Y-chromosomal characterization of the Colombian population

Maria Inês Pinto da Silva Paiva Machado
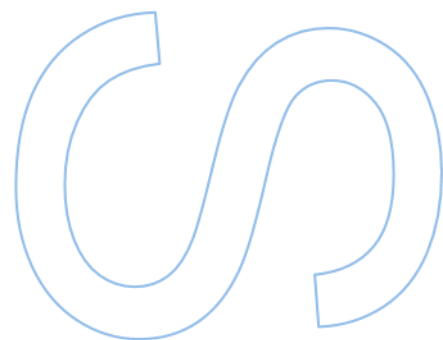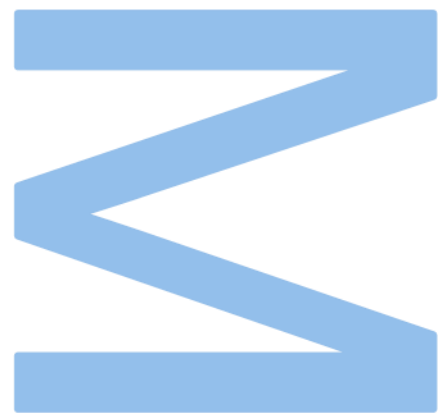Genética Forense
Departamento de Biologia
2023

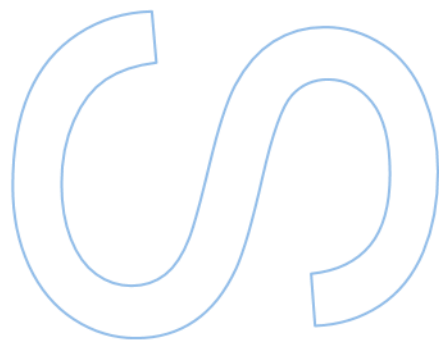**Orientador**
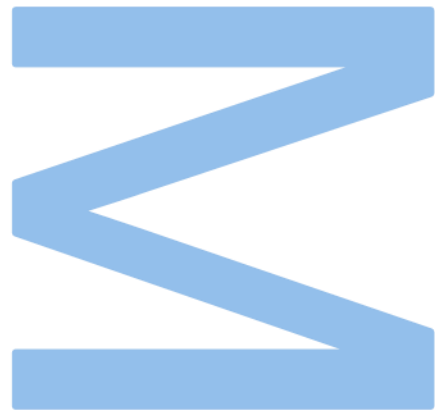Vânia Pereira, Associate Professor, Retsmedicinsk Institut, Copenhagen University

Coorientadores
Verónica Gomes, Junior Researcher, i3S, IPATIMUP
Leonor Gusmão, Associate Professor, UERJ

''Para ser grande, sê inteiro: nada
Teu exagera ou exclui.
Sê todo em cada. Põe quanto és
No mínimo que fazes.
Assim em cada lago a lua toda
Brilha, porque alta vive.''

Ricardo Reis em ''Poesia dos Outros Eus.'

Agradecimentos

À minha orientadora, Vânia Pereira, um obrigado, em primeiro lugar, por me ter permitido elaborar o trabalho laboratorial e de investigação científica ao seu lado no Instituto Retsmedicinsk em Copenhaga. Quero deixar-lhe um especial agradecimento por me ter acompanhado, diariamente, tão de perto, visando sempre pelo rigor; por me ter transmitido o conhecimento laboratorial e teórico sobre o tema em questão; por me ter incutido o espírito proativo e crítico inerente à elaboração desta dissertação e por todos os conselhos partilhados.

À, Verónica Gomes, minha coorientadora, agradeço por me ter recebido de braços abertos no i3s aquando do meu regresso a Portugal; pelo apoio incondicional; pela disponibilidade total em me ajudar, especialmente, na análise dos dados e pela transmissão de conhecimento científico.

À Professora Leonor Gusmão, minha coorientadora, um obrigado por ter aceitado fazer parte desta caminhada, partilhando sempre o seu imensurável conhecimento sobre o tema e por ter acrescentado um valor maior à minha dissertação.

A todo os elementos do Instituto Retsmedicinsk, um obrigado por me terem recebido calorosamente e por terem partilhado, diariamente, comigo todo o seu conhecimento científico e por me terem amparado com a sua sabedoria e amizade durante os meus momentos de insegurança.

À minha família, um obrigado, por, incondicionalmente, estarem presentes e apoiarem-me nos desafios a que me proponho cumprir.

Ao, Zé, meu parceiro, um obrigado por me acompanhar e amparar nesta longa caminhada.

Aos meus amigos de sempre e aqueles que se cruzaram comigo durante esta jornada, um obrigado pelo apoio incondicional.

Obrigada!

Resumo

A Colômbia é um dos países pertencentes à América do Sul cuja composição genética é pautada por uma elevada diversidade e heterogeneidade, dada a sua complexa história marcada fortemente pela colonização Europeia e, consequente, tráfico de escravos, e pelo fluxo migratório desde o século XX.

O presente estudo consistiu na caraterização das linhagens masculinas de indivíduos miscigenados pertencentes à Colômbia com o objetivo de expandir o conhecimento do fundo genético desta população e compreender os eventos históricos e demográficos caraterísticos deste país. Recorrendo à análise dos polimorfismos encontrados na região não recombinante do cromossoma Y, analisaram-se 167 indivíduos masculinos, miscigenados, não aparentados, pertencentes a duas regiões naturais da Colômbia, as Caraíbas (98 indivíduos) e os Andes (69 indivíduos). Foram genotipados 859 Y-SNPs através de sequenciação massiva paralela (MPS). Para além disso, os perfis de Y-STR destas amostras, previamente, genotipados com 23 Y-STRs no âmbito de outros estudos foram utilizados.

Os resultados obtidos com os dados de Y-STR permitiram identificar um total de 159 haplótipos singulares (92 no Caribe e 67 nos Andes) não havendo partilha de haplótipos entre as duas regiões. Para além do mais, a diversidade haplotípica calculada para ambas as regiões foi elevada (≥99.94%). Relativamente aos resultados obtidos com recurso aos Y-SNPs, 62 haplogrupos foram identificados nas populações em análise. O haplogrupo mais frequente quer na região do Caribe, quer na região dos Andes foi o haplogrupo Euroasiático R (39.80% e 62.32%, respetivamente). As origens das linhagens encontradas no Caribe foram, maioritariamente, Eurasiáticas (78.57%), seguida de linhagens com origem Africana Subsariana (16.33%) e, finalmente, apenas 5.10% dos haplogrupos foram de origem Nativa. Por sua vez, nos Andes, a origem das linhagens mais frequentes foi Eurasiática (89.86%), seguida de linhagens com origem Nativa (7.25%) e, com apenas 2.90% de linhagens com origem Africana subsariana. As distâncias genéticas baseadas nos valores de R_{ST} e F_{ST} e os respetivos valores de p, foram calculadas entre as duas populações.

As análises comparativas efetuadas entre a nossa amostra e outras populações miscigenadas da América do Sul mostraram que a região do Caribe encontra-se, apesar da alta influência Europeia, mais próxima de populações com maior número de linhagens Africanas Subsarianas comparativamente aos Andes. Em contraste, os Andes colombianos encontram-se muito próximos de outras populações miscigenadas

altamente impactadas pela influência Europeia. No que toca à comparação com populações eurasiáticas, o Caribe e os Andes encontram-se muito próximos de populações pertencentes ao Sul da Europa tais como Portugal, Espanha e França. Ambas as populações colombianas apresentam-se muito afastadas de populações pertencentes ao Médio Oriente. Um estudo mais intensivo de linhagens R-M269 foi executado demonstrando que quer no Caribe, quer nos Andes, os indivíduos pertencentes a linhagens de R-M269 são altamente influenciados pela linhagem R-S116, muito frequente na Península Ibérica e França.

Os resultados obtidos neste trabalho ilustram bem a diversidade genética das populações colombianas. Apesar de ambas as regiões serem altamente influenciadas pela componente Europeia, existem diferenças significativas no que toca à componente Africana mais frequente no Caribe. Pode-se afirmar que os resultados demonstrados estão de acordo com o contexto histórico do país, corroborando a clara influência da colonização Ibérica no processo de miscigenação do país e atestando, a influência demarcada do tráfico de escravos nas regiões portuárias do Caribe.

Palavras-chave: Genética Populacional, Cromossoma Y, Sequenciação Massiva Paralela, Y-STRs, Y-SNPs, Colômbia, Populações miscigenadas, América do Sul

Abstract

Colombia is a South American country with a highly diverse and heterogeneous genetic background, given its complex history, which has been strongly marked by the European colonization and, consequently, the slave trade, as well as by more recent migratory flows since the 20th century.

The aim of this study was to characterize the male lineages of admixed individuals from Colombia, in order to expand our knowledge of the genetic background of this population and understand the historical and demographic events characteristic of this country. A sample of 167 unrelated, admixed male individuals from two natural regions of Colombia, the Caribe (98 individuals) and the Andes (69 individuals) was analysed for polymorphisms located in the non-recombinant region of the Y chromosome. A total of 859 Y-SNPs were genotyped using massively parallel sequencing (MPS). In addition, the Y-STR profiles of these samples, previously genotyped with 23 Y-STRs in the context of other studies, were used in data analyses.

The results obtained with the Y-STR data allowed us to identify a total of 159 unique haplotypes (92 in the Caribe and 67 in the Andes), with no shared haplotypes between the two regions. The haplotypic diversity calculated for both regions was high (≥99.94%). With regard to the results obtained using Y-SNPs, 62 haplogroups were identified in the populations under analysis. The most frequent haplogroup in both the Caribbean and Andean regions was the European haplogroup R (39.80% and 62.32%, respectively). Most lineages found in the Caribe were of European origin (78.57%), followed by lineages with sub-Saharan African origin (16.33%) and, finally, only 5.10% of the haplogroups were of Native origin. In the Andes, most lineages were from European origin (89.86%), followed by lineages with Native origin (7.25%) and, finally, only 2.90% of the lineages have sub-Saharan African origin. Genetic distances based on R_{ST} and F_{ST}, and the corresponding non-differentiation p-values, were calculated between the two populations.

Comparative analyses carried out between our dataset and other admixed populations in South America showed that, despite being highly influenced by the Eurasian component, the Caribbean region is closer to populations with a greater number of sub-Saharan African lineages than the Andes. The Colombian Andes is close to other admixed populations highly impacted by European influence. When it comes to comparisons with Eurasian populations, the Caribe and the Andes are very close to populations belonging to Southern Europe such as Portugal, Spain and France. Both Colombian populations are far from populations belonging to the Middle East. A more

exhaustive study of R-M269 sub-lineages was carried out, showing that in both the Caribe and the Andes most individuals belong to haplogroup R-S116, which is very common in the Iberian Peninsula and France.

The results obtained in this study illustrate the genetic diversity of Colombian populations. Although both regions have high European ancestry, there are significant differences when it comes to the African component, which is more frequent in the Caribe. These results are aligned with the country's historical context, corroborating the influence of Iberian colonization on the admixture process and attesting to the marked influence of the slave trade in the Caribe port regions.

Keywords: Population Genetics, Y Chromosome, Massively Parallel Sequencing, Y-STRs, Y-SNPs, Colombia, Admixed populations, South America

Index

List of Figures

List of Tables

List of Abbreviations

ANC-A	"Southern Native American" or "Ancestral A"
ANC-B	"Northern Native American" or "Ancestral B"
Bp	Base pairs
Cal.BP	Calibrated years before the present
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
emPCR	Emulsion Polymerase Chain Reaction
FET	Field-Effect Transitor
gDNA	Genomic DNA
ISOGG	International Society of Genetic Genealogy
ISPs	Ion Sphere Particles
MDS	Multidimensional Scaling
MPS	Massively Parallel Sequencing
MSY	Male-Specific region of Ychromosome
mtDNA	Mitochondrial DNA
NRY	Non-recombing region of Y chromosome
PAR 1	Pseudoautosomal region 1
PAR 2	Pesudoautosomal region 2
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
pM	Pico Molar
qPCR	Quantitative Polymerase Chain Reaction
RFLP	Restriction Fragment Length Polymorphism
RNA	Ribonucleic acid
SMM	Single-Step Mutation Model
SNP	Single Nucleotide Polymorphism
SRY	Sex-determining Y gene

STD1	Standard 1
STD2	Standard 2
STR	Short Tandem Repeat
UDG	Uracil-DNA glycosylases
VNTR	Variable number of tandem repeats
YCC	Y Chromosome Consortium
YHRD	Y Chromosome Haplotype Reference Database
Y-SNP	Y-Chromosomal Single Nucleotide Polymorphism
Y-STR	Y-Chromosomal Short Tandem Repeat

1. Introduction

1.1. Population and Forensic Genetics

Understanding what we are genetically and how that affects our adaptation to the environment has always been a topic of interest among the scientific community. Human migrations, to already occupied habitats, give rise to the interbreeding of individuals from different populations (admixture), contributing to an increase in human genetic diversity. This diversity provides the ability to resist and constantly adapt to the different environments to which humans are exposed.

Population Genetics is a biological field that gives insights about how genetic variability among individuals and populations of the same species occur and evolve across time and space, as noted by Millstein & Skipper, 2006 and Griffiths et al., 2008. Population Genetics aims to study the genetic composition of populations and infer the mechanisms that shaped the observed variability. The evolutionary forces responsible for the genetic variability within and between populations consist of mutation, migration, genetic drift, natural selection and non-random mating (Motoo et al., 2011; Staub et al., 2009). This knowledge enhances our comprehension of the historical and present genetic background of populations, which have an impact on other scientific disciplines such as, for example, anthropology (Hodgson et al., 2010), medicine (Assyuhada et al., 2023), phylogeography (Underhill et al, 2001), archaeogenetics (Sokal, 2001), and forensic genetics (National Research Council (US) Committee on DNA Forensic Science, 1996).

Forensic Genetics is a science field that applies "genetics to human and non-human material for the resolution of legal conflicts" (Amorim, 2019; Forensic Science International: Genetics, 2007). The ABO blood group system, identified by Karl Landsteiner in 1900, was the first known human genetic polymorphism[1] that allowed individuals to be distinguished based on their blood system classification (Landsteiner et al., 1940). Within the forensic field, Lattes, in 1915, described the use of ABO genotyping to resolve a paternity case and the method became a standard technique for identification in forensic laboratories (Goodwin et al., 2007).

As time went by and knowledge in this field progressed, deoxyribonucleic acid (DNA) polymorphisms (also known as DNA markers) were identified with the development of more refined laboratory techniques. Variable Number of Tandem Repeat

---

[1] A variation in the DNA sequence that occurs in a population with a frequency of 1 % or higher (Brookes AJ, 1999).

(VNTR) loci[2] (Nakamura et al., 1987), due to their high variability, provide greater discriminatory power, enabling human identification in routine forensics (Weir et al., 1993). This type of marker is genotyped by restriction fragment length polymorphism (RFLP) analysis (Botstein et al., 1980; Kruglyak, 1997; Wyman & White, 1980). Using this technique, variations in DNA homologous sequences are identified by fragments of different lengths after the DNA samples have been digested with restriction endonucleases (Botstein et al., 1980).

One of the major milestones was the invention of Polymerase Chain Reaction (PCR), first discovered and published by Mullis and collaborators in 1986. PCR allows the amplification of a specific DNA fragment from a biological sample generating millions of copies of the same genetic material (Butler, 2005). Human identification has evolved exponentially since then, with the identification of various genetic markers. Short Tandem Repeats (STRs), which are small tandemly repetitive DNA sequences (Tautz, 1998) and Single Nucleotide Polymorphisms (SNPs) that are genomic variants at a single base position in the DNA (Kruglyak, 1997; Wang, 1998), allow, until this day, the generation of DNA profiles in different research and casework fields such as population and forensic genetics.

In order to assess and comprehend genetic diversity, various panels of genetic markers, featuring different combinations of STRs or SNPs, are routinely used in population genetics. These tools allow to investigate the mechanisms underlying diversity in order to connect and contextualize the history of the human population. Genotyping these markers in a biological sample provides a DNA profile belonging to a given individual. In forensic investigations, when a biological sample is collected from a crime scene, its DNA profile is generated and compared with a reference sample's profile (e.g., it can be the sample of a given suspect) (Bond J.W., 2007; Marano & Fridamn, 2019). Population genetics plays a crucial role in the statistical evaluation of genetic evidence obtained in the context of a forensic investigation (Amorim & Budowle et al., 2016). Therefore, DNA databases that accurately describe the composition of reference populations are essential for estimating the frequency of a genetic profile and correctly communicating the weight of evidence (National Research Council (US) Committee, 1996; Pannerchelvam & Norazmi, 2003; Vullo et al., 2015).

---

[2] Position within the chromosome where a particular gene or genetic marker is located (Nakamura et al., 1987).

1.2. Recombining and non-recombining genomes

Meiotic recombination is a molecular mechanism that contributes to the genetic diversity in sexual reproducing organisms such as humans (Hunter, 2015). Characterized by a rearrangement of the genetic material during meiosis it allows an almost endless number of alleles[3] combinations (Hunter, 2015). This rearrangement occurs between generations and is, in part, one of the underlying causes of each individual carrying a unique genome, with the exception of monozygotic twins (Ayub et al., 2016; Pereira & Gusmão, 2016). Considering recombination, it is possible to distinguish different types of DNA markers. There are recombination markers that can discriminate and identify a single person (e.g. autosomes and X chromosome markers)[4], highlighting that in males X-chromosomal recombination is limited only to the two tips that are homologous to the Y chromosome (Gomes et al., 2020). On the other hand, the non-recombining markers provide information that is shared among related individuals as they are haploid entities, which is the case of the mitochondrial DNA (mtDNA) and the male-specific region of the Y chromosome (MSY) (Butler et al., 2005; Pereira & Gusmão, 2016). The lack of recombination during meiosis allows the MSY and mtDNA to be inherited from generation to generation as a single locus (haplotype) (Charlesworth et al., 1996). Differently from the autosomes and the X chromosome (Fig.1), the Y chromosome is only transmitted from father to son, whilst in mtDNA the inheritance occurs from the mother to all offspring (Butler, 2005; Wallace D.C., 2007).
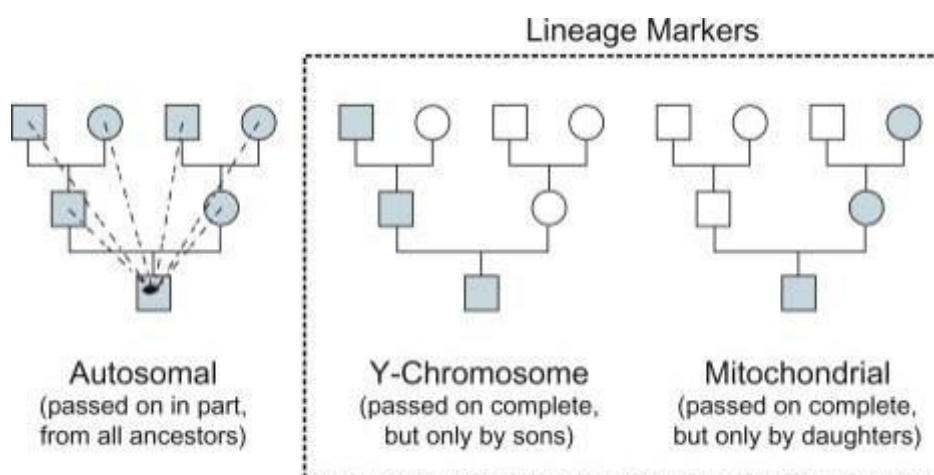
Figure 1: Diagram of the inheritance patterns of recombinant and non-recombinant genetic markers. On the left-hand side of the figure there is a diagram of how autosomal markers are passed on to offspring. Each individual receives (1/2) of the paternal information and (1/2) of the maternal information as they recombine with each other. On the right-hand side lineage markers are represented :1) markers belonging to the non-recombinant portion of the Y chromosome, are transmitted uniparentally, without modification unless a mutation occurs, from father to son. On the other hand, mitochondrial DNA markers are represented. It is transmitted only from mother to the offspring without suffering recombination (Adapted from Butler, 2012).

---

[3] A variant form of a gene which is located at the same locus on a chromosome.
[4] It is important to highlight that here only the recombinant segments of the X-chromosome and autosomes are taking into account. In their full extension both present regions that do not suffer recombination.

The Y chromosome and mtDNA share a low effective population size compared to the autosomes (1/4 of them) and the X chromosome (1/3) (Jobling & Tyler-Smith, 2003; Pereira & Gusmão, 2016). As a result, they are more susceptible to the effects of genetic drift, with higher genetic distances between populations. This enables uniparental markers to provide a different perspective on the composition and structure of populations compared to autosomal analyses (Carvalho Gontijo et al., 2019; Frudakis, 2010). Since uniparental markers do not undergo recombination, they can provide a clearer pattern of past historical events (Bisso-Machado et al., 2012). In the last 20 years, mtDNA and the Y chromosome polymorphisms have been widely used to address a plethora of questions concerning forensic and population genetics (Alonso-Morales et al., 2018; Martínez et al., 2020; Mogollón Olivares et al., 2020).

Owing to their patrilineal inheritance pattern, Y-chromosomal markers are used in paternity testing when the sample of the putative father is not available. In this case, Y-chromosomal markers are analysed on relatives of the same patrilineage as they share the same Y chromosome genetic information, unless mutation occurs (Jobling & Tyler-Smith 2003). Moreover, in cases of sexual assault, these markers allow the isolation and identification of the DNA profile of the offender in mixtures of male and female genetic material (Jannuzzi et al., 2020; Jobling & Tyler-Smith, 2003; Kayser, 2017).

1.2.1. Y chromosome

In humans, the Y and X chromosomes constitute the sex chromosomes of the genome and both evolved from a pair of ancestral autosomes (Charlesworth et al., 1996; Carey et al., 2022; Quintana-Murci et al., 2001; Snell et al., 2018). The Y chromosome of mammals is thought to have arisen 166 million years ago (Snell et al., 2018) as mutations on the proto-Y chromosome led to the formation of the testis-determining gene SRY (sex-determining region of Y chromosome) (Gubbay et al., 1990; Sinclair et al.,1990). Upon acquiring the SRY, the proto-Y chromosome gathered a number of mutations characterized by local inversions that suppressed recombination between proto-X and proto-Y (Bachtrog et al., 2013; Snell et al., 2018; Wright et al., 2016). Moreover, over time, the lack of recombination led to the accumulation of repetitive DNA sequences and a short-term decrease in the size of the Y chromosome (Bachtrog et al., 2013). This eventually resulted in large deletions and explains the relatively diminutive size of the Y chromosome in many mammals (Hughes et al., 2015). As a result, the Y chromosome had a major loss in gene content and expression, contrary to the X

chromosome that retained most of the original gene content and size (Bachtrog et al., 2013).

Nowadays, the Y chromosome is one of the smallest human chromosomes, with an estimated average size of 60 million base pairs (bp) (Gusmão et al., 2008; Quintana-Murci et al., 2001). It is comprised of several regions (Fig. 2): the pseudoautosomal regions (PAR1 and PAR2), the heterochromatic and the euchromatic regions. During meiosis, the PAR 1 and PAR 2, located at the end of short (Y_p) and long arms (Y_q) of the chromosome, respectively, exchange genetic material with the homologous regions at the tips of the X chromosome (Graves et al., 1998). These regions, comprising about 5% chromosome, represent the exceptions to the haploid genome (Freije et al., 1998; Quintana-Murci et al., 2001).

