0.00000
POL	0.10851	0.1196	0.11631	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
POR	0.0244	0.02851	0.02611	0.10978	*	0.02317	0.01475	**0.05158**	0.02445	0.00000	0.00000	0.00020	0.00000
AR_SP	0.05814	0.05783	0.03305	0.1625	0.01308	*	**0.10346**	**0.77477**	0.00010	0.00010	0.00000	0.00000	0.00000
AS_SP	0.03598	0.03789	0.02357	0.15173	0.01288	**0.00279**	*	**0.66944**	0.00069	0.00000	0.00000	0.00000	0.00000
MA_SP	0.04774	0.04851	0.02883	0.16528	**0.01074**	**-0.00258**	**-0.0017**	*	0.00069	0.00000	0.00000	0.00000	0.00000
SWI	0.03076	0.04838	**-0.00043**	0.11589	0.01376	0.02529	0.01911	0.02513	*	0.00000	0.00000	0.01485	0.00000
WAL	0.13981	0.1486	0.07523	0.23457	0.08857	0.03696	0.04937	0.04715	0.07936	*	0.00000	0.00000	0.00000
MAC	0.06188	0.09299	0.10228	0.10904	0.09938	0.1717	0.14835	0.1606	0.10541	0.25153	*	0.00000	0.00000
DEN	0.02946	0.05428	0.0111	0.08502	0.04219	0.07139	0.05394	0.06974	0.01043	0.13253	0.07986	*	0.00000
LEB	0.03092	0.03151	0.10157	0.13133	0.06459	0.11679	0.10105	0.10957	0.08173	0.22004	0.07933	0.07641	*

Table 26: F_{ST} genetic distances results (bellow the diagonal) and its P-Values (above the diagonal) with 12 Y-SNPs between our six populations that presented Eurasian haplogroups (Guainía, Paeces, Cundinamarca, Huila, Santander del Norte and Tolima) and other European populations and two groups from Middle East. Non-statistically significant values are in bold (Continues in the next pages).

	GUA	PAE	CUN	HU	SN	TOL	AST	BC	GAL	PT	POL	GER	CRO
GUA	*	0.04940	**0.57479**	**0.21750**	**0.19325**	**0.13642**	**0.39659**	0.00129	**0.05653**	**0.27859**	0.00693	0.00337	0.02802
PAE	0.24744	*	**0.19157**	**0.58351**	**0.66785**	**0.80705**	**0.33135**	**0.08554**	**0.31363**	**0.17216**	0.00000	0.00000	0.00030
CUN	**-0.03704**	**0.04405**	*	**0.40382**	**0.30660**	**0.41441**	**0.90941**	0.00010	**0.14197**	**0.83705**	0.00000	0.00000	0.00010
HU	**0.11774**	**-0.02791**	**-0.00019**	*	**0.75230**	**0.82150**	**0.74250**	**0.09197**	**0.63063**	**0.42243**	0.00000	0.00000	0.00010
SN	**0.09412**	**-0.03029**	**0.00692**	**-0.03886**	*	**0.93545**	**0.54915**	0.00594	**0.42125**	**0.31492**	0.00000	0.00000	0.00000
TOL	**0.11496**	**-0.05164**	**-0.00404**	**-0.05685**	**-0.03766**	*	**0.78764**	0.02940	**0.79250**	**0.48688**	0.00000	0.00000	0.00010
AST	**0.01639**	**0.01128**	**-0.03672**	**-0.03076**	**-0.01101**	**-0.03099**	*	0.00020	**0.69795**	**0.88892**	0.00000	0.00000	0.00010
BC	0.56406	**0.04703**	0.30701	**0.06341**	0.13538	0.11178	0.20929	*	0.00000	0.00000	0.00000	0.00000	0.00000
GAL	**0.13461**	**0.00429**	**0.02183**	**-0.0198**	**-0.00275**	**-0.02326**	**-0.01285**	0.10907	*	**0.10029**	0.00000	0.00000	0.00000
PT	**0.01622**	**0.03413**	**-0.02151**	**-0.00203**	**0.00429**	**-0.00796**	**-0.01953**	0.18981	**0.01066**	*	0.00000	0.00000	0.00000
POL	0.36249	0.39015	0.31868	0.41756	0.37245	0.34715	0.34478	0.55223	0.3705	0.33248	*	**0.53836**	0.00000
GER	0.40537	0.43887	0.35587	0.46599	0.41453	0.39382	0.38632	0.61237	0.40396	0.36022	**-0.002**	*	0.00000
CRO	0.20161	0.24018	0.17031	0.23748	0.22062	0.1825	0.16393	0.43355	0.21028	0.18691	0.0931	0.12549	*

(Continues in the next 3 pages)