The male-specific region of the Y chromosome (MSY) comprises the non-recombinant portion, consisting of heterochromatin and euchromatin. Heterochromatin is a ''tightly packed'' (Clark & Pazdernik, 2016) form of DNA, mainly consisting of genetically inactive sequences, and is associated with several important cell functions such as gene regulation and silencing of repetitive DNA (Liu et al., 2020). On the other hand, the euchromatin is a constant size region that includes sequences homologous to the X chromosome, Y-specific repetitive sequences and all the genes identified in the Y chromosome (Gusmão et al., 2008). More specifically, the euchromatic region of the Y chromosome is divided into three distinct parts with different genomic sequences: (i) the X-transpose region, which contains genes homologous to the X chromosome which it can recombine with; (ii) the X-degenerate region, contemplating the portion with ancestral autosomal traces from which both the X and Y chromosomes have been derived; and (ii) the ampliconic region, which gathers MSY-specific genes(Gusmão et al., 2008; Hallast et al, 2021; Krauz & Casamonti, 2017; Skaletsky et al., 2003; Xu & Pang, 2022).
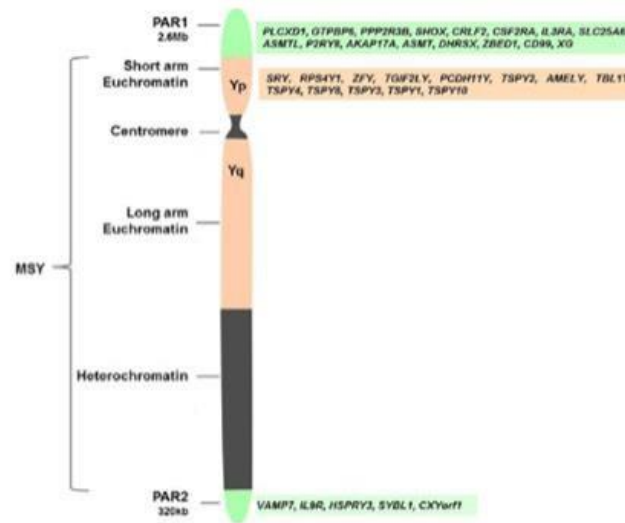
Figure 2: Structure of the human Y chromosome. The Y chromosome is made up of several genomic sequences with different functions. The recombinant portions of this chromosome are found at the ends of the chromosome (PAR 1 and PAR2) where the genes that each one encodes are represented in the green boxes. The Male Specific Region of Y-chromosome (MSY) is mainly composed by euchromatin and heterochromatin. Euchromatin contains the short (Y_p) and long (Y_q) arms of the chromosome, separated by the centromere. The genes encoded on the short arm are shown in peach tones. Heterochromatin is the mostly functional inactive region of the Y chromosome (Adapted from Colaco et al., 2018).

After the discovery of the first genetic polymorphisms on the human Y chromosome, in 1985 (Casanova et al., 1985), different types of Y-chromosomal genetic markers have been described and are currently used in forensic and population genetics. Nowadays, Short Tandem Repeats (Y-STRs) and Single Nucleotide Polymorphisms (Y-SNPs) are the most used.

1.2.1.1. Y-STRs

STRs or microsatellites are highly polymorphic multi-allelic markers with mutation rates at ~ 10^{-3} (Ellegren, 2004; Kayser et al., 2000). Arranged in tandem, they are small repetitive units between 2 to 6 bp (Pereira & Gusmão, 2016). Due to high mutation rate, the STRs have a high discriminatory power amongst individuals. The combination of several STRs on the Y chromosome shapes haplotypes that enable different paternal lineages to be discerned (Kayser, 2017). The origin of mutations in STRs is the result of a mechanism called replication slippage (Ellegreen, 2004; Leclercq, S., Rivals, E., & Jarne, P., 2010) (Fig. 3). This mechanism is characterized by a mismatch between the template strand and the strand being synthesized, resulting in a variation of the STR length (Dieringer & Schlötterer, 2003; Ellegren, 2000a; Ellegreen, 2004; Leclercq, S., Rivals, E., & Jarne, P., 2010). According to the Stepwise Mutation Model (SMM) (Kimura & Ohta, 1978), mutations during DNA replication occur by inclusion or subtraction of a repeat unit. The mutation rate is inversely correlated with the number of bases

comprising the repeat, i.e., the lower the number of bases, the higher the mutation rate (Birnkmann et al., 1998).
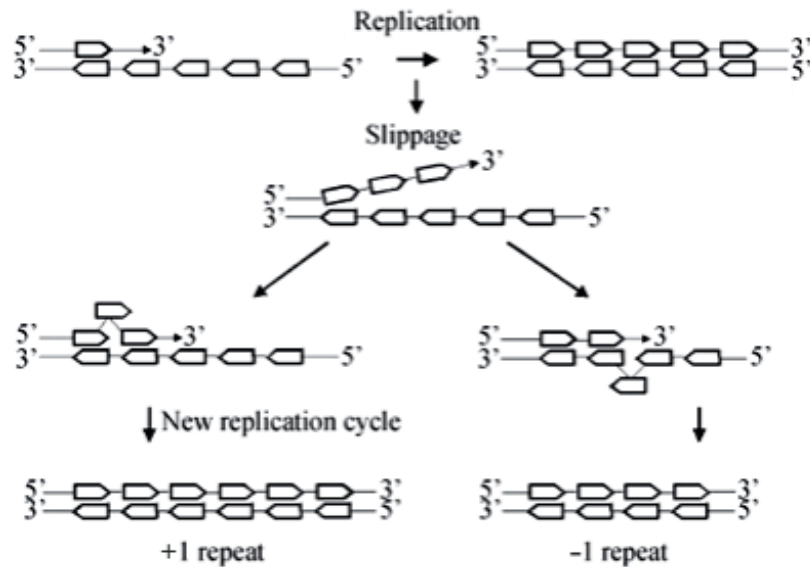
Figure 3: Representation of slippage during DNA replication. During the replication of DNA due to a mutation, a slippage on the growing strand of the replicate DNA can cause the gaining or the loss of bases (Adapted from Goldsteind & Schlotterer, 1999).

In population genetics, STRs are useful for assessing population diversity, as they are hypervariable, highly mutable and ubiquitous (Consortium IHGS, 2004; Rower et al., 1996; Silva et al., 2012; Weber & Broman et al., 2001). These characteristics have driven the use of this type of polymorphism to characterize human genetic diversity patterns and population structure (Silva et al., 2012). Their use is highly valuable to disclose recent evolutionary events rather than long past events as they are poorly stable due to their high mutation rates.

As previously mention, establishing valid databases of various populations from different locations are needed in evolutionary, demographic and forensic studies. Population databases aim to gather haplotype frequencies from worldwide populations that, then, allow performing and calculating comparisons between populations (Diegoli, 2015; Roewer, 2009). The Y Chromosome Haplotype Reference Database (YHRD) (https://yhrd.org/; Willuweit & Roewer, 2007) is a database for Y-STRs, available online, which contains haplotypes from 37 metapopulations with several populations within. Nowadays, this database also gathers Y-SNPs using phylogenetic tree from van-Oven et al.,2014. Recommended by the International Society of Forensic Genetics, the YHRD is the reference database ''most widely'' (Roewer, 2016) used to obtain the data needed to the interpretation of Y-STR results in forensic analysis (Roewer, 2016).

1.2.1.2. Y-SNPs

SNPs are the smallest and most abundant type of human DNA polymorphisms (Brookes AJ, 1999; Collins & Chakravarti, 1998; Sims et al., 2007). This form of polymorphism is the main source of genetic variation within the human genome and is produced by a single nucleotide substitution, insertion or deletion (Budowle & van Daal, 2008; Butler, 2005; Pereira & Gusmão, 2016). They are characterized as point, mostly, bi-allelic (two allelic variants) mutations with low mutation rates (10^{-9}) (Xue et al., 2009). Although most reported SNPs are categorized as bi-allelic, occurrence of tri-allelic and tetra-allelic SNPs have been reported by Phillips and coworkers, in 2020, which identified 271,934 tri-allelic markers. Furthermore, they conclude that SNPs with three alleles ''upgraded the discrimination power comparing to the bi-allelic SNPs'' (Phillips et al, 2020). SNPs present valuable advantages in forensic casework when compared to STRs. In mass disasters and missing person cases, DNA samples can be highly fragmented not allowing STR amplification due to the large size of amplicons (100 bps to 400 bps) (Budowle et al., 2009). In these types of samples, SNPs are more effective since their amplicons are shorter (around 50-80 bps) and, therefore, their amplification may be more successful (Budowle, 2004; Budowle & van Daal, 2008; Divne & Allen, 2005; Jobling & Gill, 2004). It is well recognized that past or recent human populations movements can leave their imprint on the genome and influence genetic diversity. Thus, the analysis of the genetic variation in current populations can provide information about migration events (Jobling MA, 2012; Underhill et al., 1997). The low mutation rates of Y-SNPs compared to Y-STRs, makes them highly effective for studying population movement events. Combinations of Y-SNPs define clusters of haplotypes with a common ancestor known as haplogroups.

Because the Y-SNPs are more stable than the Y-STRs, the haplogroups have a specific geographical distribution (Fig.4) enabling the paternal biogeographic ancestry assessment through Y-SNP genotyping (Brion et al., 2005; Jobling and Tyler-Smith, 2003; Kayser 2017; Larmuseau et al., 2017; Ralf et al., 2019; Underhill et al., 2003).
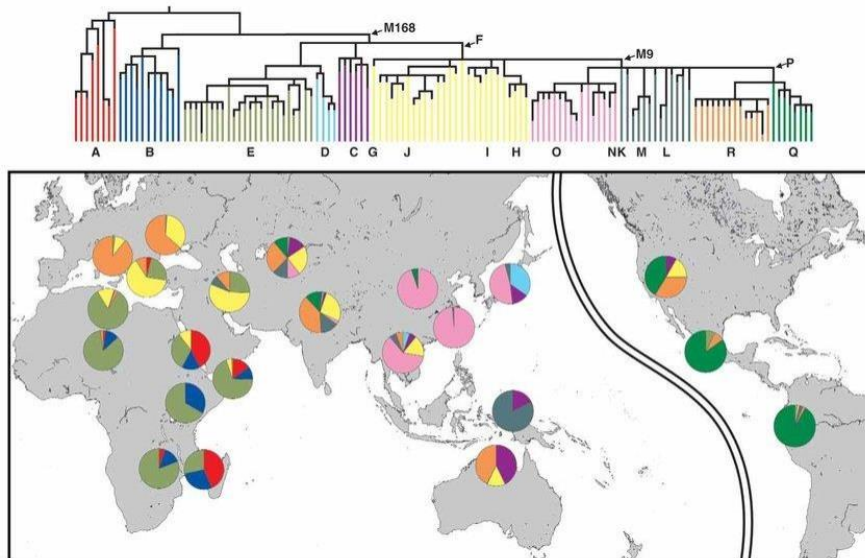
Figure 4: Geographic distribution of the major Y-chromosome haplogroups in 22 global geographic regions. Major haplogroups are represented above (A, B, E, D, G, J, I, H, O, NK, M, L, R, Q) with an attributed colour. On the map pie charts are represent with the occurrence of those major haplogroups according to geographic location (Adapted from Underhill et al., 2003).

The phylogenetic relationships between different Y-haplogroups have been used to construct hierarchical trees of Y-chromosomal lineages (Hammer et al., 2001; Karafet et al., 2008; van Oven et al., 2014). Due to the growth of new Y-SNPs and in order to establish a standard nomenclature about Y-SNP haplogroups, in 2002, the Y Chromosome Consortium created a Y chromosome phylogenetic tree gathering all 153 haplogroups[5] identified by 245 Y-SNPs. In 2008, Karafet and coworkers revised this tree and incorporated 600 new biallelic markers. The discovery of massively parallel sequencing (MPS) has led to the exponentially emergence of new biallelic markers in the past years. This increased the need of, once again, an update of this phylogenetic tree in order to establish a more consensual nomenclature. Therefore, there are, currently, several online platforms with different levels of nomenclature complexity and number of SNP markers, namely PhyloTree (https://www.phylotree.org/Y/tree/index.htm) (van Oven et al., 2014), ISOGG Y-DNA Haplogroup Tree (https://isogg.org/tree/) and Yfull tree (https://www.yfull.com/tree/).

---

[5] Each major haplogroup was identified by a different capital letter (e.g., haplogroup A)

1.3. DNA typing

Over the past few decades, there have been remarkable developments in methods and technologies designed to assist in the detection of nucleic acid sequences found in biological samples. These advances have resulted in accurate sequencing of both DNA and ribonucleic acid (RNA), leading to significant impact in a multitude of research areas such as forensic genetics (Haddrill, 2021). For instance, regarding the identification of the major Y haplogroups, SNaPshot© assays is one of the most used multiplex[6] techniques designed specifically to determine haplogroups (Bouakaze et et al., 2007; Geppert et al., 2011; van Oven et al., 2012). However, despite being considered a method with a ''high level of sensitivity and a great multiplex capacity'' (Metha et al., 2017; Wang et al., 2016), this technique has a major gap when it comes to achieving a high level of resolution in the identification of the main haplogroups (Ralf et al., 2019). In order to genotype a large number of biallelic markers, it is necessary to distribute them in different runs, as there is a limit in the number of markers that can be genotyped in a single run (Ralf et al.,2019). As Ralf and his team reported in 2019, as forensic samples are collected with a very low amount of genomic DNA (gDNA), this technique is "practically disadvantageous" to use in the daily routine of forensic casework. However, currently, there are more advanced techniques which allow a larger number of SNPs to be genotyped in a single multiplex ensuring a high level of resolution. (Ralf et al., 2019; Koksal et al., 2022).

1.3.1. Massively Parallel Sequencing

The latest developments in sequencing technologies that have been introduced in research, diagnostic and forensic laboratories have significantly enhanced the quality of nucleic acid analysis (Carratto et al., 2022). MPS methods ''effectively allow the sequencing of all types of nucleic acids, using either a whole genome or a targeted approach, with both DNA and RNA'' as reported by Balanovsky O. and coworkers, in 2015. The main difference from previous sequencing methods, is that MPS stems from sequencing output, allowing the processing of millions of reactions in parallel, resulting in high-throughput, higher sensitivity, speed, and effectiveness (Bruijns et al., 2018; Satam et al., 2023; Ralf. et al., 2019).

Prior sequencing methods could only sequence a single DNA fragment at a time generating only limited amounts of data, which in the end, is more complex and time-

---

[6] Inclusion of several pairs of primers in the same amplification reaction. Each pair of primers is designed to amplify a specific region.

consuming compared to MPS if multiple loci need to be analysed. Although demanding a complex workflow, MPS sequences millions of fragments of DNA simultaneously per run (Heather & Chain, 2016; Nguyen & Burnett, 2014).

Firstly, in this sequencing method, gDNA samples must be converted into a library of fragments suitable to be interpreted by the specific sequencer used (Goodwin et al., 2016; Head et al., 2014). After extraction, biological samples must be fragmented, end-repaired, and covalently linked to adapters-barcodes using ligation methods.

In target sequencing (e.g., amplicon sequencing) it is not necessary to fragment the DNA, as this method contains specific primers to amplify the target fragment. However, short-read sequencing technologies (e.g., Illumina) require long strands of DNA to be cut into smaller, readable pieces (Head et al.,2014). This fragmentation process can be physical fragmentation (e.g., sonication), chemical fragmentation (combination of metal cations and heat) and enzymatic fragmentation (e.g., restriction endonucleases) (Head et al., 2014). During fragmentation as the DNA is ´´broken´´, the end of each strand of molecule can have 3´and 5´overhangs. It is then mandatory to remove it as adapters-barcodes must be, next, covalently attached to the strand´s end. Adapters flank either side of the sequence of interest having attached to them barcodes that enable to individualize each library allowing to pool and sequence several individuals in a single run (Head et al., 2014). After this, it is for the upmost importance to purify the DNA suspension to ensure that there are no remaining oligonucleotides and small fragments in it. Magnetic beads (iron oxide particles) are often used to clean the DNA suspension (Berensmeier S, 2006). Its function is to create a magnetic field that attracts the DNA libraries immobilizing them on the surface of the beads when the libraries tubes are put in contact with a magnet (Berensmeier S, 2006). This separates the libraries from the residues that remain in suspension. The supernatant with the residues is then discarded and washed with ethanol while the libraries remain attached to the beads. It is only by the addition of a specific elution buffer that the filtered DNA libraries separate from the beads, becoming suspended again and ready to proceed to quantification and analysis. This quantification is performed by quantitative PCR (qPCR). This method combines the amplification steps of regular PCR, simultaneously monitoring the amplification directly within the reaction tube (Coleman, 2017). Using fluorescently-labeled probes, the exponential phase of amplification is detected in real time. Thus, the amount of amplified product is directly proportional to the amount of fluorescence emitted (Wilhelm & Pingoud, 2003). The detection chemistries for qPCR use various types of fluorescently labeled probes such, for example, TaqMan®, which are highly specific to bind to the adapter-barcodes that is about to be sequenced. TaqMan® probes are combined with a

reporter fluorochrome and a quencher dye (Heid et al.,1996; Jannetto et al., 2004). Once the target sample start to be amplified, the Taqman® probe is degraded by Taq polymerase and both the reporter and the quencher are cleaved (Heid et al.,1996; Jannetto et al., 2004). The separation triggers the emission of fluorescence by the reporter fluorochrome that was suppressed by the quencher. As reported by Heid and collaborators in 1996, before hybridization of the probe ''the reporter dye fluorescent emission is absorbed by the quenching dye''.

The amplification curve analysis will allow to quantify the amplicon amount by, comparison to a known standard proceeding, then, to the library normalization. The concentration of DNA libraries is variable. To ensure that the DNA is read evenly, it is therefore necessary to equalize the amount of all the libraries before proceeding to pooling (Chiniguy et al., 2020). The same volume of each library is combined in a single tube and this total volume with the combination of all the libraries is loaded onto the chip to proceed to sequencing.

1.3.1.1. Ion Torrent Technology

Currently, one of the most commonly used MPS platforms is Ion Torrent technology. One of the main advantages of this technology is that it incorporates Ion chef system which provides automated library preparation, template preparation and chip loading (Ballard et al., 2020).

Ion-Chef (Thermo Fisher Scientific) is a sample preparation robot that automatizes template preparation. In this instrument, libraries are subjected to clonal amplification by emulsion PCR (emPCR) (https://www.thermofisher.com/order/catalog/product/4484177) (Fig.5). EmPCR is based on compartmentalization of DNA fragments in a water-in-oil emulsion (Chai, 2019). Ideally, each droplet contains Ion Spheres™ Particles (ISP), one amplicon molecule and all other reagents necessary for the PCR reaction (buffer, dNTPs, polymerase). These spheres are the structures where the library templates will be amplified on, containing thousands of primers in their surface that are complementary to the DNA libraries adapters added during the library preparation. Moreover, it is important to remind that the primers on the surface of the ISP are specifically complementary to each adapter. Thus, every droplet functions as an ''isolated PCR microreactor'' (Zanoli & Spoto, 2012) leading to generation of numerous copies of the bound templates facilitating signal detection (Griffiths & Tawfik, 2006; Zeng et al., 2010). It is of upmost importance to add the right concentration of the combined libraries in the beginning of the template preparation. If the concentration is too low, not enough amplicons will be

amplified on the ISPs, and it will not generate enough data for further analysis. On the other hand, if the concentration is too high, there is a possibility of more than one sample amplicon ending up in the droplet of oil, resulting in more than one amplicon amplified in the same ISP (polyclonal).
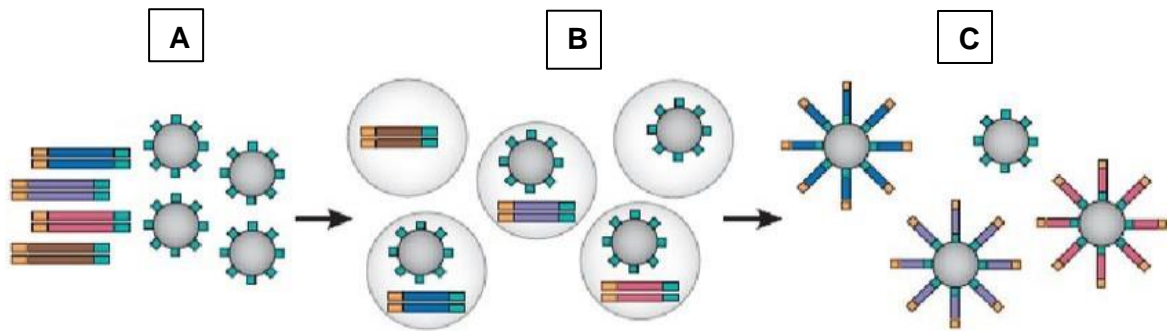
Figure 5: Overview about the emulsion PCR process. A) Adapters are used to capture single molecules of template onto microbeads. (B) Beads are incorporated into an emulsion, in which each bubble constitutes a microreactor containing DNA template, primer and reagents for PCR. (B) Following amplification, each bead is coated with clonally amplified molecules (Adapted from Kim, 2012).

The amplified DNA, which is bound to the sphere, is separated in a unique microwell of a chip (Fig. 6A). This chip is a ''complementary metal-oxide-semiconductor sensor array'' (Sun et al.,2021) that can perform all the data collection necessary for sequencing. Beneath the wells is an ion-sensitive layer followed by a proprietary ion-sensor (Fig. 6B). This is the oxide sensor array chip with the sensor surface present at the bottom of the well plate (Rothberg et. al, 2011). Such chips can perform millions of simultaneous sequencing reactions that the instrument will process as a signal that will be converted to base calls (captures natural biological activity and translates the information directly into digital nucleotide base calls) (Fig. 6C) (Rothberg et. al, 2011). Subsequently, a single type of dNTP is flowed on the chip at a time. As a dNTP is incorporated into the elongating strand by the DNA polymerase, hydrogen ions are released, resulting in a decrease in pH that is detected by the pH-sensitive chip (Goodwin et al.,2016). A greater decrease in pH indicates consecutive dNTPs being incorporated. The transistor-based detection of the released H+ it is known detected by a pH FET (pH-sensitive field effect transistor) (Sun et al.,2021). The local H+ concentration creates a positive voltage near the gate region of a transistor, which will result in a change in the

current flowing through the transistor that will be ultimately converted as base call by the sequencer (Merriman et al., 2012; Rothberg et. al, 2011; Sun et al., 2021).

A)

C)

B)
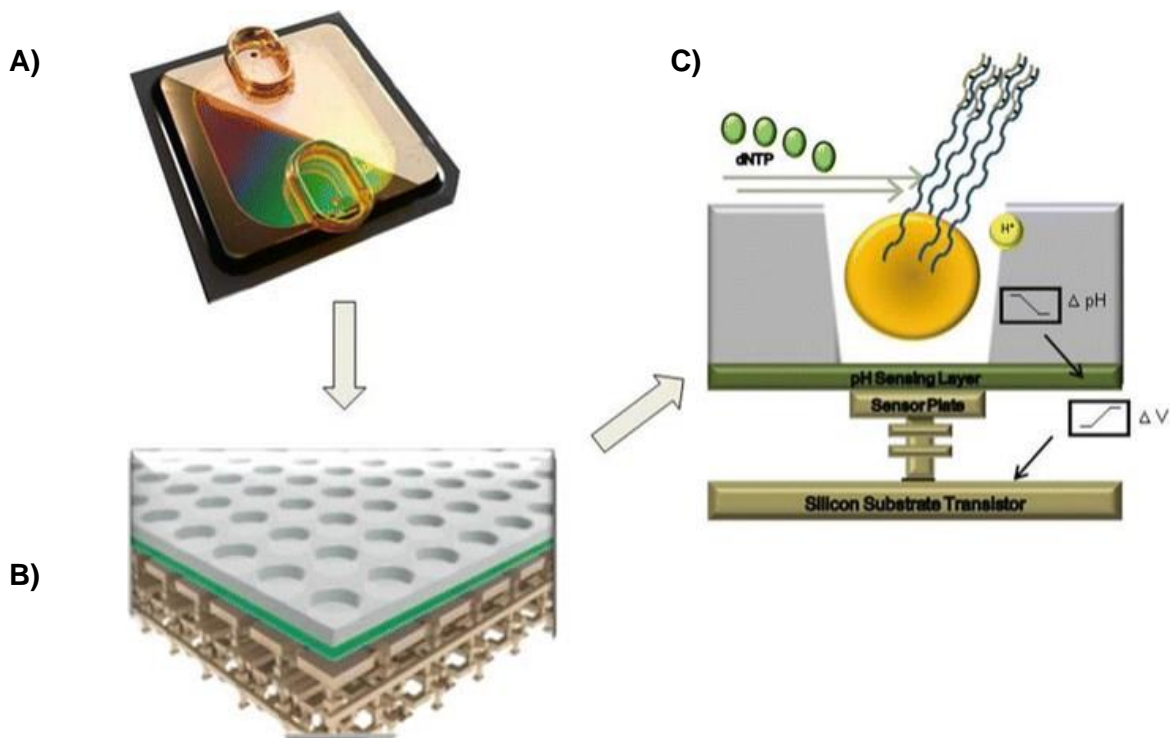
Figure 6: Overview representation of the sequencing chip, its functionalities and how the signal is converted in a base call. A) Ion-semiconductor sequencing chip; B) Ion Chip contains millions of ion-sensitive field-effect transistor (ISFET) sensors that allow parallel detection of multiple sequencing reaction; C) The instrument will detect the change pf the pH and will process as a signal that will be converted to base calls (Adapted from Del Chierico et al., 2015).