	HUN	DEN	UK	GRE	FR	N_IT	S_IT	SIC	TK	LEB
GUA	0.00881	**0.05108**	**0.44075**	**0.14117**	**0.08890**	**0.07039**	**0.44996**	**0.59717**	**0.68845**	**0.71379**
PAE	0.00079	0.02178	**0.11019**	0.00109	**0.39729**	**0.28532**	0.00317	0.01297	0.00069	0.00000
CUN	0.00000	0.00307	**0.93387**	0.04158	**0.10682**	**0.37976**	**0.45273**	**0.80725**	**0.09910**	0.01168
HU	0.00010	0.02198	**0.23265**	0.00119	**0.59598**	**0.54846**	0.01208	**0.05108**	0.00178	0.00000
SN	0.00000	0.00139	**0.12603**	0.00030	**0.52104**	**0.32017**	0.00109	0.01059	0.00020	0.00000
TOL	0.00020	0.02594	**0.38056**	0.00228	**0.83150**	**0.81655**	0.01673	**0.05029**	0.00198	0.00000
AST	0.00000	0.01119	**0.89338**	0.02366	**0.65875**	**0.92011**	**0.08870**	**0.22245**	0.00535	0.00010
BC	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
GAL	0.00000	0.00000	**0.12375**	0.00000	**0.18850**	**0.27195**	0.00000	0.00000	0.00000	0.00000
PT	0.00000	0.00000	**0.57895**	0.00099	0.02525	**0.05762**	0.00040	0.02525	0.00020	0.00000
POL	**0.28007**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
GER	**0.12771**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CRO	0.00208	0.00020	0.00089	0.00030	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

	GUA	PAE	CUN	HU	SN	TOL	AST	BC	GAL	PT	POL	GER	CRO
HUN	0.26511	0.28604	0.21062	0.32008	0.27097	0.24263	0.23877	0.54466	0.28058	0.23659	**0.0023**	**0.01068**	0.05938
DEN	**0.17542**	0.14072	0.11826	0.11773	0.13055	0.08353	0.07864	0.2696	0.10729	0.11501	0.21764	0.25408	0.04432
UK	**-0.00317**	**0.06355**	**-0.04215**	**0.02871**	**0.03322**	**0.00369**	**-0.03383**	0.32798	**0.0226**	**-0.00794**	0.28261	0.3207	0.11344
GRE	**0.08135**	0.21893	0.06413	0.19052	0.17563	0.15079	0.07426	0.5111	0.147	0.10273	0.30078	0.33441	0.11888
FR	**0.11657**	**-0.00529**	**0.02968**	**-0.01754**	**-0.00694**	**-0.02541**	**-0.01171**	0.10563	**0.00528**	0.02263	0.36887	0.40286	0.21647
N_IT	**0.11165**	**0.00904**	**-0.0007**	**-0.01384**	**0.00296**	**-0.02322**	**-0.01991**	0.11506	**0.002**	**0.01163**	0.34202	0.3691	0.19853
S_IT	**-0.006**	0.12939	**-0.00202**	0.09168	0.08988	0.06446	**0.02324**	0.27525	0.08329	0.03232	0.25609	0.27213	0.13519
SIC	**-0.0408**	0.11354	**-0.01832**	**0.06336**	0.06369	**0.04789**	**0.01007**	0.26779	0.06787	0.01673	0.26744	0.28488	0.1463
TK	**-0.04653**	0.20074	**0.03231**	0.15321	0.14533	0.12344	0.08174	0.46078	0.16405	0.07665	0.25085	0.27423	0.14207
LEB	**-0.04355**	0.2759	0.07165	0.21234	0.20624	0.19428	0.12521	0.476	0.21159	0.12085	0.31337	0.3296	0.19635

	HUN	DEN	UK	GRE	FR	N_IT	S_IT	SIC	TK	LEB
HUN	*	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
DEN	0.15879	*	**0.01851**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
UK	0.17672	**0.0677**	*	**0.27918**	**0.06405**	**0.34096**	**0.81645**	**0.84041**	**0.08197**	0.00743
GRE	0.20393	0.12571	**0.01171**	*	0.00000	0.00010	0.03950	0.00871	0.00119	0.00238
FR	0.27962	0.11433	**0.03527**	0.17544	*	**0.09633**	0.00000	0.00000	0.00000	0.00000
N_IT	0.25403	0.10788	**0.00196**	0.13168	**0.00821**	*	0.00000	0.00000	0.00000	0.00000
S_IT	0.17608	0.11599	**-0.01507**	0.03002	0.09387	0.06408	*	**0.28977**	0.04099	0.00010
SIC	0.18178	0.1176	**-0.01626**	0.05223	0.07672	0.05141	**0.00121**	*	0.04376	0.00010
TK	0.16989	0.1554	**0.03093**	0.07459	0.17255	0.14183	0.01668	0.01796	*	**0.28987**
LEB	0.23631	0.2131	0.06598	0.07235	0.22224	0.18918	0.03642	0.04276	**0.0031**	*