1.3.1.2. Y-SNPS panels for MPS

The highly repetitive nature of the MSY presents a challenge to its successful sequencing. Unique regions comprise 8.97 Mb and are distributed across the Y chromosome intermixed by non-unique regions (Wei et al., 2013). Concerning the analyses of Y-chromosome polymorphisms, the development of MPS techniques and the increased resolution of the developed kits will expand the path to identifying new lineages of Y-SNPs. This is a pivotal step that will help reveal new insights into the human population trajectory, as well as revealing new patterns of more recent migrations (Ralf et al, 2019; Kivisild, 2017).

Thermo Fisher Scientific developed the first kit for MPS presenting a kit for autosomal STR genotyping named Ion Torrent HID STR 10-plex. (Fordyce et al., 2015). After this major breakthrough, several panels have been developed by different companies to access and genotype different DNA markers. For accessing the biogeographical ancestry and human identification, Thermo Fisher Scientific developed

the HID-Ion AmpliSeq Identity Panel to target 124 different SNPs (including 90 autosomal SNPs and 34 Y-SNPs). Recently, Claerhout and coworkers., (2021) published an article describing a Y-chromosomal specific MPS panel named CSYseq that targets 15,611 Y-SNPs and 202 Y-STRs. In the present study, the Ion AmpliSeq™ HID Y-SNP Research Panel v1 (Ralf et al., 2019) was used. Until recently, it was the largest-scale Y-SNP typing panel targeting 602 amplicons with 859 SNPs that allows to define 640 Y-haplogroups. A total of 602 amplicons were designed with an average amplicon size of 130bp (Ralf et al., 2019). Moreover, this panel was designed to have most of the sequenced bases located within unique regions of the Y chromosome (Wei et al.,2013; Köksal et al., 2022) including SNPs which are geographically specific (Ralf et al., 2019; Köksal et al., 2022).

1.4. Human migration: conquering the New World

America was the last continent to be colonized by modern humans (Cavalli-Sforza & Feldman, 2003) and, for a long time, the scientific community has been debating about how the colonization process took place.

1.4.1 Entering the Americas: Beringia Strait

During the late Pleistocene glaciations, when global sea levels dropped, the continental shelf regions between Northeastern Siberia and Western Alaska became landlocked, forming a land bridge between the two continents (Goebel et al., 2008; O'Rourke & Raff, 2010). This land mass connected the Arctic Ocean to the Bering Sea, separating the continents of Asia and North America. Not only did this land bridge allow many species of plants and animals to pass between the two continents, it was essentially the only land connection between the two continents (Jiang et al., 2019; Elias & Crocker, 2008). According to present-day geography, Beringia extends from the Verkhoiansk mountain on Eastern Siberia side to the Mackenzie River in Northwestern Canada (Graf & Buvit, 2017; Hoffecker et al., 2016). In 1590, the Spanish Jesuit Priest Jose de Acosta, in *Natural and Moral History of the Indies*, was the first to speculate on the origin of the ancestors of the Native American populations (De Acosta & López-Morillas, 2002; Mazières, 2011). Jose de Acosta suggested that Native Americans derived from Asian populations, entering the continent by an overland via (Mazières, 2011; O'Rourke & Raff, 2010) that connected Asia to North America, lately, found to be the Beringia Strait.

1.4.2. Colonizing the Americas

The entry of modern humans into the Americas via Beringia is, nowadays, a more consensual topic in the scientific community. However, many doubts persist about the specific date of entry and the migratory routes of the first humans in the Americas. Analysis of modern and ancient nuclear data have shown that Native Americans derive from the same founding Beringian population that entered North America, known as Paleoindians (Grugni et., 2019). Moreover, the analyses also showed that the split between Northern and Southern Native Americans occurred in the South of the North American ice sheets (Moreno-Mayar et al., 2018; Kleman et al., 2010). After this event, the initial Native Americans spread across North and South America by two possible routes, one coastal and one inland (Mandryket al., 2001; Nielsen et al., 2017) constituting the two main models of migration routes. The first route, known as the ice-free corridor between the Cordilleran and Laurentide ice, has contributed to the settlement of North America, while the second, the North Pacific coast facilitated rapid southward expansion along the Pacific coastal regions of the double continent inland (Fig. 7) (Grugni et al. 2019; Hoeffecker et al., 2016; Pedersen et al., 2016; Potter et al., 2018).
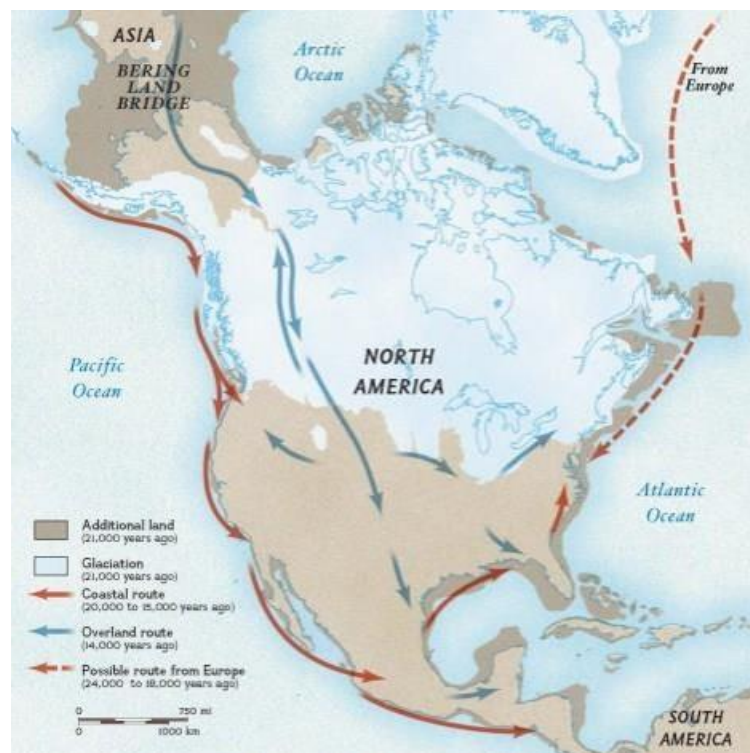
Figure 7: Migration and expansion routes taken by modern Humans in the colonization of the Americas. Representation of two possible routes taken by the aboriginals after entering America via Beringia strait. The coastal is represented by the all-coloured dark red arrows and the inland route is represented in blue arrows (Adapted from National Geographic website).

In 1930, stone projectile points were found at archeological sites spread through American Southwest and the Great Plains (O'Brien, 2019; Potter et al., 2017; Waters, 2019; Waters et al., 2020).Those gave clues to the scientific community of a single human migration that was then named Clovis, due to an 11 200-year-old human site found in Clovis, New Mexico (Dillehay, 2008). At the time, no older sites were known to exist in the Americas, and, indeed, archaeologists thought that the Clovis people were the first to step in into the American Continent (Dillehay, 2008; Eren et al., 2015; Llamas et al., 2016; O'Brien & Buchanan, 2017; Sanchez et al., 2014; Smith & Goebel, 2018; Sholts et al, 2017). In recent years, much has been discussed about the peopling of the Americas due to new discoveries in archaeology, history, linguistics, genetics and paleoanthropology (Dillehay, 1998; Dillehay, 2009; Gibbons, 2014; Hamilton & Buchanan, 2010; Skoglund et al., 2015). New evidences based on founded sites of human habitation at South of the ice sheets traces back to the presence of humans prior to the Clovis Era. Dillehay and collaborators, in 2015, reported ''human presence in the Monte Verde area between at least 18,500 and 14,500 cal BP.'' (calibrated years before present (cal BP). This turned attention to the coastal route of the North Pacific. This route ran along the coastal regions of the Pacific and allowed for a rapid migration southward (Davis & Madsen, 2020; Dillehay et al.,2015; Praetorius et al., 2023).

1.4.2.1. Genetic Evidence

Genetic studies demonstrate that Asia gave rise to the first Beringia's and Americans (Tamm et al., 2007). Thus, revealing that the ancestors of all Indigenous people had descended from five maternal lineages (haplogroups A, B, C, D, and X) (Tamm et al., 2007). These lineages also indicated that the founding population came from Asia and experienced a severe genetic bottleneck, in which a small number of people with limited genetic diversity gave rise to all Indigenous people who occupied the continent before European arrival (Torroni et al., 2000).

Identifying Native American founding lineages of MSY has been a challenge due to the post-Columbian by the high rate of male-mediated admixture into Native American communities (Bataglia et al.,2013; Colombo et al., 2022; Grugni et al., 2019). Moreover, Native American Y-chromosomal DNA analyses, also, exhibits low diversity (Geppert et

al.,2011; Geppert et al., 2015). However, findings revealed that only two paternal lineages, Q and C, gave rise to all Indigenous people (Zegura et al.,2004). Recent studies of the Y chromosome revealed the splitting into two distinct lineages of the ancestral components (ANC-A and ANC-B) (Moreno-Mayar et al.,2018; Scheib et al.,2018) which occurred in Eastern Beringia before entering America via two possible routes: the coastal route - followed by the ANC-A lineage and the internal route followed by the ANC-B lineage (Grugni et al.,2019). This fact resulted in the separation between the founding lineages of the New World and their Asian sister lineages (Skoglund et al.,2015; Skoglund, & Reich, 2016). Although, the majority of studies have been in accordance with a single common origin of Native American groups from Central and South America, some have suggested present-day Native Americans sharing traits with present-day Australasians (indigenous groups in Australia, Melanesia, and island southeast Asia) as skulls recovered from geological sites in America showed similarities between both populations (González-José, et al., Neves et al., 2003; Neves et al.,2004; Skoglund et al., 2015).Further genetic studies from Raghavan et al. (2015) and Skoglund et al. (2015) described those Native Americans as not being the directly related to Australasians, but from a named ''Population Y''. Unlike Native Americans, whose ancestors came from Siberia, Population Y was more closely related to Melanesians and Australian Aborigines. Although a lot of doubts persist about this topic, genetic evidence of Australasian ancestry was found in present-day Amazonian populations (Matisoo-Smith et al., 2010).

Thus, the crossing of the American continent allowed the entire continent to be progressively colonized by indigenous peoples. This first group of Natives living in the American continent remained isolated for a period, interbreeding, maintaining their Native genetic background until European colonization (Bartosik-Vélez, 2014; Grugni et., 2019).

1.5. Colombia

Colombia, located in the Northwestern part of South America, represents the entry point of the initial migrations into the South American continent by modern humans (Ongaro et al., 2019). Together with its diversity richness, Colombia turns out as an interesting object of study within the field of population genetics applied to the forensic field.

With an estimated population of around 52 million people, Colombia is the third-most populous country in South America, being one of the most urbanized. Colombia comprises 1140 00 km^2 divided in 32 administrative departments where each of them has capital of district (Censo Nacional de Población y Vivienda (CNPV, 2018) -

https://www.dane.gov.co/). Bogotá is the capital of the country, and it is located in the Cundinamarca department. Each department varies greatly in area, population, and urban centers. Colombia geography is divided in five natural regions. From the Andean region covering all Andean highlands , representing the most urbanized and populated region; to Caribbean coastal region covering the area adjacent to the Caribbean Sea , classified as the second most populated and economically active region of the country; to the Pacific region adjacent to the Pacific Ocean; the Orinoquia consisting of savannas and forest areas; to the Amazonian Region, which comprises the Amazon rainforest being the largest in area, but the less populated and, finally; the Insular Region which comprises San Andrés, Santa Catalina and Providencia islands in the Atlantic ocean (Fig. 8).
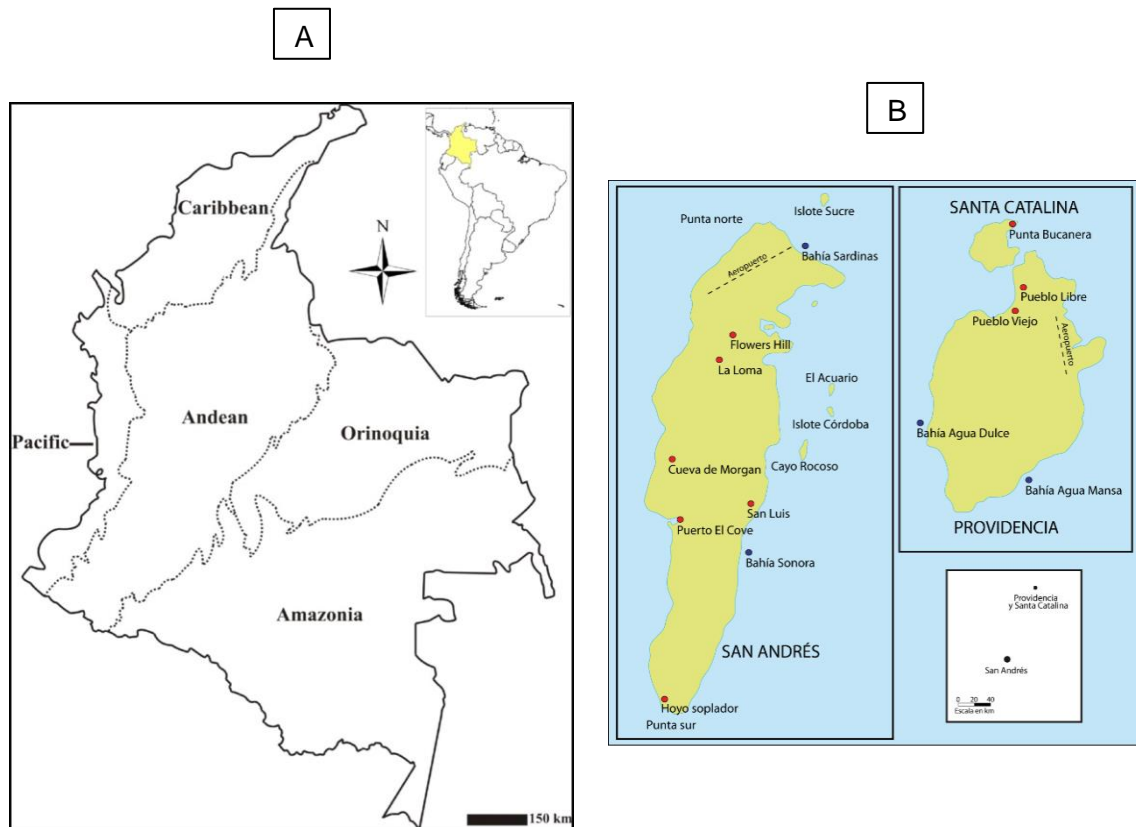
Figure 8: Representation of South American continent, highlighting the location of Colombia in the Northwestern part of the continent. A) Colombia divided into its five inland natural regions (Adapted from Hazzi, et al., 2013). B) The Insular region comprises three islands named San Andres, Santa Catalina and Providencia (Adapted from Google Images licensed under the Creative Commons attribution- share alike 4.0 International).

1.5.1. Colombia first settlers

In pre-Columbus times, many Native groups settled in the territory that corresponds to nowadays Colombia. The region referred to as the Isthmus-Colombian

area lies nestled between Mesoamerica and the Central Andes. It is considered a potential entry point through which Paleoindian migrations from North to South America may have occurred (Capodiferro, et al., 2021; Keyeux et al., 2002). According to Rothhmmer & Dillehay (2009), the first settlers made their way into South America by accessing '' Andean highlands by way of the Cauca and Magdalena River valleys which flow from south to north in Colombia''. Furthermore, they, also, reported that the most probable time of entering South America was between at'' least 15,000 and 13,500 B.P.'' (Rothhmmer & Dillehay (2009). Due to its geographical location, Colombia assumes a key role as a meeting point between Central and South America, facilitating interactions among diverse people.

1.5.2. European arrival, colonial exchange and recent migratory events

In the end of 15th century, an expedition led by Christopher Columbus marked the beginning of the European settlement in the South American continent, as newcomers from Spain and Portugal stepped the sub-continent (Bartosik-Vélez, 2014). During the early stages, male immigrants were predominant. Historical records attest to a general imbalance in the number of men and women arriving, especially during the early phase of European colonization. For example, early Iberian immigrants were mostly (>80%) male (Boyd-Bowman, 1976). The arrival of these immigrants led to a forced reshaping of the Native populations, as well as their reduction, due to the decimation of indigenous communities, victims of slavery work, and the propagation of diseases brought over from European immigrants (Collen et al., 2022; Fortes-Lima & Verdu, 2021).

The arrival of a Spanish military expedition led by Gonzalo Jimenez de Quesada (Francis, 2006), took the Spanish military strategically crossed the Magdalena River (Fig. 9) that rises in the south and runs through the entire Colombian territory.
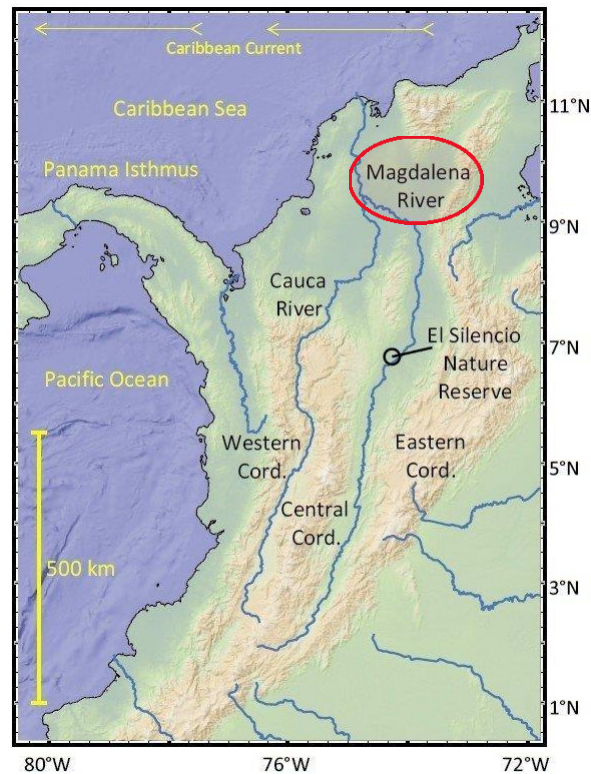
Figure 9: The course of the Magdalena River throughout Colombia. It rises between Central and Eastern Andes and floods into the north in the Caribbean Sea (Adapted from Ryan et al., 2009).

Along with the settlement of the Spanish conquerors, thousands of Africans, ''forcibly removed from their homelands'', arrived in the port city of Cartagena (Dawson, 2018). This port city was one of the largest destinations of ships transporting enslaved Africans who would be distributed and sold across colonial Colombia, Peru, Venezuela, Mexico, Panama, Bolivia, Ecuador, Chile, and Argentina (Wheat, 2011). After several rebellions against the colonial rule, more independence movements emerged throughout the Spanish colony. In 1819, the independence of the Gran Colombia state was declared, which agglomerated the now known territories of Colombia, Ecuador, Venezuela, and parts of other countries such as Panama, Peru, and Brazil (Arango, 1998; Arango, 2004).

Although European immigration to Colombia was relatively modest in the early 19th century, it surged significantly between the mid-19th and mid-20th centuries (King & Zontini, 2000). Between 1930 and 1940 Colombia received in its land a wave of Jews from Europe, most specifically a group of German Jews enter the country in 1936 (Leal Villamizar, 2022; Neuman, 1941). After World War II, Colombia witnessed an influx of Europeans (Rausch, 2021). Other immigrant communities, including the Spanish, French, Italians, and Americans migrate more recently to Colombia. Specially, a large wave of people from the Levant came into Colombia: from Arab countries, along with refugees from Syria, Lebanon, Jordan and Palestine escaped from the Turkish Ottoman

where a lot of violent conflicts arose due to the independency (Bruckmayr, 2010; Fawcett et al., 1997). The arrival of Middle Easterners reconfigured Colombia's communities as they established, foremost, in coastal cities (Bruckmayr, 2010). The convergence of Native Americans, Europeans, Asians and Africans led to extensive admixture throughout the country. The patterns of genetic diversity observed today are influenced by factors such as geography, the timing and scale of population migrations, and various social elements.

1.5.3. Colombian genetic diversity

Colombia is characterized by a great heterogeneity of ethnic groups/populations inhabiting different geographic regions (Rojas et al., 2010). Their patterns of diversity were modelled by the vaguely known ancient history of the people that lived in what is now known as Colombia. After the events that followed the European colonization, Colombia became an overall admixed country of Native, European and African lineages. Four well-defined ethnic groups are present in Colombia: Native Americans, Afro-Colombians (including the Raizales and Palenqueros), Mestizos and Gypsies (Censo Nacional de Población y Vivienda (CNPV, 2018) - https://www.dane.gov.co/). These ethnic groups differ in their physical features, culture, native language, tradition, and the geographical region they occupy. Approximately 83% of the Colombians are self-recognized as Mestizos, making this ethnic group the most representative in Colombia (Censo Nacional de Población y Vivienda (CNPV, 2018) https://www.dane.gov.co/)**.** The Mestizos are admixed populations that emerged across the entire country through interbreeding between Native and non-Native individuals, with diverse patterns prevailing in different regions. As a consequence, each population has its own distinct admixture pattern, as revealed by Salas and collaborators (2008), with the analyses of mtDNA of admixed Colombian population, and by Rojas and coworkers (2010), with the analyses of Colombian populations with X-chromosome, mtDNA, and Y-chromosomal markers. Generally, Colombian Mestizo populations consist of an admixture of Native American and European ancestries, with a smaller fraction of African ancestry (Salas et al., 2008, Rojas et al., 2010, Ibarra et al., 2014). However, the proportions of these ancestries vary significantly between different populations in Colombia (Ossa et al., 2016). The analysis of uniparental markers (mtDNA or Y chromosome lineages) provides valuable insights into the origin of women and men involved in the admixture processes. Genetic studies show that most Mestizos in Colombia resulted from admixture primarily involving Native or African women and European men (Rojas et al., 2010). This sex bias in the ancestry of Colombian admixed populations is also commonly

observed across South America (Wang et al., 2008). Admixed populations from the Caribe and Southwest Coasts often exhibit higher levels of African ancestry compared to populations from the Northeast, where Amerindian ancestry can surpass European ancestry. In contrast, more European lineages are observed in the North and Central regions of the country (Rojas et al., 2010, Ibarra et al., 2014).

1.5.4. Y-chromosome studies in Colombia

Several Colombian populations have been characterized for Y-chromosomal lineages (Alonso-Morales, et al, 2018; Noguera et al., 2014; Rojas et al., 2010). Comparing the haplotype and haplogroup composition of different populations helps to provide insights into migratory events, including the peopling of South America and the more recent demographic transformations that arose with the European colonization. Findings revealed by Noguera and team workers (2014) demonstrated a high diversity of Y-haplogroups in the department of Bolivar, Colombia (Caribbean Region). The sum of European male lineages exceeded 57% as R1b1-P25 was the most common Y-haplogroup encountered. E1b1a*-M2 ranked second, which represents the most widespread haplogroup in sub-Saharan African Niger-Congo populations, both in Bantu and non-Bantu groups (de Filippo et al., 2011). Finally, representing the Native American haplogroup only 7% of the Y-chromosomes were identified within Q major haplogroup Q1a3*-M346 and Q1a3a*-M3. Accordingly, Alonso-Morales and coworkers (2018) studied populations from the Andean Region, and the results showed that the European-specific R1b-M269 lineages were the most representative (57.83%). Native American paternal ancestry was also detected based on the presence of the Q1a2-M3*(xM19, M194, M199) and Q1a2-M346*(xM3) lineages. However, the most common sub-Saharan African lineage, E1b1a-M2, or other African lineages, were not detected in this study.

Although some studies have been done focusing on the Y-chromosomal characterization of Colombia, most were limited to Y-STRs and only a few included Y-SNP markers (e.g., Mogollón Olivares et al., 2020; Martínez et al., 2020). The combined analysis of Y-chromosomal markers such as Y-STRs and Y-SNPs is of special interest when it comes to determining bio-geographical ancestry, admixture and population substructure at different scales.

2. Aims

Colombia's location in northwestern South America suggests it might have been one of the entry points for early human migrations into the South American sub-continent. Its rich genetic diversity and historical context make it an intriguing subject for population genetics research. Studies on the characterization of the Y chromosome in Colombia are still lacking when it comes to the analysis of Y-SNPs and their haplogroups. Most of the articles published are limited to the characterization of Y-STRs and fail to explore the SNPs' potential for understanding the existing diversity in highly admixed countries. Therefore, an in-depth study of paternal lineages present in Colombia is crucial to increase knowledge about Human migrations that occurred in South America.

The main objective of this work was to extend the genetic coverage of Colombian male lineages belonging to Caribbean and Andean regions, to understand how events such as colonization, the transatlantic slave trade, and immigration may have contributed to the current genetic diversity patterns in Colombia. It will be of utmost importance to compare the results obtained in this study with other populations relevant to Colombian history to reveal patterns of miscegenation in Colombia and South America and to understand how each event has impacted and influenced the Colombian genetic composition.

To achieve this main objective, we propose to:

- Characterize Colombian male lineages by genotyping bi-allelic markers (859 Y-SNPs) using a MPS method;
- Complementing Y-SNPs data with Y-STRs haplotypes data in order to obtain a more complete characterization of the male lineages;
- Study the genetic stratification and admixture patterns in Colombia by comparing populations from different regions;
- Compare Colombian male lineages with those of other regions in South America and with populations from Europe, Asia, and Africa to infer the influence from each continent on Colombian populations;
- Interpret the data collected considering the historical events that shaped the Colombian demographic panorama.

3. Material and Methods

A total of 182 gDNA extracts from unrelated male individuals living in different districts of Colombia were studied. The DNA was previously extracted using standard Chelex and phenol-chloroform extraction methods, from blood samples collected on FTA cards (Whatman Inc.) at the University of Cartagena, Colombia. Samples were collected upon written informed consent signed by all participants. The study was approved by the Ethics Committee of the University of Cartagena (approval numbers: 013 and 040). Y-STR genotyping was earlier performed at the State University of Rio de Janeiro (UERJ), Brazil. Considering the Y-SNP genotyping, the workflow performed in this study is represented in Figure 10. Detailed information about each step will be further explained below.
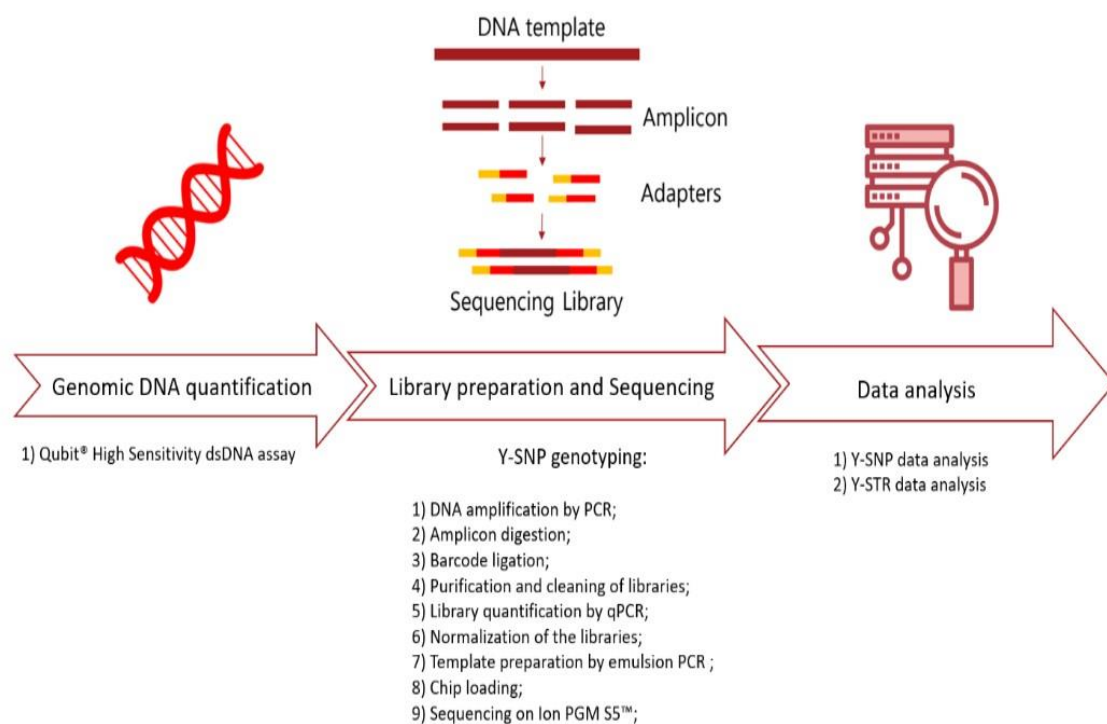
Figure 10: Overview of the workflow performed in this project. First, the DNA of all samples were quantified, followed by the Y-SNP genotyping process, which requires some essential steps (1-9) and, finally, culminating with the analyses of the data collected of Y-SNP. Moreover, Y-STR data genotyped within the scope of another project was used in the present study.

3.1. Genomic DNA quantification

Genomic DNA concentrations of our samples were measured by the Qubit® High Sensitivity dsDNA assay and Qubit® 3.0 Fluorometer (Thermo Fisher Scientific), which measures the natural absorption of light by DNA using fluorescent dyes that bind to the DNA molecule. The Qubit® working solution was prepared combining (199 x n) μL of Qubit® dsDNA HS Buffer mixed with (1 x n) μL of Qubit® dsDNA HS reagent, where n corresponds to the total number of samples plus 2 standards (Figure 11). It is essential to perform a calibration before quantifying the samples. Each standard (STD1 and STD2) was prepared by combining 190 μL of dsDNA HS Buffer with 10 μL of STD1 and STD2, respectively (Fig. 11).
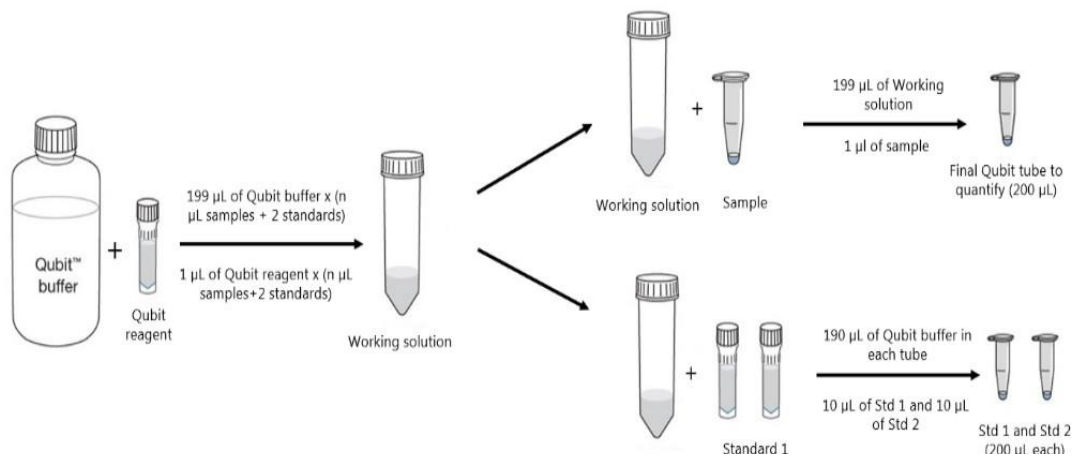
Figure 11: Qubit® High Sensitivity dsDNA assay workflow. Preparation of working solution and the standards to quantify the gDNA in the samples (Adapted from Quick Reference Qubit Assays PDF file (Pub. No. MAN0017210),Thermo Fisher).

3.2. Y-SNP genotyping

The Y-SNP genotyping of 859 Y-SNPs (Table S1, Appendix) was performed using the Ion AmpliSeq™ HID Y-SNP Research Panel v1 developed for MPS in the IonS5 System (Thermo Fisher Scientific). The protocol comprises the amplification of targets and preparation of libraries for sequencing. DNA amplification was performed by combining 4 μL of Ion AmpliSeq™ HiFi Mix, 10 μL of Ion AmpliSeq™ primer pool (2X),

1-5 μL of gDNA template (1-5ng) and 1-5 μL of water. A negative control was included in each PCR reaction to confirm the absence of contamination. The PCR reaction plates were placed in a thermal cycler and subjected to the conditions described below (Table 1).

Table 1: PCR amplification conditions to amplify the DNA.

Stage	Cycles	Temperature	Time
Denaturation[1]		99°C	2 minutes
Anneling[2]		99°C	15 seconds
Extension[3]	19x	60°C	4 minutes
Hold		10°C	Hold (for up to 1 hour)

[1]Double-stranded DNA templates are heated to separate the strands.

[2]Primers bind to flanking regions of the target DNA.

[3]DNA polymerase extends the 3′ end of each primer along the template strands.

After the PCR, non-incorporated primers and the ends of amplicons were digested by adding 2 μL of FuPa Reagent (Thermo Fisher Scientific) to the PCR products and the plate was placed in a thermal cycler using the conditions described below (Table 2).

Table 2: Thermal cycler conditions after amplicons digestion. (Information from Ion AmpliSeq™ Library Kit 2.0 User Guide (Pub.No. MAN0006735)

Temperature	Time
50°C	10 minutes
55°C	10 minutes
60°C	20 minutes
10°C	Hold (for up to 1 hour)

To perform the ligation of the sequencing adapters and barcode labels to the PCR products, 4 μL of Switch Solution (Thermo Fisher Scientific), 2 μL of DNA Ligase and 2 μL of diluted Ion Xpress™ Barcode adapter (Thermo Fisher Scientific) were added to the PCR products. To dilute each barcode, 2 μL of Ion P1 Adapter; 2 μL of Ion Xpress Barcode X and 4 μL of

nuclease-free water were combined. The ligation reaction was performed using the described conditions below (Table 3).

Table 3: Thermal cycler conditions after the ligation reaction. (Information from Ion AmpliSeq™ Library Kit 2.0 User Guide (Pub.No. MAN0006735)

Temperature	Time
22°C	30 minutes
68°C	5 minutes
72°C	5 minutes
10°C	Hold (for up to 24 hours)

Following the ligation reaction, barcoded libraries were manually purified and cleaned up using 45 µL (1.5X the sample volume) of AMPure XP magnetic beads (Beckman Coulter) and 150 µL of freshly prepared 70 % ethanol (2x). Each library was combined with magnetic beads (Figure 12, step 1). Afterwards, the tubes were placed in a magnetic rack and incubated for 2 minutes (or a little more) until the solution cleared (Fig. 12, step 2). The supernatant was carefully discarded without disturbing the pellet that contained the DNA bonded to the beads (Fig. 12, step 3). After repeating the cleanup process, 25 µL of LOW TE buffer (Elution buffer) was added and mixed with the beads (Fig. 12, step 4), to release the DNA fragments from the magnetic beads (Fig. 12, step 5). A total of 20 µL of the supernatant solution was transferred to another tube (Fig. 12, step 6).
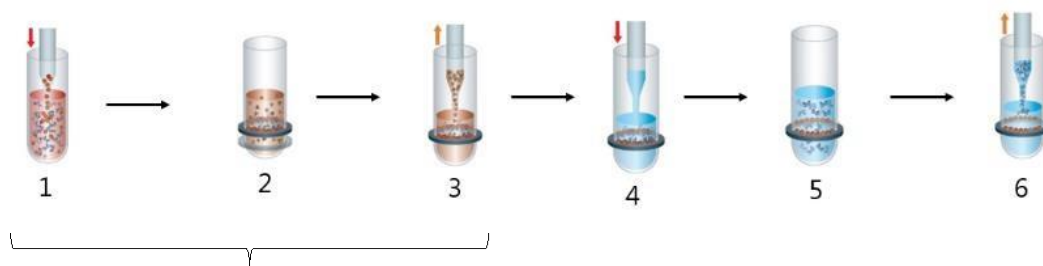
Repeat steps 1, 2 and 3 (2x)

Figure 12: Overview of the purification and cleaning process of barcoded libraries. Libraries in solution are first combined with the magnetic beads (1) and, after, placed in the magnetic rack (2). In this representation the magnet is surrounding the tube which will be coated with libraries. Then ethanol is added to clean the solution, discarding the supernatant (3). These previous steps must be repeated twice. In order to remove the libraries attached to the magnet, a buffer solution is added (4 and 5) and the solution containing the purified libraries is collected (6). (Adapted from SPRI Bead Technology (https://www.mybeckman.dk/).

Eluted libraries were quantified using the TaqMan™ Library Quantitation Kit (Thermo Fisher Scientific) and pooled to a final concentration of 50-80 pM. First, a 100-fold dilution was prepared using 2 µL of the eluted library and adding 198 µL of nuclease-free water. To use this quantification method, three 10-fold serial dilutions of the *E. coli DH10B* Control Library at 6.8 pM, 0.68 pM, and 0.068 pM were prepared as standards (Table 4).

Table 4: Three 10-fold serial dilutions of the E. coli DH10B Control Library. (Information from Taqman_Quantitation_Kit_User Guide protocol (Pub.No. MAN0015802)

Standard	Library input	Water	Fold dilution	Concentration
1	5 µL (undiluted)	45 µL	0.1	6.8 pM
2	5 µL (from STD1)	45 µL	0.01	0.68 pM
3	5 µL (from STD2)	45 µL	0.001	0.068 pM

Each sample, the standard and the negative control were quantified in duplicates. Sufficient reaction mix was prepared to replicate reactions by combining 20 µL of 2X Ion Library qPCR Master Mix and 2 µL of Ion Library TaqMan® Quantitation Assay, 20X. A total of 11 µL of this mixture was pipetted into each well on a plate. Finally, 9 µL of the diluted libraries (1:100) were added to each well of the plate. The plate was placed on the real-time qPCR instrument and subjected to the conditions described below (Table 5).

Table 5: Thermal cycling conditions for Applied Biosystems™ real-time PCR Instrument. (Information from Taqman Quantitation_Kit_UG protocol (Pub.No. MAN0015802).

Stage	Temperature	Time
Hold (UDG[7] incubation)	50°C	2 minutes
Hold (polymerase activation)	95°C	20 seconds
Cycle (40 cycles)	95°C	1 second
	60°C	20 seconds

---

[7] Uracil-DNA glycosylases (UDGs) are DNA repair enzymes that prevent cross-contamination of amplification products and primers used in previous PCR experiments.

Libraries with concentrations below 80 pM were re-amplified with 8 cycles according to the manufacturer's manual (Precision ID Library Kit on the Ion Torrent system by Thermo Fisher Scientific) to guarantee enough genomic material for normalization (Table S2, Appendix).

To perform template preparation and chip loading, libraries must be diluted to the same concentration to ensure an even distribution of sequencing reads for all samples (normalization). Thirty-two normalized libraries (including the negative control) were set up for each sequencing run, and 25 µL of the pooled library were used in the subsequent steps. Before sequencing, template preparation and chip loading were automated performed on the Ion Chef™ Instrument (Thermo Fisher Scientific). This instrument enables the pooled library to be clonally amplified by an emulsion PCR with Ion Sphere™ Particles (ISPs). Afterwards, 32 libraries were loaded into each Ion 530™ Chip (Thermo Fisher Scientific). Sequencing was performed on the Ion PGM S5™ instrument (Thermo Fisher Scientific) with 650 run flows, using Ion S5™ Sequencing Kit (Thermo Fisher Scientific).

3.3. Y-STR genotyping

The PowerPlex® Y23 System (Promega) was used, which allowed the simultaneously amplification of 23 Y-STR loci (DYS19, DYS385ab, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS481, DYS533, DYS549, DYS570, DYS576, DYS635, DYS643, YGATAH4). The alleles of each locus were separated on an ABI3500 Genetic Analyzer (Applied Biosystems) and analysed on GeneMapper™ Software v5.0 (Applied Biosystems) according to manufacturer's instructions.

3.4. Y-SNP sequencing data analysis

Sequencing analysis was performed on the Torrent Suite Server v.5.10.1 (Thermo Fisher Scientific). This server provided run statistics and quality information about the performed run, including information about the alignment to the *Homo Sapiens* reference genome (Hg19). The FileExporter v5.10.0.3 (Thermo Fisher Scientific) plugin was used to generate the output files from each run in BAM format for further analysis.

The coverage Analysis v5.10.0.3 (Thermo Fisher Scientific) plugin was used to check the sequence coverage of the targeted genomic regions.

3.4.1 Haplogroup determination

Y-leaf software version 3.0 was ran for Y-SNP variant calling and to infer the haplogroups using BAM files. The threshold criteria for the haplogroup inference were set as follows: minimum of 20 reads for each base, quality threshold of 20 for each base read, and majority base threshold of 95%. For comparison purposes, the Y-STR haplotypes previously generated were uploaded to the Y-DNA Haplogroup Predictor NevGen to (https://www.nevgen.org/) assign the most likely haplogroup to each sample.

3.5. Statistical analyses

Haplogroup frequencies were determined by direct counting. The number of shared and unique haplotypes among individuals; haplotype and haplogroup diversities (genetic diversities); pairwise F_{ST} and R_{ST} genetic distances, together with the Bonferroni-corrected p-value were calculated using Arlequin 3.5.2.2 software (Excoffier & Lischer, 2010). Pairwise genetic distances were calculated by removing DSY385ab and the number of repeats in DYS389I was subtracted for DYS389II.

The data obtained were compared with previously reported data from other South American, African and European populations (Tables 3-7, Appendix). Pairwise R_{ST} and F_{ST} genetic distances were visualized in Multi-Dimensional Scaling (MDS) graphs using the STATISTICA v.10 software (www.statistica.com).

Median-joining networks of Y-STR haplotypes (with the common set of STR between all studies) were constructed using the Network 10.0 software (https://www.fluxus-engineering.com/sharenet.htm) in order to assess genetic relationships of haplotypes within specific haplogroups. According to Qamar et al., 2002, a differential and inversely proportional to the variance of each microsatellite weight was applied in order to obtain a more parsimonious network.

In order to understand the pattern of interpopulation differentiation of haplogroups in the studied regions of Colombia, the relative frequencies of the main haplogroups were compared individually and differences between proportions were assessed by a Chi square statistical test (https://www.socscistatistics.com/tests/chisquare2/default2.aspx).

4. Results and Discussion

4.1. Sequencing analysis

4.1.1. Run overview

Data from 182 samples from different regions of Colombia were processed on the Torrent Suite Server v.5.10.1 (Thermo Fisher Scientific). The samples were divided into six chips, each one with 30-32 samples. The Torrent Suite server reported various parameters regarding the quality of each run, namely: total reads sequenced per chip, percentage of chip loading (i.e., how many chip wells contain beads), percentage of enrichment (i.e., how many beads contain DNA fragments), information on clonality of the library, and percentage of the final library. The results of the six runs are summarized in Table 6.

The percentage of the final library was low for the first two runs compared to the others. This value is a reflection not only of the amount of library loaded in the chip, but also the quality of the sequencing data generated. According to the recommended values reported in the sequencing metrics report, although chip one had a higher loading rate (≥ 50%), a similar enrichment rate (98%) and higher clonal rate (≥ 50%) than the remaining chips, the final library percentage was effectively reduced due to the percentage of low-quality reads (62%). On the other hand, although chip two had a low loading of ISPs on the chip wells, the enrichment percentage, and the percentage of monoclonal DNA template were high (60%). Contrasting to the 62% of low-quality reads of chip one, only 56% were filtered out as low-quality reads in chip two, resulting in a higher final library than chip one. Enrichment and clonal percentages are parameters whose output is intertwined with emPCR performance. In Ion S5™ sequencing runs, the number of wells per chip that can be loaded with ISPs define the number of possible reads. The primers attached to the ISPs bind to the adapters that were added to the DNA template, being replicated throughout the ISP. It is crucial to ensure adequate ratio of both input and ISPs to provide sufficient data. However, an excess of DNA input can generate polyclonal ISP (more than one amplified fragment on the same sphere). Readjustment of the pool concentration on chips three, four, five and six from 50 pM to 80 pM increased the amount of sequencing data produced in those runs (total reads) and overall contributed to higher percentages of useable final libraries.

Table 6: Run overview from the six chips reporting number of total reads; loading percentage; enrichment percentage; clonal percentage and final library percentage.

RUN	N	Pool concentration	Final volume to pool	Total reads	Loading percentage	Enrichment percentage	Clonal percentage	Final library percentage
1	31	50 pM	25 µL	4,743,986	60%	98%	60%	36%
2	30	50 pM	25 µL	3,002,222	34%	94%	60%	42%
3	32	80 pM	25 µL	6,231,939	55%	97%	62%	50%
4	30	80 pM	25 µL	4,924,570	46%	94%	59%	51%
5	31	80 pM	25 µL	7,177,523	58%	96%	64%	53%
6	31	80 pM	25 µL	7,302,536	65%	97%	61%	50%

4.1.2. Coverage analysis plugin

In the primary analysis on the server, all reads of each sample were mapped to the reference genome (Hg19 *Homo sapiens* reference genome). Afterwards, the Coverage Analysis v5.10.0.3 plugin (Thermo Fisher Scientific) was ran to obtain statistics on the number of mapped reads per sample, percentage of on target reads, and depth of coverage for each of the 602 amplicons. Table 7 summarizes the results for the six runs. Overall, the percentage of on target reads (those aligned to the Y chromosome), and the average depth of coverage was similar amongst the six runs, with the numbers reflecting the amount of library input used and the percentage of final library available.

Table 7: Coverage analysis overview reporting average values for mapped reads, average of the on target read percentage, and average depth of coverage for the 602 amplicons.

RUN	N	Pool concentration	Mapped reads	On target	Mean Depth
1	31	50 pM	145,219	54%	140
2	30	50 pM	94,527	56%	92
3	32	80 pM	187,604	58%	185
4	30	80 pM	187,604	58%	185
5	31	80 pM	223,504	58%	214
6	31	80 pM	224,691	54%	199

4.1.2.1. Failed amplicons

The information on the depth of coverage for each of the 602 amplicons in each sample was investigated further, to understand the balance of the PCR and sequencing of these fragments. It was important to determine which amplicons were not genotyped (or had few reads sequenced), and therefore, which Y-SNPs in the used panel would not be considered for analysis. To trim the data, a threshold of 20 reads per amplicon was set. A total of seven amplicons were below the defined threshold for 90% of the samples and were excluded from further analysis. This result is in accordance with findings in the literature, where six out of these seven amplicons (SP_84.270639; SP_85.1.1990847; SP_85.1.417472; SP_87.210087; SP_122.112152; SP_291.643377) were also reported to have low performance (Köksal et al., 2022). As in Köksal et al. (2022), no allele calls were considered for the 16 Y-SNPs located in the amplicons that did not meet the quality criteria (Table S8, Appendix). The exclusion of these markers from the analysis, however, does not necessarily mean that the lineages that they define were excluded. Indeed, for seven markers, there are alternative Y-SNPs in the panel that can be used to determine the same specific sub-haplogroup. For eight markers without an alternative Y-SNP for the same branch, it was still possible to define the haplogroup of a sample, since the panel contained Y-SNPs located downstream of the failed ones.

4.2. Final dataset

Samples with 20% or more missing data (N=7) were excluded for further analysis. Furthermore, as only eight samples from Amazonian individuals were analysed, they were also excluded for not being a sufficient sample number to infer any conclusions about this particular region. The final dataset comprised 167 samples from unrelated male individuals residing in the Caribe (N=98) and Andes (N=69) Colombian regions.

4.3. Y-chromosomal diversity in Colombia

Both Y-chromosomal haplotypes and haplogroups assigned to the 167 samples, together with NEVGEN (based on Y-STR profiles) and Y-leaf prediction results (based on Y-SNP information), are available in Table S9 in Appendix.

4.3.1. Y-STR data

In the two studied Colombian regions, 159 unique haplotypes were identified. More specifically, in Caribe 94% (N=92) of the Y-chromosomes presented unique haplotypes, showing this region the highest haplotype sharing between individuals (three haplotypes shared by two individuals each). In Andes, 97.1% (N=67) of the haplotypes were unique and only one haplotype was shared by two individuals. No haplotype was shared among the two regions. High levels of haplotype diversity and mean number of pairwise differences (Table 8) were observed for both regions of Colombia. Since there was a low number of shared haplotypes, it was expected a high haplotype diversity value in both regions. Moreover, it was already demonstrated that in urban areas from South America, as the two Colombian regions analysed in present study, high levels of admixture and uniparental-lineage diversity are expected due to the history of colonization, slave trade and, more recently, the extensive waves of immigration (Alonso-Morales et al., 2018, Ossa et al., 2016, Rojas et al., 2010).

Table 8: Haplotype diversity and mean number of pairwise differences values obtained on the male samples of the present study.

Region	N	Haplotype diversity	Mean number of pairwise differences
Caribe	98	0.9994±0.0016	13.3538 ± 6.0563
Andes	69	0.9996±0.0026	12.3474 ± 5.6431

Haplotype diversities obtained in our samples were compared with those from two other mestizo populations from Colombia. Romero and coworkers (2017) reported Y-STR profiles of individuals from seven Colombian Caribe states (Atlántico, Bolívar, Cesar, Córdoba, Guajira, Magdalena and Sucre). On the other hand, Alonso-Morales and team (2018) characterized the diversity and distribution pattern of Y chromosome lineages in the departments of Tolima and Huila, part of the Andean region of Colombia. Using the data from these two studies, diversities were calculated by grouping the samples into the macro-regions of interest: Caribe and the Andes, and considering 16 Y-STRs common among the three studies (DYS19, DYS385a/b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, GATA-H4, and DYS635).

Table 9: Haplotype diversity values obtained in our study and in other published studies from South American admixed populations. Number of shared haplotypes between different individuals from the same region were, also, calculated (SH*).

Region	N	AmpFISTR® Y-filer™ kit HD	SH*	References
Caribe	98 (admixed population)	0.9994 ± 0.0016	3	Present study
Andes	69 (admixed population)	0.9991 ± 0.0027	2	Present study
Caribe	293 (admixed population)	0.9999± 0.0003	3	Romero et al., 2007
Andes	83 (admixed population)	0.9979 ± 0.0022	7	Alonso-Morales et al., 2018

* Each haplotype was shared by two individuals from the same region.

It was possible to observe (Table 9) that data from the present study showed similar diversity values to those already published (Alonso-Morales et al., 2018¸ Romero et al., 2007). Moreover, Alonso-Morales et al., (2018) showed the lowest haplotype diversity.

Pairwise R_{ST} genetic distances between our two Colombian populations were calculated. A high, statistically significant genetic distance was observed between these two Colombian regions (R_{ST}=0.07098; p=0.00000) revealing that as both regions are highly diverse and heterogenous within each region, they are not genetically close to each other.

4.3.2. Y-SNP data

The data was first examined with the Y-leaf software for variant calling and haplogrouping. As mentioned before, it was determined which amplicons were not typed (or had few reads sequenced), and therefore, which SNPs in the panel would be excluded from the analysis. Thus, 16 Y-SNPs located in amplicons that were not typed in 90% of the samples were eliminated; two Y-SNPs located in regions outside of amplified amplicons were also excluded, and finally, 569 Y-SNPs with ancestral status for all samples were also excluded. The final dataset contained 272 variable Y-SNPs (i.e., with ancestral and derived states in the 167 samples), that define 62 haplogroups.

Table 10: Haplogroup diversity values found in the studied samples from each region.

Region	N	Haplogroup diversity
Caribe	98	0.9520 ± 0.0096
Andes	69	0.9037 ± 0.0317

Haplogroup diversity values were compared with other ones reported in the literature for Colombia and other admixed populations from South America (Table 11).

Table 11: Haplogroup diversity values obtained in our study and in other published studies based on 13 SNPs.

Region	N	HD	References
Caribe	98	0.8016 ± 0.0233	Present study
Andes	69	0.6121 ± 0.0641	Present study
Colombia (Caribe and Andes)	167	0.7462 ± 0.0283	Present study
Colombia	121	0.7353 ± 0.0261	Campos Ribeiro, R., 2022
Argentina	250	0.7151 ± 0.0259	Corach et al., 2010
Ecuador	527	0.7290 ± 0.0100	Villaescusa et al., 2021
Bolivia	226	0.6434 ± 0.0306	Cardenas et al., 2015
Patagonia	196	0.7429 ±0.0257	Manuel Fernandes Rodrigues, P., 2021
Brazil	1260	0.8295 ± 0.0053	Resque et al., 2016

The diversities of these seven populations were calculated considering 13 common Y-SNPs between all studies: A-M13, A4-SRY10831.1*, E-M96*(xM35), E-M35, FM213*(xM201, M170, 12f2a, M9), G-M201, I-M170, J-12f2a, KLNT-M9, Q-M242*(xM3),

Q-M3, R-M207*(xM343) and R-M343. Overall, all admixed populations from South America presented very similar haplogroup diversities (Table 11), which is in accordance with the fact that most American admixed population share a highly diverse paternal pattern along the sub-continent. For instance, Bolivia had the lowest haplogroup diversity, which was already expected as its population has a higher frequency of Native Y-chromosomal lineages compared to the other populations. It is important to note that reducing the number of SNPs in the analysis biases the result, since, in order to have a more or less balanced comparison, it is necessary to use markers with greater resolving power within the lineages depending on the sample universe being analysed (e.g., admixed or native). Thus, the analysis of the admixed samples along with 859 SNPs kit reveals a greater diversity of haplogroups because the SNP genotyping resolving power within each branch is higher.

Pairwise F_{ST} genetic distances between populations from Andes and Caribe were calculated (F_{ST}=0.03509; p=0.00000). The genetic distance observed between the two populations reveals statistically significant differences within haplogroups. This is in accordance with the reported data from other Admixed populations in South American countries revealing different patterns of Native American, European and African ancestries along different regions from the same country (Vullo et al., 2015; Salzano, 2004; Sans, 2000).

4.3.3. Colombian Y-chromosomal major haplogroups

Colombia (Caribbean and Andean populations) exhibited a high European influence (83.23%). The remaining male samples showed sub-Saharan African (10.78%) and Native American ancestry (5.99%). According to the literature, the Spanish colonization shaped the population from Colombia, with the introduction of European individuals and African slaves, which led to major changes in the demographics of Native population groups (Alonso-Morales et al., 2018; Páchon et al, 1996). Today, the genetic heritage from Colombia is the result of admixture of Native American, African and European contributions (Alonso-Morales et al., 2018; Xavier C et al., 2015; Censo Nacional de Población y Vivienda - CNPV, 2018) - https://www.dane.gov.co/). To understand the haplogroup diversity pattern inside these two regions, the relative frequencies of the main source lineages (European, African and Native) (Figures 13 and 14) were compared and differences between proportions were assessed by a Chi-square statistical test. Most noticeably, the frequency of sub-Saharan African lineages (E-L485*(xCTS5038); E-U209*(xZ37284, U290); E-U174*(xZ37817, CTS490, L372); E-U290*(xU181); E-U181) in the Caribbean population (16.33%) was higher than in the Andes (2.9%).

The Chi-square test revealed a significant difference in the proportions of these haplogroups between the two regions (X^2=5.7623, *p*=0.016374). On the other hand, the frequencies of the European and Native lineages showed no significant differences between the two regions. It can therefore be concluded that the greater haplogroup diversity achieved in the Caribe (0.9520 ± 0.0096) compared to the Andes (0.9037 ± 0.0317) is correlated with the greater frequency of African lineages found in the Caribbean region.

Macro-haplogroup frequencies found in the Caribbean region
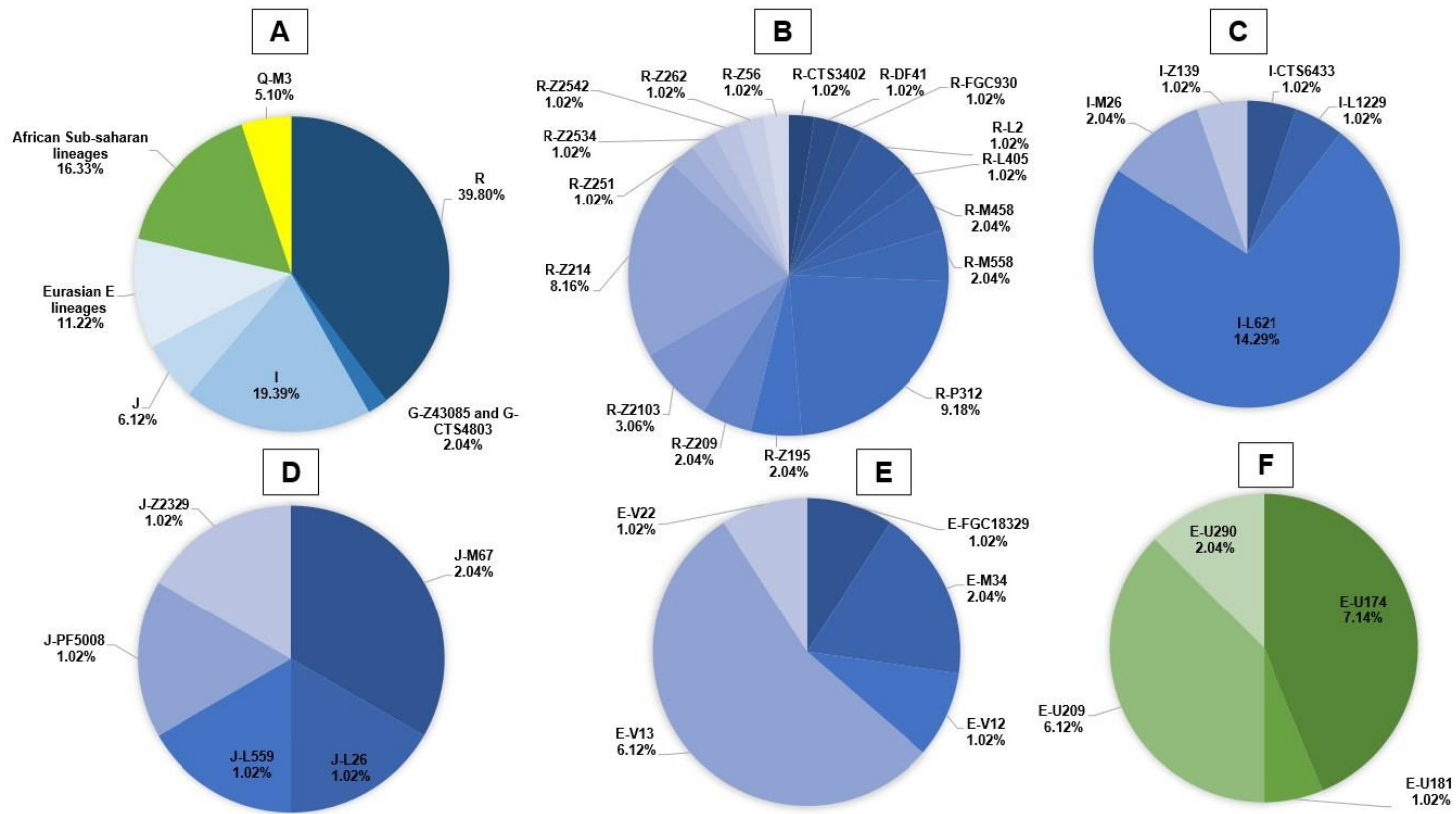
Figure 13: Overview of macro-haplogroup frequencies found in the Caribbean region. A) Frequency of all haplogroups and their respective lineages: in yellow are the Native haplogroup; in green are the sub-Saharan haplogroups and the Eurasian lineages are represented in different shades of blue; B, C, D, E, F) Frequency of sub-lineages inside haplogroups R, I, J and E (Eurasian), and sub-Saharan E lineages, respectively.

Macro-haplogroup frequencies found in the Andean region
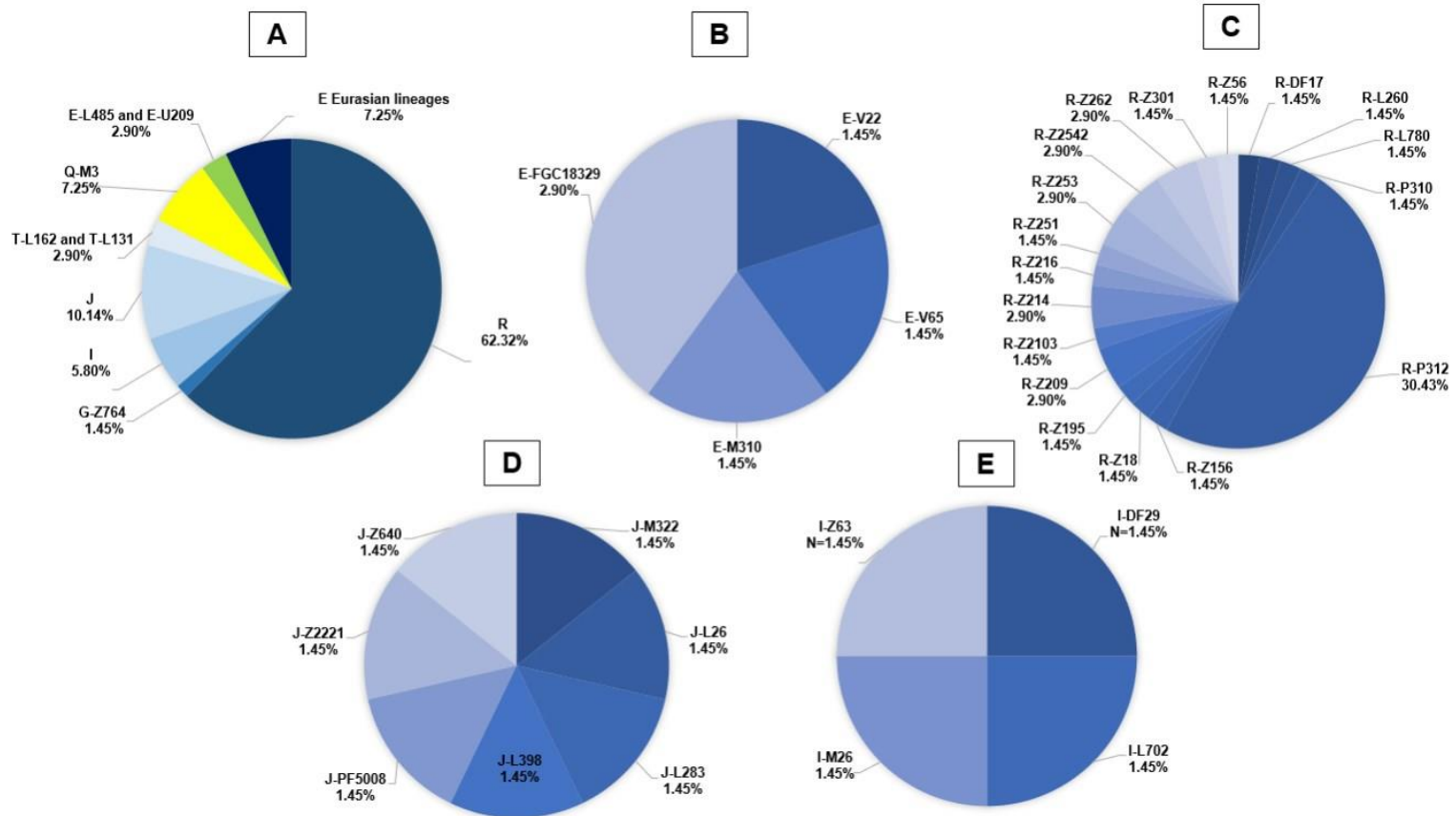
Figure 14: Overview of macro-haplogroup frequencies found in the Andean region. A) Overall macrohaplogroups frequencies found in Andes. Eurasian major haplogroups are represent in blue; Native haplogroup is represent in yellow; the sub-Saharan African haplogroups are in green; B; C; D; E) Represent E Eurasian, R, J and I lineages frequencies, respectively.

4.3.3.1. Haplogroup R

Of all the inferred haplogroups, the most prominent in both Colombian regions was the macrohaplogroup R defined by the M173 marker. In the study, 62.32% of the Andean samples and 39.80% of the Caribe were assigned to this major haplogroup, belonging to two main subclades (Chiaroni, et al., 2009): R-M420 (R1a) typically found in Eastern European and R-M343 (R1b), more frequent in Western and Central European populations (Myres et al., 2011; Underhill et al., 2015; Kayser et al., 2005) (Fig. 15).

Final Haplogroup	Caribe(N)	Andes(N)
R-M458*	2	0
R-L260	0	1
R-M558*	2	0
R-CTS3402*	1	0
R-L780*	0	1
R-P310*	0	1
R-Z18*	0	1
R-Z156*	0	1
R-Z301*	0	1
R-FGC930	1	0
R-P312*	9	21
R-Z195*	2	1
R-Z209*	2	2
R-Z216*	0	1
R-Z214	8	2
R-DF17	0	1
R-Z262	1	2
R-L2*	2	0
R-Z56*	1	1
R-Z2542*	1	2
R-DF41	1	0
R-Z251	1	1
R-Z253*	0	2
R-Z2534	1	0
R-Z253*	3	1
R-L405	1	0
TOTAL (N)	**39**	**43**
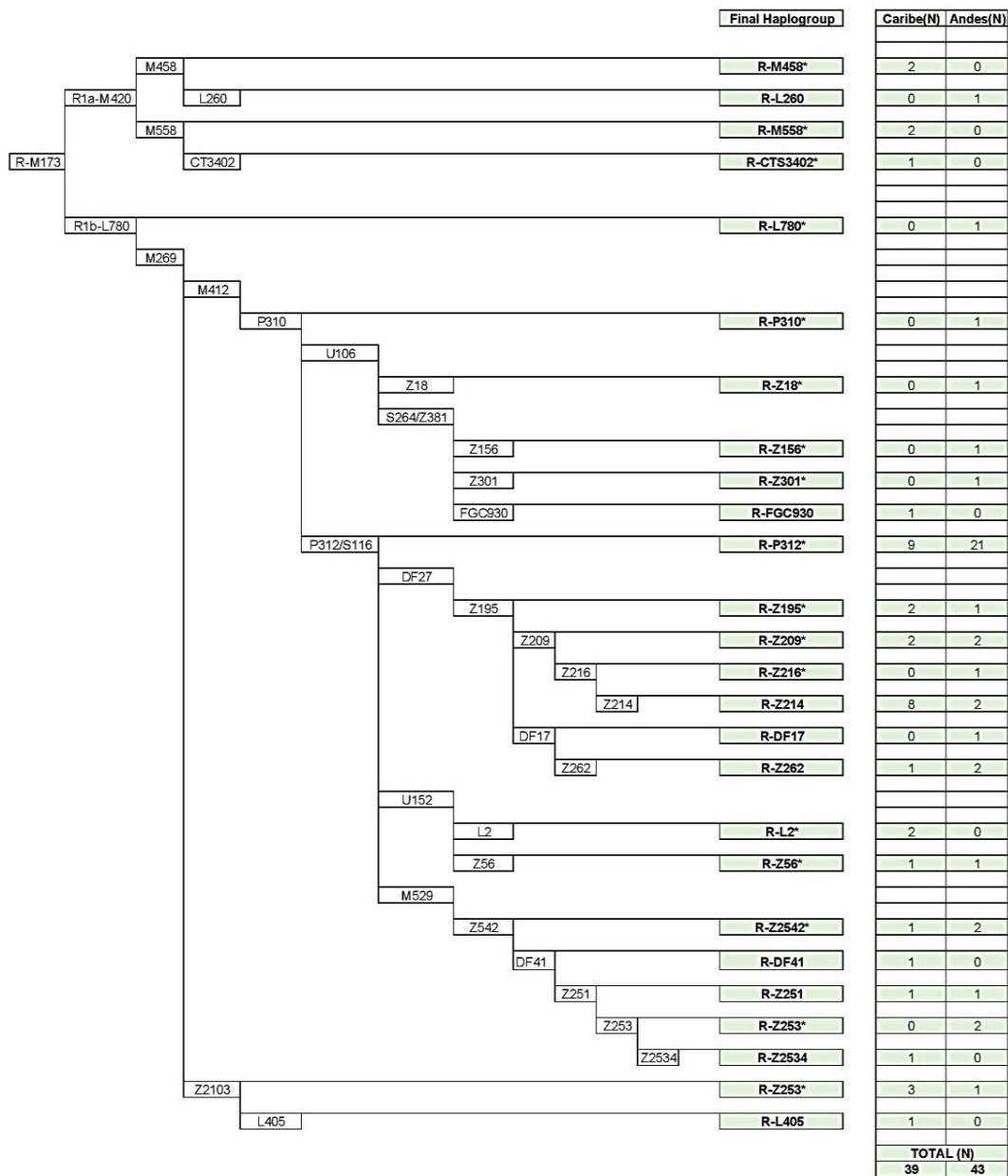
Figure 15: A simple hierarchical representation of the major haplogroup R. Only genotyped Y-SNPs that had a haplogroup assigned to samples in the dataset are represented. Also, a table is showed with the number of samples assigned to each haplogroup in Caribe and Andes.

From R-M420 (R1a) lineage, four different sub-haplogroups were identified in the dataset. The more upstream sub-haplogroup found was R-M458* (N=2). Inside this sub-haplogroup a sample from Andes was assigned to R-L260. R-M558* parallel to R-M458* was assigned to two Caribbean samples, sharing the same haplotype. Downstream to M458*, one sample was assigned to sub-haplogroup R-CTS3402*. According to the literature, R-M458* and R-M558* appear in Central Europe, Russia, Ukraine, and the Balkan Peninsula (Underhill et al., 2009; Underhill et al., 2015).

Haplogroup R-M343 (R1b) was identified in this study by an Andean sample with an equivalent marker R-L780*. Downstream, other R1b samples from the dataset were allocated to the main sub-haplogroup R-M269. The most upstream sub-lineage within R-M269 was R-P310* assigned to a sample from the Andes. Downstream, inside R-U106* haplogroup, its sub-haplogroups R-Z18*, R-Z156*, R-Z301* and R-FGC930 were allocated to one sample each from the Andean region, with the exception of the R-FGC930 lineage, which was found in a sample from the Caribe. Parallel to R-U106, the R-P312/S116 sub-haplogroup had the highest prevalence in the Caribe and Andes. (33.67% and 53.62% respectively). The P312 branch breaks into three major sub-lineages: R-DF27; U152 and M529 (Solé-Morata et al., 2017; Valverde et al., 2016). Figure 16 shows all the sub-lineages found within DF27; U152 and M529, as well as the number of samples attributed to each lineage. Most samples of Caribe and Andes were found belonging to R-DF27 lineages. Solé-Morata and collaborators, in 2017, reported a high frequency of DF27 in Western Europe, more precisely in Asturias and Portugal.

Caribbean region   **Andean region**
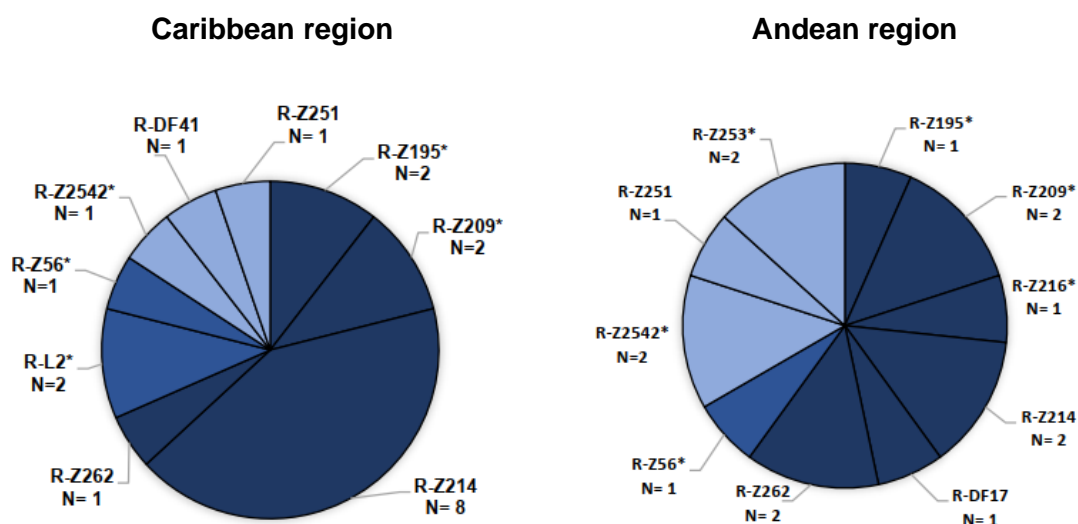
Figure 16: Number of individuals from each Colombian population belonging to R haplogroup sub-lineages R-DF27, R-U152 and R-M529. Different shades of blue were used to address the diversity of the identified lineages found inside of these haplogroups: dark blue: R-DF27 lineages; medium blue: R-U152 and light blue for R-M529 lineages.

The most downstream lineages found within R major haplogroup were R-Z103 assigned to three samples in Caribe and one in Andes and, further, R-L405 in one sample in Caribe. As far as haplogroup R-M269 and its sub-lineages are concerned, these Y chromosomal lineages are particularly relevant when it comes to understanding the genetic history of Western Europe. According to the literature, R-U106* is very common in the Netherlands, Germany, Switzerland, Austria, Western Poland and throughout Northern Europe (Busby et al., 2012; Cruciani et al., 2011; Lucotte, 2015; Myres et al., 2011; Zieger & Utz, 2020). Meanwhile, the Y-SNP marker P312/S116 defines one of the most common haplogroups in Western Europe and is mostly prevalent in the Iberian Peninsula (Myres et al., 2011; Valverde et al., 2016).

Concerning the historical data of Colombia, the presence of these lineages in the nowadays admixed Colombian males is, most certainly, connected to the European colonization and, more recently, due the genetic influx from the European migrations into this country since the 20th century. These results, regarding haplogroups and high diversity sub-lineages within R1b-M269, reflect the migration dynamics characterized by the arrival of settlers from various parts of Europe, mainly from the Iberian Peninsula (i.e., Spain and Portugal). Indeed, all lineages belonging to the R clade, especially R-S116* and R-DF17*, are likely the result of the Spanish and Portuguese legacy in the Americas.

4.3.3.2. Haplogroup E

Haplogroup E was more frequent in Caribe (27.55%) than in Andes (10.14%) (Fig. 17). In Caribe the most frequent major sub-haplogroup was defined by M2 marker (N=16). On the other hand, Andes showed a higher frequency of lineages inside major sub-haplogroup E-M35 (N=5). Defined by the M96 mutation, haplogroup E lineages are mostly spread on the African continent (Karafet et al., 2008). However, it also encompasses some subclades that are also spread across Southern Europe, the Middle East and, more rarely, in Central and South Asia (Karafet et al., 2008, Cruciani, et al., 2004).

Considering E-M2 haplogroup, the sub-lineage E-L485* was only found in Andes Inside of E-L485*, seven samples from Caribe were assigned to sub lineage E-U174. Parallel to E-L485*, E-U209* branch had the highest frequency within both regions. Here, six samples from Caribe were identified by U209 Y-SNP. Downstream, other sub-haplogroups were also detected: E-U290* and E-U181. E-M2 is distributed in sub-Saharan Africa and it has been connected to the Bantu expansion (Cruciani et al., 2002; Berniell-Lee et al., 2009). This expansion was responsible to reshape the populations

from Central, Eastern, and Southern Africa through the spread of the agriculture, the Bantu languages and, more importantly, the Bantu genetic pool (Bostoen, 2018). The presence of E-M2* lineages in the Colombian data set most likely resulted from Transatlantic slave trade. These Africans who were brought over to Colombia were mainly from Bantu groups of sub-Saharan Africa, which includes territories along the Atlantic Sea in Central-Western and Western Africa, as well as Mozambique on the east coast.
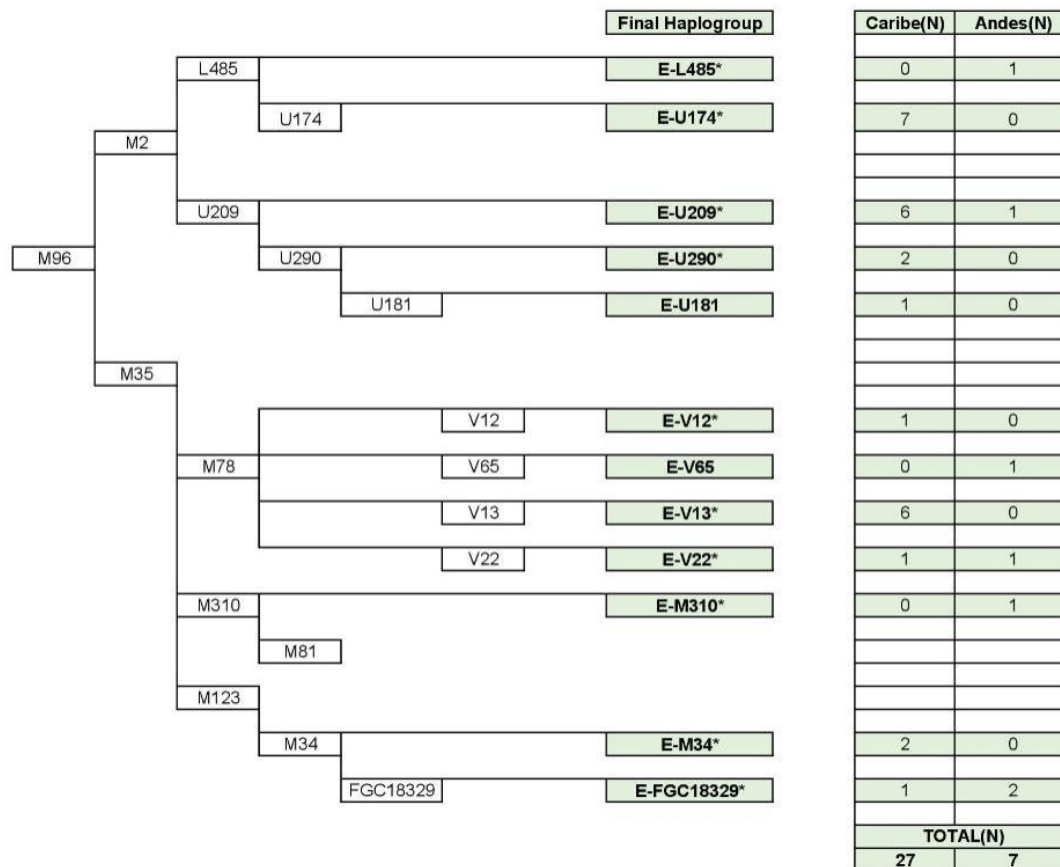
Figure 17: A simple hierarchical representation of the major haplogroup E. Only the genotyped Y-SNPs that had a haplogroup assigned to samples in the dataset are represented. Also, a table is represented with the number of samples assigned to each haplogroup in Caribe and Andes.

Concerning E-M35 branch, three major sub-lineages were identified (defined by M78; M310 and M123 markers). The most upstream lineage of E-M35 was E-M78. Inside of it, four sub-haplogroups were identified (E-V12*; E-V65; E-V13* and E-V22*). Caribbean population had the highest frequency of E-M78 lineages in which, one sample from E-V12* and six others from E-V13* lineage. On the other hand, the Andes population only presented one sample belonging to E-V65 and another to E-V22* (Figure 18). Cruciani and coworkers, in 2017, proposed the possibility of the E-M78 haplogroup having an origin in Northeastern Africa, more specifically, in the region of Lebanon and

Libya. In addition, the E-V13* lineage, also found in this study, indicated a later expansion to the European continent (Cruciani et al., 2017).

Since E-M78 lineages are present in Africa and Eurasia, in order to clarify the most likely origin of the lineages found in the Colombian samples, a network was constructed using 10 Y-STRs common to all studies (Fig. 18).
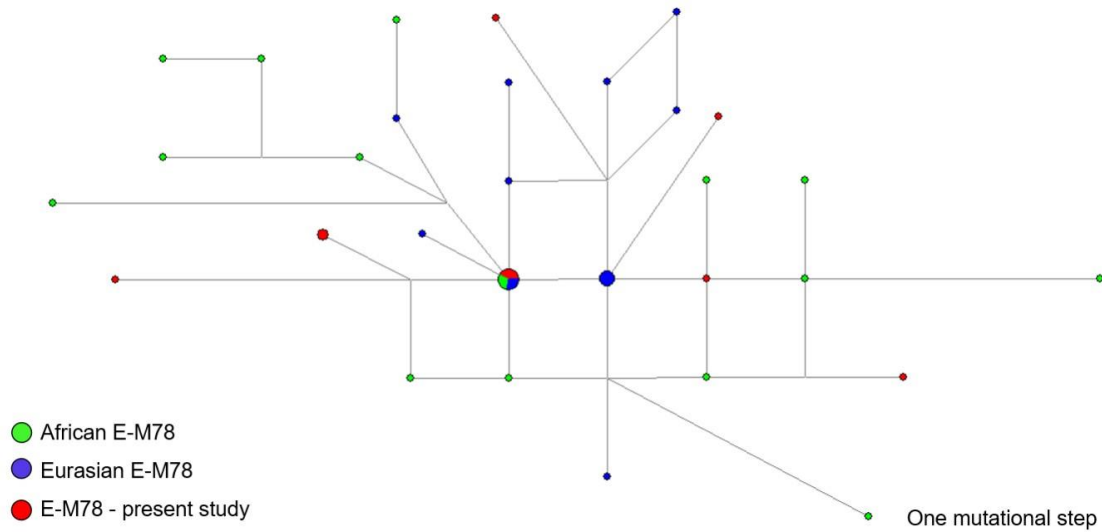
Figure 18: Haplogroup E-M78 phylogenetic network. Network was constructed with 10 Y-STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439) common to all studies. In the network are included 16 samples from the present study dataset from lineages E-M78; 14 Europeans (Adams et al., 2008; Boattini et al., 2013; Rębała et al., 2013), and 16 African (De Filippo et al., 2011; Larmuseau et al., 2015b; Haber, et al. 2011) belonging to E-M78. Circle size is proportional to the number of haplotypes.

Analysing the network (Fig. 18), it was not possible to clearly separate the European from sub-Saharan African lineages. As it is clear in the figure, three samples (one reference African, one reference Eurasian and one from the present study) shared the same haplotype. Thus, considering the historical background of Colombia it is most probable that E-M78 chromosomes present in this study have a Eurasian origin, providing by European colonizers or brought by immigrants.

Two parallel sub-haplogroups to E-M78 were found. One Andean sample was assigned to E-M310* sub-haplogroup. E-M123 sub-lineages were also identified. E-M34* sub-lineage was identified in two Caribbean samples and the sub-lineage E-FGC18329* was found in one sample from Caribe and in two samples from Andes. E-M123 is common in East Africa, being brought to Eurasia by a Middle Eastern expansion (Cruciani et al., 2004; Semino et al., 2004). The most probable origin of E-M123 lineages in our sample is Europe given the demographic and historical context of Colombia, as stated before. However, as E-M35* lineages are frequent in both Eurasian and Africa,

the origin of E-M310* and E-M123 lineages are unclear. In order to assess the most likely origin of these lineages found in Colombia, E-M35* network was constructed (Fig. 19). In Figure 19, it is evident that the Y-chromosomes E-M310* and E-M123 are spread out in the middle of the samples belonging to the Eurasian and African E-M35* lineages. However, as mentioned above, E-M123 lineages are especially common around the Horn of Africa (East Africa), and not in West Africa. Together with the results observed in the Network, it is doubtful that the Colombian E samples show a sub-Saharan origin owing to the latest migratory flows. Moreover, Semino et al., in 2004, reported that European E-M123 lineages indicate ''expansions from the Middle East toward Europe''. For these reasons, the samples were considered as part of the M35* lineages introduced in Colombia by Europeans.
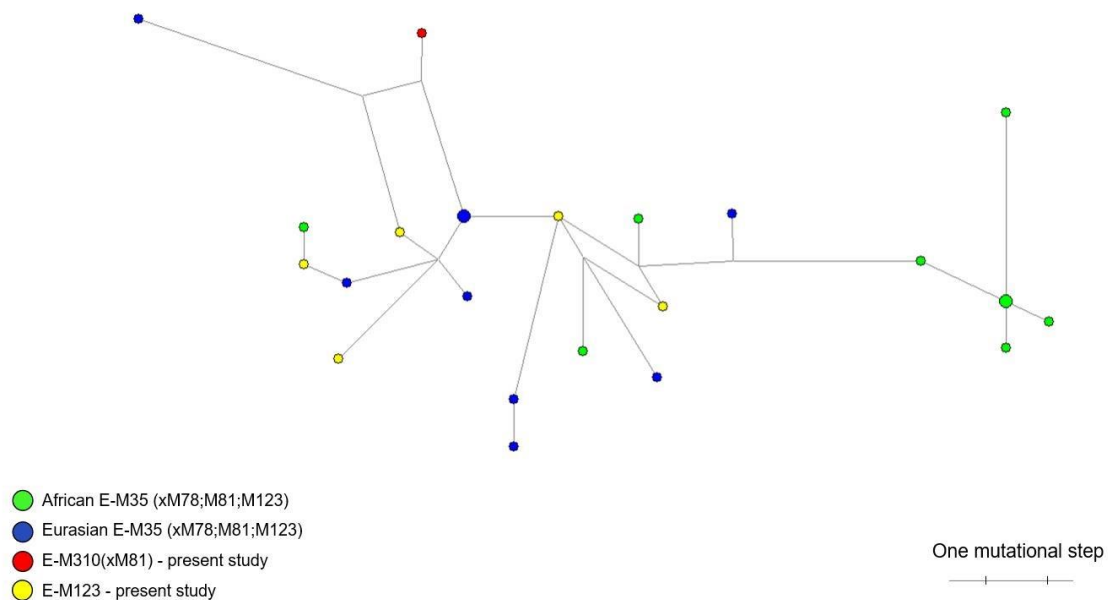
Figure 19: Haplogroup E-M35*(xM78; M81; M123) phylogenetic network. Network constructed with 10 Y-STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, and DYS439) common to all studies. In the network are included 6 samples from the present study dataset from lineages E-M310*(xM81) and E-M123; 9 Europeans (Adams et al., 2008; Boattini et al., 2013; Rębała et al., 2013), and 9 African (De Filippo et al., 2011; Larmuseau et al., 2015b; Haber, M., et al.,2011) belonging to E-M35*. Circle size is proportional to the number of haplotypes.

Overall, the Caribe population showed a higher diversity of E sub-lineages when comparing with Andes group (Fig. 20). Moreover, 16.33% of the Caribbean Y chromosomes were from sub-Saharan African origin, contrasting with only 2.90% of the Andean lineages. This is in accordance with the results from Ossa and coworkers, in 2016, which also reported a significant African ancestry (23%) in the Caribbean Region for autosomal markers. The Caribbean populations are known to have a strong African ancestry due to Cartagena's role in the slave trade that strongly shaped the genetic background of this population. Moreover, Noguera et al., in 2013, reported that 25% of the Y chromosomes from the Northern coastal Colombian region presented a very high sub-Saharan ancestry when compared to other admixed populations from South America.
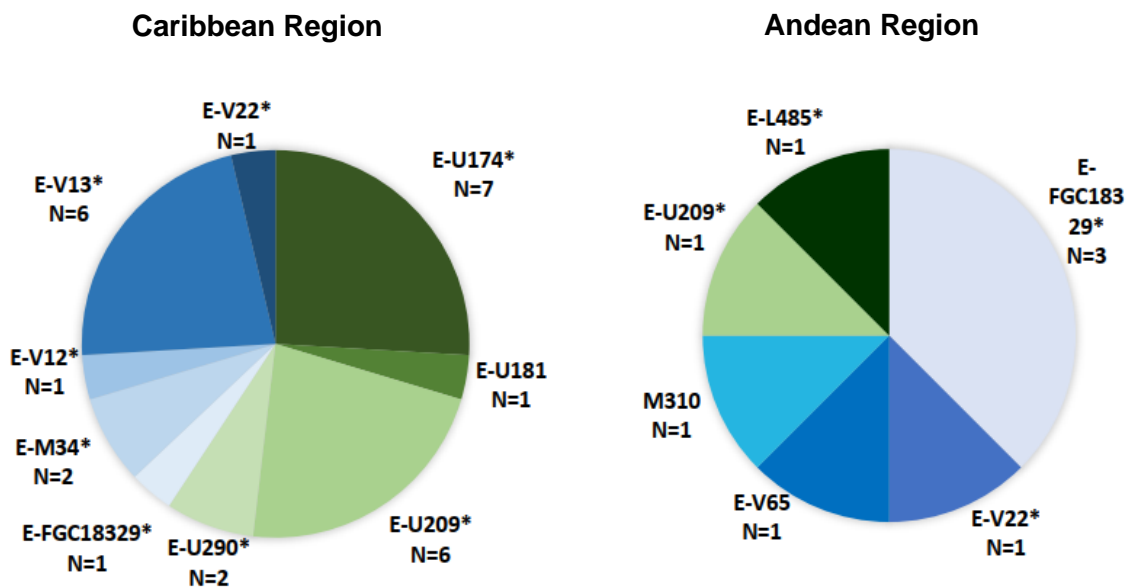
Figure 20: E haplogroup sub-lineages present in Caribbean and Andean regions. Sub-Saharan African lineages are represented in different shades of green, and Eurasian lineages are represented in different shades of blue.

4.3.3.3. Haplogroups G, I and J

Major haplogroup G defined by M201 SNP marker has a possible origin in the Middle East (Rootsi et al., 2012; Semino et al., 2000). Three samples from haplogroup G were assigned to the G2a2 sub-haplogroup (one sample from the Andes and two samples from the Caribe), representing 1.80% of our dataset. The haplogroup G marks its presence, as stated by Rootsi et al, in 2012, most commonly in the region of Caucasus. However, it is also frequently found in the Middle Eastern and South Europe (Balanovsky et al., 2011; Gonçalves et. al., 2015; Hammer et al., 2009).

With regard to haplogroup I, the dataset reveals a frequency of 13.77% of Y chromosomes belonging to this main haplogroup. Haplogroup I, defined by the M170 marker, is a European lineage that is very frequent in the Balkans, Eastern and Northern Europe (Karafet et al., 2008). All samples were allocated to the I-M253 and I-M438 sub-haplogroups. The results showed that two samples from the Andes were assigned to haplogroup I-M253: one to I-DF29* and another to I-Z63*. The haplogroup I-M253 appears mostly in Northern Europe, with its highest frequencies in Scandinavian populations (Rootsi et al., 2004). Only one sample belonging to haplogroup I-Z63 was found in the Caribe. Sub-haplogroup I-M438 was the most frequent. A total of three samples (two from the Caribe and one from the Andes) were identified in downstream lineages as I-M26. This sub-haplogroup has a very high occurrence within the island of Sardinia in Italy (Semino et al., 2000). Furthermore, it has been detected, in lower frequencies, in Castile, Bearnais (France) and in the Basques (Rootsi et al., 2004). Sub-haplogroup I-L621 had the highest incidence in our dataset. It was only detected in the Caribe, in 14 samples. This haplogroup is typical of Slavic populations and has high frequencies in Southeastern Europe. Four more I2 sub-haplogroups were identified [I-CTS9183*, I-L1229, I-L702*, I-CTS6433*] with only one sample assigned to each sub-haplogroup.

In the Colombian populations, 7.78% of the samples were identified as belonging to haplogroup J. This major haplogroup is defined by M304 Y SNP marker commonly found in Anatolia and the Middle East (Cinnioglu et al., 2004; Di Giacomo et al., 2004; Semino et al., 2004). Here 10 sub-haplogroups were identified inside major branches J-M267 and J-M172 clades (Cinnioğlu et al., 2004; Underhill et al., 2000). From J-M267 clade, two sub-haplogroups were identified, one in Caribe (J-Z2329*) and one in Andes (J-Z640). From J-M172 clade, eight different haplogroups were assigned to eleven Y-chromosomal samples from both Caribe and Andes: J-L559*; J-L26*; J-M322; J-M67*; J-L398*; J-PF5008*, J-L283 and J-Z2221.

Although haplogroups G, I and J can be found in other regions, in the context of the history of Caribbean and Andean populations, it was considered that all the sub-lineages detected were carried by Europeans to Colombia.

4.3.3.4 Haplogroup T

The less frequent haplogroup found in this study was the haplogroup T-M70 (1.20%). From this major haplogroup only two Andean samples were identified with T-L162 and T-L131 lineages. Haplogroup T has a Middle Eastern origin from where it expands, being present in nowadays populations from Africa and Europe (Elkamel et al., 2021; Mendez

et al., 2011). The two samples from Andean region classified as T-M70 sub-lineages most probably came from Europe as result of European migrations to Colombia.

4.3.3.5. Haplogroup Q

The major founder haplogroup of Native Americans is the haplogroup Q defined by the Y-SNP M242, a transition from C to T identified in Native Americans and Siberian Eskimos (Karafet et al., 2008; Seielstad et al., 2013; Underhill et al., 1996). Within the clade Q, only Q-M3* was found in the present study with a low frequency in the total dataset (5.99%). Only five samples were identified in each Caribbean and in Andean region. The haplogroup Q is found in approximately 85% of Native Americans (Jota et al., 2016). Inside this haplogroup, the Q-M3 and Q-CTS1780 are the main and founder lineages, being ubiquitous throughout the American continent (Grugni et al., 2019; Zegura et al., 2004). The low frequencies in Caribe and Andes are in accordance with the historical records that show a low Y-chromosomal Native ancestry due to the interbreeding between Native women and non-native men, mostly European males.

4.4. Colombian Ancestry

Andes had the highest frequency of European lineages (89.86%) followed by Native American lineages (7.25%) and 2.90% of sub-Saharan lineages. On the other hand, Caribe presented 78.57% of European lineages, followed by 16.33% of African sub-Saharan lineages and 5.10% of Native American lineages. The lineages from haplogroups R and I (63% of the total) were inferred as introduced exclusively by European males during the Iberian colonization, later, by the more recently European migrations into the country, as stated before. These results are in accordance with the literature that shows that European lineages are the main contributor of the paternal ancestry in Colombia due to the Iberian colonization of this country (Alonso-Morales et al.,2018). On the other hand, a Middle Eastern source of lineages from haplogroups E, G, J and T cannot be fully excluded. However, concerning the historical and migratory movements to Colombia, the introduction of these lineages was most certainly mediated by European males. Thus, in the present study we considered that Europe was the source of these lineages brought to the Caribe and Andes. The Caribbean region had the highest frequency of sub-Saharan African lineages, which, as mention before, can be explained by the role of Cartagena as one of the most important ports in South America during the slave trade. Regarding the Native American component, studies of

uniparental markers revealed a distinct ancestry distribution in the mtDNA or Y chromosomal lineages, as most Mestizos resulted from admixture involving predominantly Native women and European men (Marcheco-Teruel et al., 2014; Rojas, W., et al., 2010). Therefore, a low frequency of Native lineages was expected in our dataset.

4.5. Population comparisons

As previously mentioned, Colombian data set (Caribe and Andes) was compared with Y-STR and Y-SNP data collected from other Admixed South American, Eurasian and African populations available in the literature (Tables S3-S7, Appendix). For both Y-STR and Y-SNP comparisons, the set of markers was reduced to those common to all studies.

4.5.1. Comparisons between admixed South American Populations

The South American countries have a close genetic composition, as a result of a similar geopolitical and demographic history. Nonetheless, genetic makeup varies among current populations as a result of different genetic dynamics and migratory flows. Thus, in the scope of this work, different admixed populations from this subcontinent were compared, in order to understand and frame the similar or divergent influences felt throughout the South American territory.

For that reason, 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), and F_{ST} genetic distances based on 13 common bi-allelic markers (A-M13, A4-SRY10831.1*; E-M96*(xM35), E-M35, FM213*(xM201, M170, 12f2a, M9), G-M201, I-M170, J-12f2a, KLNT-M9, Q-M242*(xM3), Q-M3, R-M207*(xM343) and R-M343) (Table S10 and S11 in Appendix).

According to pairwise genetic distances (R_{ST} and F_{ST}) there are statistically significant distances ($p \leq 0.00029$ and $p \leq 0.00042$, respectively) between Colombian and other admixed populations with the Native and sub-Saharan African populations. However, Bolivia (Mestizo) was the only exception presenting a non-statistically significant distance with Native-Bolivian population ($R_{ST}= 0.01868$; $p=0.05485$). These significant distances might be explained by the higher frequency of haplogroup Q lineages in the Bolivian Native population, as well as the higher occurrence of haplogroup E in the African population compared to the lower frequencies observed in admixed

populations. According to Fig. 21A, a shared cluster between Andes, admixed populations, and the European population of Spain is detected. This is mainly due to a high Iberian influence due to European colonization of South American sub-continent in the beginning of XV century.
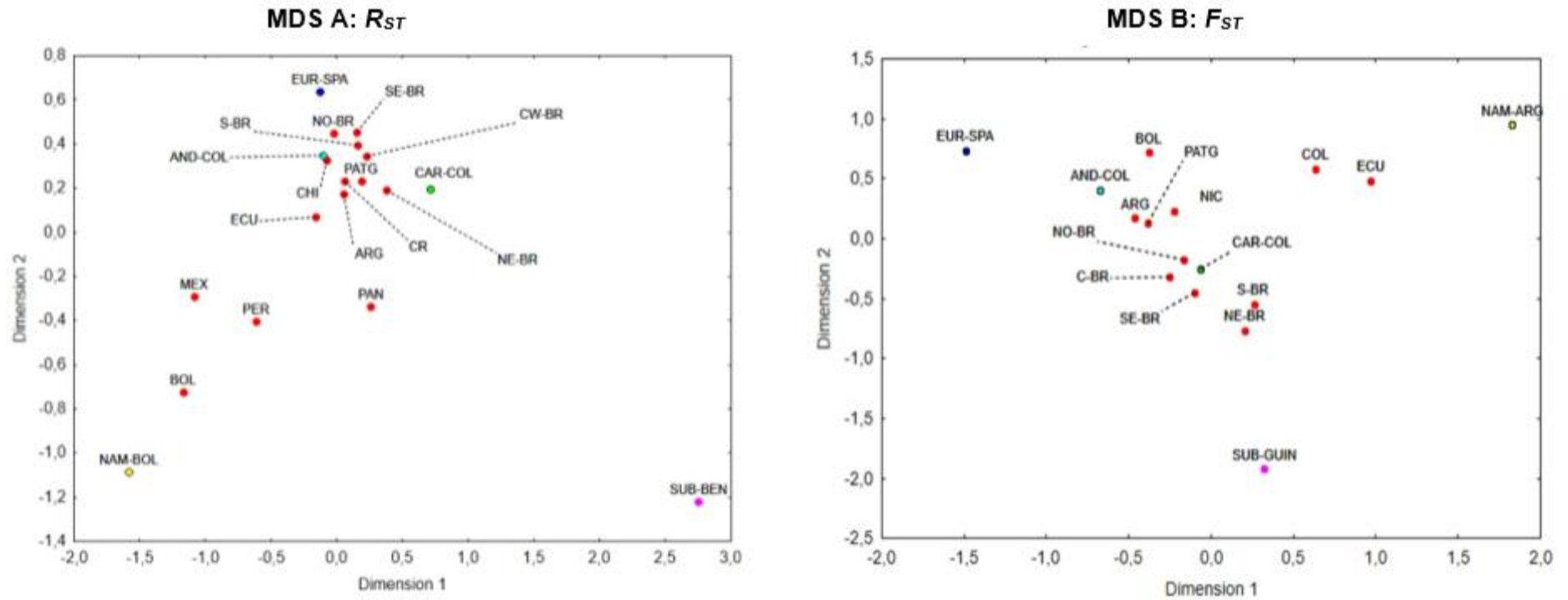
Figure 21: MDS graph of Admixed South American Populations comparisons. A) MDS plot based on R_{ST} genetic distances between the Caribe and Andes populations and other South American populations, based on 15 Y-STRs (Stress=0.03074). Caribe (CAR-COL) region is represented in green and Andes (AND-COL) in neon blue. The admixed populations of South America were plotted in red (SE-BR- Southeastern Brazil; CW-BR-Central West Brazil; NO-BR-North Brazil, S-BR-South Brazil; NE-Brazil-Northeast; PATG-Patagonia; CHI-Chile; ECU-Ecuador; ARG-Argentina; CR-Costa Rica; PAN-Panama; MEX-Mexico; PER-PERU; BOL-Bolivia). The European (EUR-SPA- Spain), Native American (NAM-BOL-Native, Bolivia), and African (SUB-BEN-sub-Saharan Benin) populations are shown dark blue, yellow and pink, respectively. (B) MDS plot based on F_{ST} genetic distances between the Caribe and Andes populations and other South American populations, based on 13 Y-SNPs (Stress= 0.0611). Caribe (CAR-COL) region is represented in green and Andes (AND-CAR) in neon blue. The admixed populations of South America were plotted in red (PATG- Patagonia, ARG- Argentina, BOL- Bolivia, NO-BR- North Brazil, NE-BR- Northeast Brazil, SE-BR- Southeast Brazil, S-BR- South Brazil, C-BR-Center Brazil; COL-Colombia, ECU- Ecuador, NICA-Nicaragua). A European population is represented in blue (EUR-SPA-Spain); a Native American is represented in yellow (NAM-ARG -Argentina) and, finally, a Sub-Saharan African population is represented in pink (SUB-GUIN-sub-Saharan Equatorial Guinea).

It is also important to highlight the clustering between the Andes and the Chilean populations (Fig. 21A). The distance measured between the two populations was very low (R_{ST}=0.00119 and p=0.29363) and not significant, which allowed to conclude that they are two similar populations. This can be explained by the fact that both populations present a very high European Y chromosomal ancestry (91.7% in Chile and 81.6% in Andes) (Toscanini et al, 2016a).Also, close to Andes, in both graphs, are Patagonia (R_{ST}=0.00990 and p=0.05792; F_{ST}=0.01463 and p=0.03970) and Argentina (R_{ST}=0.00549 and p= 0.10593; F_{ST} =0.00566 and p=0.15216), also, correlated with the fact that these three populations shared a common European paternal ancestry, which can be more specifically understood by the high occurrence of R-M243 lineages.

It is important to note that Caribe is more distanced from the Spanish population (F_{ST}= 0.18604 and p=0.00000) rather from the sub-Saharan population (F_{ST}=0.11843 and p= 0.00000), which can be seen in Fig. 21B as Caribe tends to get further away from the other admixed populations with high Spanish influence. As previously reported, Caribe revealed a significantly higher frequency of sub-Saharan lineages compared to Andes. The closest population to Caribe in MDS A is Northeast Brazil (R_{ST}= 0.01391 and p=0.00426). On the other hand, in Fig.21B, from all five analysed regions from Brazil, Northeast Brazil was the most distant from Caribe (F_{ST}=0.04833 and p=0.0000) and North Brazil was the most closed, presenting a non-statistically significant distance of F_{ST} =0.01335. Moreover, Southeast Brazil was found close to Caribe (F_{ST}=0.01814). Thus, after analysing the most frequent haplogroups in these regions, it was found that these regions of Brazil and the Caribe have similar frequencies of haplogroups belonging to haplogroup E-M96 (xM35), indicating a similar African influence in these populations. Similar frequencies of lineages within haplogroup R, reflecting the European influence in both South American countries, were also found.

4.5.2. Comparisons with Eurasian Populations

As shown previously, the European component was the most frequent in both Colombian populations. Therefore, in order to dissect and understand in greater detail the presence of these lineages in the Caribe and the Andes, a comparative analysis was carried out between European subset of the Colombian samples (Caribe, N=77; and Andes, N=61), European and two Middle Eastern populations. R_{ST} pairwise genetic distances were analysed for 21 common Y-STRs (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS481, DYS533, DYS635, DSY576, DSY549, DSY570, DSY643 and Y GATA H4), and Y-SNP data with 11 Y-SNPs: E-M35*(xM78, M81, M123),E-M78, E-M81,

EM123, G-M201, I-M170, J-12f2a, KLNT-M9, R-M207*(xSRY10831.2,M343), R.SRY10831.2 and R-M343 (Tables S12 and S13, Appendix).

The Caribe and Andes, as well the Eurasian reference populations, present statistically significant R_{ST} distances ($p\leq0.00014$) with Lebanon, with the exception of Galicia and Ireland ($R_{ST}=0.12625$, $p=0.00059$; $R_{ST}=0.12021$, $p=0.00366$, respectively). However, although the p-values are not statistically significant after Bonferroni correction, due to the low effective population size, Galicia and Ireland share a high genetic distance with Lebanon. For F_{ST} values, all Eurasian populations and the two Colombian ones showed statistically significant differences ($p\leq0.00029$) with Lebanon and Turkey. This is due to the high frequency of J-12f2a haplogroup in both Middle Eastern regions, also explaining the closeness between them ($F_{ST}=0.01821$ and $p=0.00000$).

In addition, non-statistically significant distances between Andean and Portuguese populations from North ($R_{ST}=-0.01222$ and $p=0.57955$), South ($R_{ST}=-0.01119$ and $p=0.59073$) and Center ($R_{ST}=-0.00546$ and $p=0.39501$) were observed (Fig. 22A). On the other hand, pairwise F_{ST} distances shown similar distances between Andes and Spain (Catalunya: $F_{ST}=0.01755$ and $p=0.07326$; Galicia: $F_{ST}=0.00735$ and $p=0.16790$; Castilla-La Mancha $F_{ST}=-0.00143$ and $p=0.42412$ and Andalusia: $F_{ST}=-0.00497$ and $p=0.68409$), all of them not statistically significant. These results corroborated, again, the Iberian influence in the Andean population due to the Iberian colonization. Moreover, in MDS B) it is possible to detect a closeness between Andean and French population ($F_{ST}=-0.00325$ and $p=0.63311$) that is explain by the high occurrence of R-M243 lineages in both populations, reflecting the influence of the recent European migration flow in Colombia. Concerning Caribe, this region is close to the Balkan populations (Hungary, Croatia and Czechia) (Figure 22A and 22B). On the other hand, it can also be verified in MDS B that Germany is close to Caribe. The low genetic distance observed ($F_{ST}=-0.00098$; $p=0.44857$) can be due to the high frequencies of haplogroup I-M170 lineages in both populations. As reported earlier, haplogroup I is very frequent in the Balkans, Eastern and Northern Europe (Karafet et al., 2008) reflecting the impact of the more recent migrations influxes from Europe to Caribe. It is also important to highlight the influence of the Italian population in the two Colombian regions. In MDS B, it is possible to see that Southern Italy is quite distant from the Andean and Caribbean regions ($F_{ST}=0.10666$ and $F_{ST}=0.05314$, respectively), unlike the Northern Italian region ($F_{ST}=0.01848$ and $F_{ST}=0.01805$) that is very close to both Colombian populations. Moreover, the Southern Italian region tends to get close to the Turkey population ($F_{ST}=0.02061$), revealing a possible Middle Eastern influence.
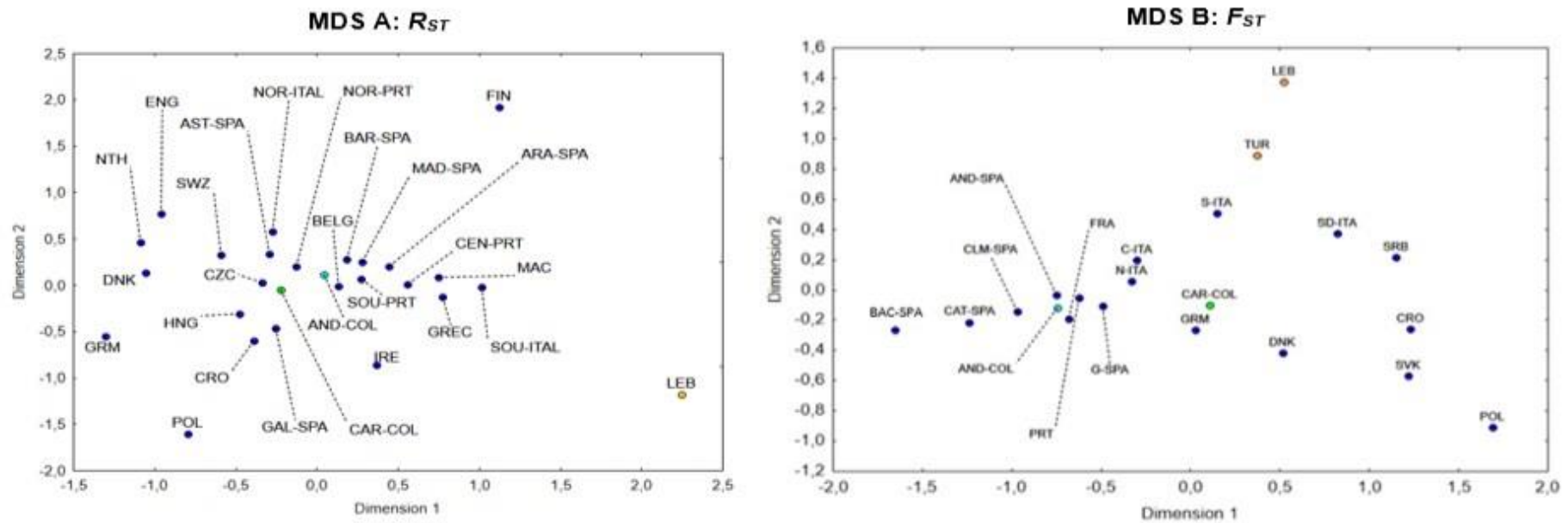
Figure 22: MDS graph of Eurasian Populations comparisons. A) MDS plotted based on the R_{ST} genetic distances from the Eurasian subset of the Colombian analysed samples calculated using the information of 21 Y-STRs (Stress= 0.141). Caribbean region (CAR-COL) is represented in green and the Andes (AND-COL) in neon blue. The European populations used as reference were plotted in blue (BELG- Belgium; CRO-Croatia; CZC-Czechia; DNK-Denmark; ENG-England; ARA-SPA-Arágon, Spain; AST-SPA-Asturias, Spain; BAR-SPA-Barcelona, Spain; GAL-SPA-Galicia, Spain; MAD-SPA-Madrid, Spain; GRM-Germany; GRE-Greece; NOR-ITAL-North Italy; SOU-ITAL-South Italy; NTH-Netherlands; POL-Poland; NOR-PRT- North Portugal; CEN-PRT-Center Portugal; SOU-PRT-South Portugal; SWZ-Switzerland; IRE-IRELAND; MAC-Macedonia; FIN-Finland.)Middle Easten population from Lebanon (LEB) was used and it is plotted in orange. B) MDS plot based on the F_{ST} genetic distances between the Eurasian subset of the Colombian analysed samples and European and Middle Eastern populations, based on 11 Eurasian Y-SNPs (Stress=0.067). Andes (AND-COL) is plotted in neon blue and Caribe (CAR-COL) in green. European populations are shown in blue (PRT-Portugal; CAT-SPA-Catalunia, Spain; GAL-SPA-Galicia, Spain; CLM-SPA, Castilla-La Mancha , Spain; BAC-SPA-Baque Country, Spain; N-ITAL-North Itay; C-ITAL-Center Italy; S-ITAL-South Italy; SD-ITAL-Sardinia, Italy; GRM-Germany; POL-Poland; SVK-Slovakia; DEN-Denmark; CRO-Croatia; SRB-Serbia and FRA-France) Middle Easten population from Lebanon (LEB) and Turkey (TUR) were used and are plotted in orange.

It is also important to highlight the influence of the Italian population in the two Colombian regions. In MDS B, it is possible to see that Southern Italy is quite distant from the Andean and Caribbean regions (F_{ST}=0.10666 and F_{ST}=0.05314, respectively), unlike the Northern Italian region (F_{ST}=0.01848 and F_{ST}=0.01805) that is very close to both Colombian populations. Moreover, the Southern Italian region tends to get close to the Turkey population (F_{ST}=0.02061), revealing a possible Middle Eastern influence.

4.5.3. Influence of R-M269 lineages

The clade R-M269 is highly prevalent in Western Europe yet being widespread in Middle Eastern too. Therefore, in order to understand more concisely the influence of this sub-lineage in the Colombian regions and to be able to infer its possible origin, F_{ST} genetic distances (Table S14, Appendix) were calculated between both Colombian populations and other populations from Europe and the Middle East with data available for haplogroups R-M269*(xL23), R-L23*(xU106, S116), R-U106, R-S116*(xU152, M529), R-U152, and R-M529.
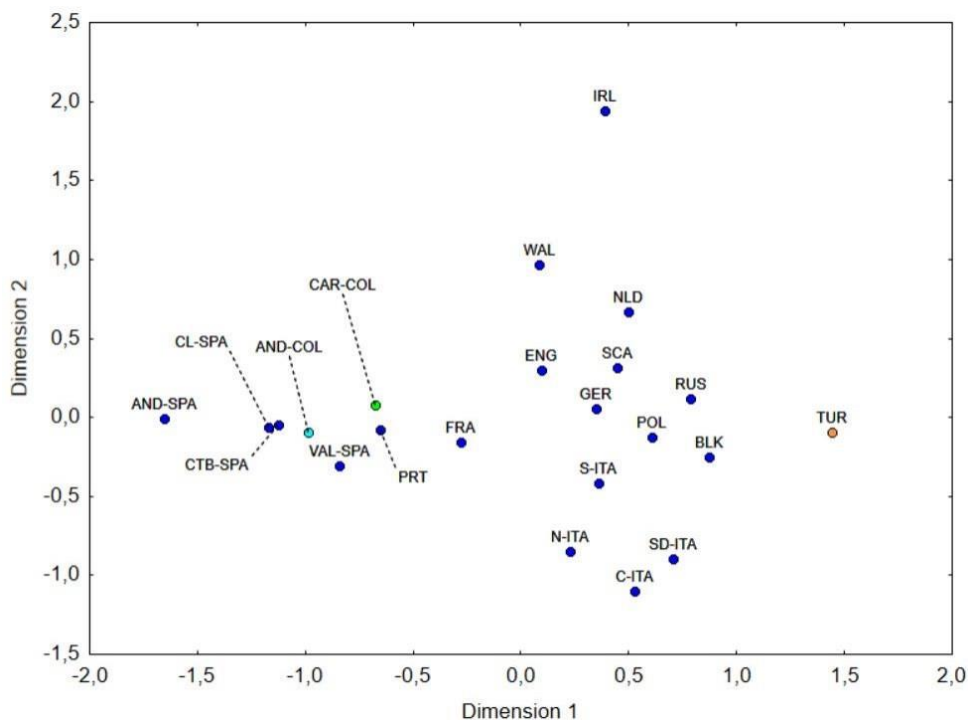
Figure 23: MDS graph of Eurasian populations with R-M269 linages. MDS plot based on F_{ST} genetic distances between the Caribbean, Andean and other South American populations, based on 6 Y-SNPs (Stress=0.1232). Caribe (CAR-COL) region is represented in green and Andes (AND-COL) in neon blue. The reference populations (European) were plotted in blue (BLK- Balkan Peninsula, GER- Germany, ENG- England, AND-SPA- Andalusia ,Spain, CL-SPA- Castilla-La Mancha, CTB-SPA- Cantabria, Spain, VAL-SPA- Valencia ,Spain, FRA- France, IRL- Ireland, N-ITA- Northern Italy, C-ITA- Central Italy, S-ITA, Southern Italy, SD-ITA- Sardinia, Italy, NLD- t Netherlands, POL- Poland, PRT- Portugal, RUS- Russia, SCA- Scandinavia and WAL- Wales). The Middle Eastern population was colored in orange (TUR-Turkey).

Regarding the frequencies of R-M269 sub-lineages, Caribe and Andes seem to be quite alike and close to each other (F_{ST}=0.00877 and p=0.26502). The Colombian samples are close to Portugal as well as to Cantabria, Castille and Valencia (Spanish populations). The French population is closer to Caribe (F_{ST}=0.03454 and p=0.02594) than to Andes (F_{ST}=0.07226 and p= 0.00079). At the same time, Andalusia is closer to Andes (F_{ST}=0.06123 and p=0.00851) but far from France (F_{ST}=0.19450 and p=0.00000). In Figure 24, it is also observed that the Italian populations are distant from both Colombian populations which fact is due the low frequency of Haplogroup R-M269 lineages in Italy.

In order to investigate which sub-haplogroups have the greatest influence in the differentiation of the populations, a PCA was performed, using the frequency of each lineage within populations, and the results obtained for the two principal components of variance were plotted (Fig. 26). The first and second dimensions account for 35.65% and 24.30% of the total variance.
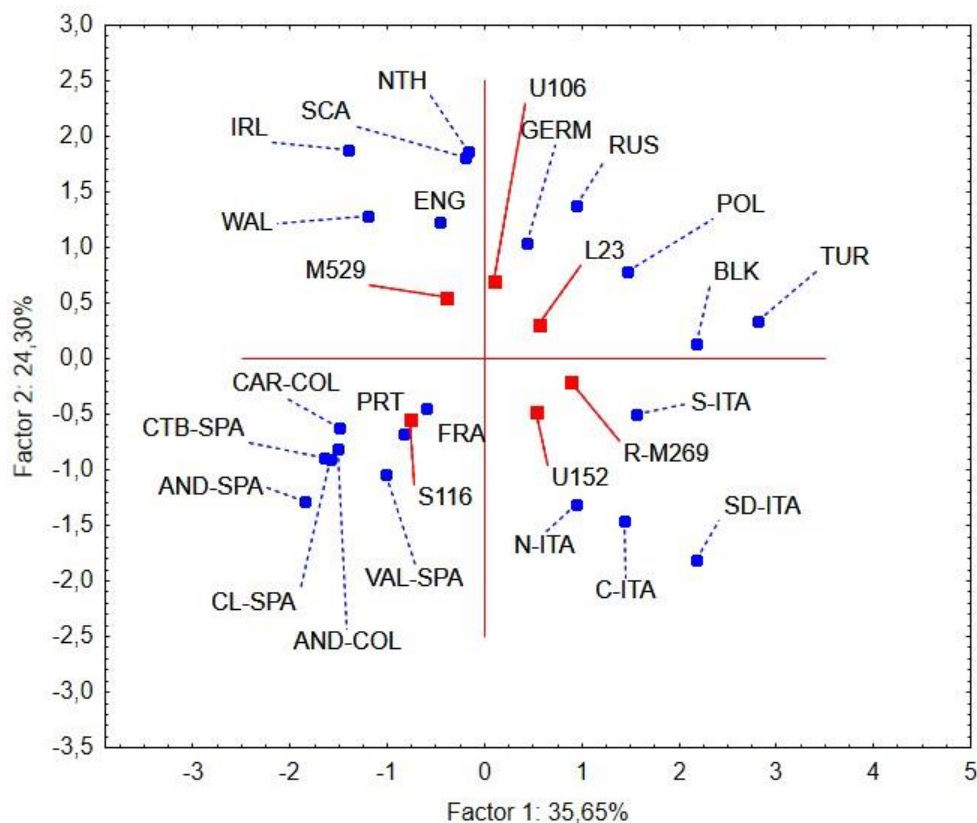
Figure 24: Principal component analysis based on R-M269 sub-haplogroups from Europe and the Middle East. The Colombian populations along with other Europeans and Middle Eastern samples are colored in blue (Caribe-COL- Caribe, Andes-COL-Andes, BLK- Balkan Peninsula, germ- Germany, ENG- England, AND-SPA Andalusia, Spain, CL-SPA- Castile and Leon, Spain; CTB-SPA- Cantabria ,Spain, SPA-VAL, Valencia ,Spain, FRA- France, IRL- Ireland, N-ITAL- Northern Italy, C-ITA- Central Italy, S-ITA- Southern Italy, SD-ITA- Sardinia ,Italy, NTH- the Netherlands, POL- Poland, PRT- Portugal, RUS- Russia, SCA- Scandinavia, TUR- Turkey, and WAL- Wales). Sub-haplogroups M529; U106; S116; L23; U-U152 and R-M269 are colored in red.

Caribe and Andes are both located in the same quadrant as the four populations of Spain, the French population and the Portuguese one, where S116 is the lineage with the highest frequency. This result confirms the high influence of the Iberian colonization in the genetic background of Colombia as well as the more recent migration flux from Europe to Southern America. Moreover, it was confirmed what was previously reported about the non-influence of Italy in haplogroup R in both Colombian populations as Italy has higher R-M269 and U152 lineages. On the other hand, Germany and the Balkans that were previously related with Caribe in Eurasian comparisons, here, they are located in another quadrant where U106 lineage is the most frequent. This result proves that R-M269 sub-linages present in Caribe are not introduced by Germany and the Balkans.

5. Conclusion

Colombia portrays and reflects its historical diversity, from its settlement to the present day, in its genetic background demarcated by an immense diversity of Natives, Africans and Europeans. In its early days, and until the European colonization, Colombia was only inhabited by the Native American people. It was in the end of 15th century that the country was invaded by European colonizers, which consequently started the slave trade to Colombia. From this moment on, the previously exclusively Native genetic background began to change, and the African and European components start being incorporated in different proportions across the various regions of the country.

The lineages observed in both Caribe and Andes were inferred to be mostly of European origin (83.23%), followed by sub-Saharan African origin (10.78%) and with less frequency of Natives (5.99%).

From genetic comparisons performed, Andes as well as several other South American populations were found to be closer to Spanish populations, being Andes very close to Argentina populations (Argentina and Patagonia reference populations), both South American admixed populations. The Caribbean region of Colombia showed that, although highly influenced by European lineages, it is closer to the sub-Saharan populations, as well as to the South American admixed populations from Brazil. Yet, both Colombian regions were found to be far from the Native American population.

Concerning the European influence, the analysis showed that the Andes is very close to the Iberian populations (Portugal and Spain) and to France. On the other hand, this analysis showed that the Caribe is closer to populations such as Germany and Balkan populations. Furthermore, the Middle Eastern populations were very far from all the European populations and the Colombian regions under study.

To dissect the European influence, a more in-depth analysis of lineages from haplogroup R-M269 was carried out. Both Colombian regions were close to Spanish Portuguese and French populations, being the R-S116 lineage a marked influence in these populations.

It is important to note that the level of resolving of both Y-STR haplotypes and Y-SNP haplogroups had to be reduced due to the scarcity of studies carried out with the same number of markers included in the present study, which influenced the results. It should therefore be noted that there is a need to continue carrying out studies in these and other South American populations with a higher number of markers to be able to infer and corroborate facts in greater detail. However, given these results, it is possible

to conclude that Colombia is a country with great male genetic diversity. Furthermore, these conclusions allowed to understand how historical and more recent events marked the genetic background of this population. It should also be noted that the genotyping of Y-SNPs using a latest generation technique will encourage the study of more populations for larger sets of Y-SNPs, enhancing knowledge in the areas of Forensic and Population Genetics.

References

Adams, S. M., Bosch, E., Balaresque, P. L., Ballereau, S. J., Lee, A. C., Arroyo, E., López-Parra, A. M., Aler, M., Grifo, M. S., Brion, M., Carracedo, A., Lavinha, J., Martínez-Jarreta, B., Quintana-Murci, L., Picornell, A., Ramon, M., Skorecki, K., Behar, D. M., Calafell, F., & Jobling, M. A. (2008). The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *American journal of human genetics*, *83*(6), 725–736. https://doi.org/10.1016/j.ajhg.2008.11.007

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. https://doi.org/10.1371/journal.pone.0207130

Altena, E., Smeding, R., van der Gaag, K. J., Larmuseau, M. H. D., Decorte, R., Lao, O., Kayser, M., Kraaijenbrink, T., & de Knijff, P. (2020). The Dutch Y-chromosomal landscape. *European journal of human genetics: EJHG*, *28*(3), 287–299. https://doi.org/10.1038/s41431-019-0496-0

Amorim, A. (2019). Nonhuman forensic genetics. *Forensic Science International: Genetics Supplement Series*, *7*(1), 44–46. https://doi.org/10.1016/j.fsigss.2019.09.019

Amorim, A. & B. Budowle. (2016). Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation (pp. 163-191). Types of genomes, sequences and genetic markers. *In World Scientific. Security Science and Technology Vol. 2.*https://doi.org/10.1142/9781786340788_0009

Andersen, M. M., & Balding, D. J. (2019). Y-profile evidence: Close paternal relatives and mixtures.*Forensic Science International: Genetics, 38, 48–53.* doi: 10.1016/j.fsigen.2018.10.00

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

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

Assyuhada, M. G. S. N., Zulkafli, Z., Noor, N. H. M., Pati, S., Kari, Z. A., Dawood, M. A. O., Chambers, G. K., & Edinur, H. A. (2023). Population genetics and human health in the genomic era. In *Journal of King Saud University - Science* (Vol. 35, Issue 4). Elsevier B.V. https://doi.org/10.1016/j.jksus.2023.102593

Ayub, Q., Haber, M., & Tyler-Smith, C. (2016). Evolutionary and Population Genetics in Forensic Science In A. Amorim, & B. Budowle (Eds.), Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation (pp. 33-60).

World Scientific. Security Science and Technology Vol. 2. https://doi.org/10.1142/9781786340788_0009

Bachtrog D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature reviews. Genetics*, *14*(2), 113–124. https://doi.org/10.1038/nrg3366

Balanovsky, O., Dibirova, K., Dybo, A., Mudrak, O., Frolova, S., … Pocheshkhova, E. (2011). Parallel Evolution of Genes and Languages in the Caucasus Region*. Molecular Biology and Evolution, 28(10), 2905–2920.* doi:10.1093/molbev/msr126

Balanovsky, O., Zhabagin, M., Agdzhoyan, A., Chukhryaeva, M., Zaporozhchenko, V., Utevska, O., Highnam, G., Sabitov, Z., Greenspan, E., Dibirova, K., Skhalyakho, R., Kuznetsova, M., Koshel, S., Yusupov, Y., Nymadawa, P., Zhumadilov, Z., Pocheshkhova, E., Haber, M., Zalloua, P. A., Yepiskoposyan, L., … Balanovska, E. (2015). Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation rates on the human Y-chromosome and reveals migrations of Iranic speakers. *PLOS ONE*, *10*(4), e0122968. https://doi.org/10.1371/journal.pone.0122968

Barać, L., Peričić, M., Klarić, I. M., Rootsi, S., Janićijević, B., Kivisild, T., Parik, J., Rudan, I., Villems, R., & Rudan, P. (2003). Y chromosomal heritage of Croatian population and its island isolates. *European journal of human genetics: EJHG*, *11*(7), 535–542. https://doi.org/10.1038/sj.ejhg.5200992

Bartosik-Vélez, E. (2014). The Legacy of Christopher Columbus in the Americas: New Nations and a Transatlantic Discourse of Empire

Battaglia, V., Grugni, V., Perego, U. A., Angerhofer, N., Gomez-Palmieri, J. E., Woodward, S. R., Achilli, A., Myres, N., Torroni, A., & Semino, O. (2013). The first peopling of South America: new evidence from Y-chromosome haplogroup Q. *PLOS ONE*, *8*(8), e71390. https://doi.org/10.1371/journal.pone.0071390

Beleza, S., Gusmão, L., Lopes, A., Alves, C., Gomes, I., Giouzeli, M., Calafell, F., Carracedo, A., & Amorim, A. (2006). Micro-phylogeographic and demographic history of Portuguese male lineages. *Annals of human genetics*, *70*(Pt 2), 181–194. https://doi.org/10.1111/j.1529-8817.2005.00221.x

Berensmeier S. (2006). Magnetic particles for the separation and purification of nucleic acids. *Applied microbiology and biotechnology*, *73*(3), 495–504. https://doi.org/10.1007/s00253-006-0675-0

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mouguiama-Daouda, P., van der Veen, L., Hombert, J. M., Quintana-Murci, L., & Comas, D. (2009). Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Molecular biology and evolution*, *26*(7), 1581–1589. https://doi.org/10.1093/molbev/msp069

Bisso-Machado, R., Bortolini, M. C., & Salzano, F. M. (2012). Uniparental genetic markers in South Amerindians. *Genetics and molecular biology*, *35*(2), 365–387. https://doi.org/10.1590/S1415-47572012005000027

Boattini, A., Martinez-Cruz, B., Sarno, S., Harmant, C., Useli, A., Sanz, P., Yang-Yao, D., Manry, J., Ciani, G., Luiselli, D., Quintana-Murci, L., Comas, D., Pettener, D., & Genographic Consortium. (2013). Uniparental markers in Italy reveal a sex-biased

genetic structure and different historical strata. *PloS one*, *8*(5), e65441. https://doi.org/10.1371/journal.pone.0065441

Bond J. W. (2007). Value of DNA evidence in detecting crime. *Journal of forensic sciences*, 52(1), 128–136. https://doi.org/10.1111/j.1556-4029.2006.00323.x

Borges de Pinto Seguro Pereira, M. (2020). Characterisation of human male lineages from Ecuador. (Master thesis). *Available in Renates. (https://hdl.handle.net/10216/132092)*

Bostoen, K. (2018). The Bantu expansion. In *Oxford research encyclopedia of African history*. https://doi.org/10.1093/acrefore/9780190277734.013.191

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American journal of human genetics*, *32*(3), 314–331.

Bouakaze, C., Keyser, C., Amory, S., Crubézy, E., & Ludes, B. (2007). First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. *International journal of legal medicine*, *121*(6), 493–499. https://doi.org/10.1007/s00414-007-0177-3

Boyd-Bowman, P. (1976). Patterns of Spanish Emigration to the Indies until 1600. *The Hispanic American Historical Review, 56(4), 580.* doi:10.2307/2514194

Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., & Rolf, B. (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. https://doi.org/10.1086/301869

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. https://doi.org/10.1007/s00414-004-0439-2

Brookes A. J. (1999). The essence of SNPs. *Gene*, *234*(2), 177–186. https://doi.org/10.1016/s0378-1119(99)00219-x

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.

Bruijns, B., Tiggelaar, R., & Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis*, *39*(21), 2642–2654. https://doi.org/10.1002/elps.201800082

Budowle B. (2004). SNP typing strategies. *Forensic science international*, *146 Suppl*, S139–S142. https://doi.org/10.1016/j.forsciint.2004.09.042

Budowle, B., & van Daal, A. (2008). Forensically relevant SNP classes. *BioTechniques*, *44*(5), 603–610. https://doi.org/10.2144/000112806

Budowle, B., Eisenberg, A. J., & van Daal, A. (2009). Validity of low copy number typing and applications to forensic science. *Croatian medical journal*, *50*(3), 207–217. https://doi.org/10.3325/cmj.2009.50.207

Busby, G. B. J., Brisighelli, F., Sanchez-Diz, P., Ramos-Luis, E., Martinez-Cadenas, C., Thomas, M. G., … Capelli, C. (2011). The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proceedings of the Royal Society B: Biological Sciences, 279(1730), 884–892.* doi:10.1098/rspb.2011.1044

Butler, J. (2012). Advanced Topics in Forensic DNA Typing: Methodology, https://doi.org/10.1016/B978-0-12-374513-2.00022-1.

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

Campos Ribeiro, R. (2022). Characterization of Y chromosome diversity in Colombian populations (Master thesis). *Available in Renates.(https://hdl.handle.net/10216/147178)*

Capodiferro MR, Aram B, Raveane A, Rambaldi Migliore N, Colombo G, Ongaro L, Rivera J, Mendizábal T, Hernández-Mora I, Tribaldos M, Perego UA, Li H, Scheib CL, Modi A, Gòmez-Carballa A, Grugni V, Lombardo G, Hellenthal G, Pascale JM, Bertolini F, Grieco GS, Cereda C, Lari M, Caramelli D, Pagani L, Metspalu M, Friedrich R, Knipper C, Olivieri A, Salas A, Cooke R, Montinaro F, Motta J, Torroni A, Martín JG, Semino O, Malhi RS, Achilli A. (2021). Archaeogenomic distinctiveness of the Isthmo-Colombian area. *Cell*, *184*(7), 1706–1723.e24. https://doi.org/10.1016/j.cell.2021.02.040

Cárdenas, J. M., Heinz, T., Pardo-Seco, J., Álvarez-Iglesias, V., Taboada-Echalar, P., Sánchez-Diz, P., Carracedo, Á., & Salas, A. (2015). The multiethnic ancestry of Bolivians as revealed by the analysis of Y-chromosome markers. *Forensic science international. Genetics*, *14*, 210–218. https://doi.org/10.1016/j.fsigen.2014.10.023

Carey, S. B., Lovell, J. T., Jenkins, J., Leebens-Mack, J., Schmutz, J., Wilson, M. A., & Harkess, A. (2022). Representing sex chromosomes in genome assemblies. *Cell genomics*, *2*(5), 100132. https://doi.org/10.1016/j.xgen.2022.100132

Carratto, T. M. T., Moraes, V. M. S., Recalde, T. S. F., Oliveira, M. L. G., & Teixeira Mendes-Junior, C. (2022). Applications of massively parallel sequencing in forensic genetics. Genetics and molecular biology, 45(3 Suppl 1), e20220077. https://doi.org/10.1590/1678-4685-GMB-2022-0077

Carvalho Gontijo C, Porras-Hurtado LG, Freire-Aradas A, Fondevila M, Santos C, Salas A, Henao J, Isaza C, Beltrán L, Nogueira Silbiger V, Castillo A, Ibarra A, Moreno Chavez F, Söchtig J, Ruiz Y, Barreto G, Rondon F, Zabala W, Borjas L, de Oliveira SF, Carracedo A, Lareu MV, Phillips C. (2019). *PIMA: A population informative multiplex for the Americas. Forensic Science International: Genetics, 102200.* doi: 10.1016/j.fsigen.2019.102200

Casanova, M., Leroy, P., Boucekkine, C., Weissenbach, J., Bishop, C., Fellous, M., Purrello, M., Fiori, G., & Siniscalco, M. (1985). A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science (New York, N.Y.)*, *230*(4732), 1403–1406. https://doi.org/10.1126/science.2999986

Castro E Silva, 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 of the United States of America*, *118*(14), e2025739118. https://doi.org/10.1073/pnas.2025739118

Cavalli-Sforza, L. L., & Feldman, M. W. (2003). The application of molecular genetic approaches to the study of human evolution. *Nature genetics*, *33 Suppl*, 266–275. https://doi.org/10.1038/ng1113

Censo Nacional de Población y Vivienda - CNPV- 2018 - https://www.dane.gov.c

Chai, C. (2019). Principle of Emulsion PCR and Its Applications in Biotechnology. *Journal of Animal Reproduction and Biotechnology*, *34*(4), 259–266. https://doi.org/10.12750/jarb.34.4.259

Charlesworth B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Current biology: CB*, *6*(2), 149–162. https://doi.org/10.1016/s0960-9822(02)00448-7

Chiaroni, J., Underhill, P. A., & Cavalli-Sforza, L. L. (2009). Y chromosome diversity, human expansion, drift, and cultural evolution. *Proceedings of the National Academy of Sciences of the United States of America, 106(48), 20174–20179.* https://doi.org/10.1073/pnas.0910803106

Chiniquy, J., Garber, M. E., Mukhopadhyay, A., & Hillson, N. J. (2020). Fluorescent amplification for next generation sequencing (FA-NGS) library preparation. *BMC genomics*, *21*(1), 85. https://doi.org/10.1186/s12864-020-6481-8

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. https://doi.org/10.1007/s00439-003-1031-4

Claerhout, S., Verstraete, P., Warnez, L., Vanpaemel, S., Larmuseau, M., & Decorte, R. (2021). CSYseq: The first Y-chromosome sequencing tool typing a large number of Y-SNPs and Y-STRs to unravel worldwide human population genetics. *PLoS genetics*, *17*(9), e1009758. https://doi.org/10.1371/journal.pgen.1009758

Clark, D. P., & Pazdernik, N. J. (2016). Aging and Apoptosis. *Biotechnology, 627–661.* doi:10.1016/b978-0-12-385015-7.00020-x

Colaco, S., & Modi, D. (2018). Genetics of the human Y chromosome and its association with male infertility. *Reproductive biology and endocrinology: RB&E*, *16*(1), 14. https://doi.org/10.1186/s12958-018-0330-5

Coleman, W. B., & Tsongalis, G. J. (2017). Laboratory Approaches in Molecular Pathology—The Polymerase Chain Reaction. *Diagnostic Molecular Pathology, 15–23.* doi:10.1016/b978-0-12-800886-7.00002-9

Collen, E. J., Johar, A. S., Teixeira, J. C., & Llamas, B. (2022). The immunogenetic impact of European colonization in the Americas. In *Frontiers in Genetics* (Vol. 13). Frontiers Media S.A. https://doi.org/10.3389/fgene.2022.918227

Collins, F. S., Brooks, L. D., & Chakravarti, A. (1998). A DNA polymorphism discovery resource for research on human genetic variation. *Genome research*, *8*(12), 1229–1231. https://doi.org/10.1101/gr.8.12.1229

Colombo, G., Traverso, L., Mazzocchi, L., Grugni, V., Rambaldi Migliore, N., Capodiferro, M. R., Lombardo, G., Flores, R., Karmin, M., Rootsi, S., Ferretti, L., Olivieri, A., Torroni, A., Martiniano, R., Achilli, A., Raveane, A., & Semino, O. (2022). Overview of the Americas' First Peopling from a Patrilineal Perspective: New Evidence from the Southern Continent. *Genes*, *13*(2), 220. https://doi.org/10.3390/genes13020220

Corach, D., Lao, O., Bobillo, C., van Der Gaag, K., Zuniga, S., Vermeulen, M., van Duijn, K., Goedbloed, M., Vallone, P. M., Parson, W., de Knijff, P., & Kayser, M. (2010). Inferring continental ancestry of argentineans from Autosomal, Y-chromosomal and mitochondrial DNA. *Annals of human genetics*, *74*(1), 65–76. https://doi.org/10.1111/j.1469-1809.2009.00556.x

Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Colomb, E. B., Zaharova, B., Lavinha, J., Vona, G., Aman, R., Cali, F., Akar, N., Richards, M., Torroni, A., Novelletto, A., & Scozzari, R. (2004). Phylogeographic analysis of haplogroup E3b (E-M215) y chromosomes reveals multiple migratory events within and out of Africa. *American journal of human genetics*, *74*(5), 1014–1022. https://doi.org/10.1086/386294

Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V., Wallace, D. C., Oefner, P. J., Torroni, A., Cavalli-Sforza, L. L., Scozzari, R., & Underhill, P. A. (2002). A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *American journal of human genetics*, *70*(5), 1197–1214. https://doi.org/10.1086/340257

Cruciani, F., Trombetta, B., Antonelli, C., Pascone, R., Valesini, G., Scalzi, V., Vona, G., Melegh, B., Zagradisnik, B., Assum, G., Efremov, G. D., Sellitto, D., & Scozzari, R. (2011). Strong intra- and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152. *Forensic Science International: Genetics, 5(3), e49–e52*

Daniel James Dawson. (2018). The Devil in Cartagena: Slavery, religion and resistance in seventeenth-century Caribbean Colombia. (Thesis, James Madison University). *Available in https://commons.lib.jmu.edu/master201019/567*

Davis, L. G., & Madsen, D. B. (2020). The Coastal Migration Theory: Formulation and Testable Hypotheses. *https://www.elsevier.com/open-access/userlicense/1.0/*

De Acosta, J., & López-Morillas, F. M. (2002). *Natural and Moral History of the Indies* (J. E. Mangan, Ed.). Duke University Press. https://doi.org/10.2307/j.ctv11cw7z1

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

Del Chierico, F., Ancora, M., Marcacci, M., Cammà, C., Putignani, L., & Conti, S. (2015). Choice of next-generation sequencing pipelines. *Methods in molecular biology (Clifton, N.J.), 1231*, 31–47. https://doi.org/10.1007/978-1-4939-1720-4_3

Di Giacomo, F., Luca, F., Popa, L. O., Akar, N., Anagnou, N., Banyko, J., Brdicka, R., Barbujani, G., Papola, F., Ciavarella, G., Cucci, F., Di Stasi, L., Gavrila, L., Kerimova, M. G., Kovatchev, D., Kozlov, A. I., Loutradis, A., Mandarino, V., Mammi', C., Michalodimitrakis, E. N., … Novelletto, A. (2004). Y chromosomal haplogroup J as a

signature of the post-neolithic colonization of Europe. *Human genetics*, *115*(5), 357–371. https://doi.org/10.1007/s00439-004-1168-9

Diegoli T. M. (2015). Forensic typing of short tandem repeat markers on the X and Y chromosomes. *Forensic science international. Genetics*, *18*, 140–151. https://doi.org/10.1016/j.fsigen.2015.03.013

Dieringer, D., & Schlötterer, C. (2003). Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome research*, *13*(10), 2242–2251. https://doi.org/10.1101/gr.1416703

Dillehay, T. D. (2009). Probing deeper into first American studies. *Proceedings of the National Academy of Sciences, 106(4), 971-978.*

Dillehay, T. D., & Collins, M. B. (1988). Early cultural evidence from Monte Verde in Chile. *Nature, 332(6160), 150-152.*

Dillehay, T. D., Ocampo, C., Saavedra, J., Sawakuchi, A. O., Vega, R. M., Pino, M., Collins, M. B., Scott Cummings, L., Arregui, I., Villagran, X. S., Hartmann, G. A., Mella, M., González, A., & Dix, G. (2015). New Archaeological Evidence for an Early Human Presence at Monte Verde, Chile. *PLOS ONE, 10(11), e0141923.*

Divne, A. M., & Allen, M. (2005). A DNA microarray system for forensic SNP analysis. *Forensic science international*, *154*(2-3), 111–121. https://doi.org/10.1016/j.forsciint.2004.09.134

Elias, S. A., & Crocker, B. (2008). The Bering Land Bridge: a moisture barrier to the dispersal of steppe-tundra biota? *Quaternary Science Reviews*, *27*(27–28), 2473–2483. https://doi.org/10.1016/j.quascirev.2008.09.011

Elkamel, S., Marques, S. L., Alvarez, L., Gomes, V., Boussetta, S., Mourali-Chebil, S., Khodjet-El-Khil, H., Cherni, L., Benammar-Elgaaied, A., & Prata, M. J. (2021). Insights into the Middle Eastern paternal genetic pool in Tunisia: high prevalence of T-M70 haplogroup in an Arab population. *Scientific reports*, *11*(1), 15728. https://doi.org/10.1038/s41598-021-95144-x

Ellegren H. (2000a). Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature genetics*, *24*(4), 400–402. https://doi.org/10.1038/74249

Ellegren H. (2000b). Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in genetics: TIG*, *16*(12*), 551–558.* https://doi.org/10.1016/s0168-9525(00)02139-9

Ellegren H. (2004). Microsatellites: simple sequences with complex evolution. *Nature reviews. Genetics*, *5*(6), 435–445. https://doi.org/10.1038/nrg1348

Eren, M. I., Buchanan, B., & O'Brien, M. J. (2015). *Social learning and technological evolution during the Clovis colonization of the New World. Journal of Human Evolution, 80, 159–170.* doi: 10.1016/j.jhevol.2015.01.002

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.

Fordyce, S. L., Mogensen, H. S., Børsting, C., Lagacé, R. E., Chang, C. W., Rajagopalan, N., & Morling, N. (2015). Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™. *Forensic science international. Genetics*, *14*, 132–140. https://doi.org/10.1016/j.fsigen.2014.09.020

Fortes-Lima, C., & Verdu, P. (2021). Anthropological genetics perspectives on the transatlantic slave trade. *Human molecular genetics*, *30*(R1), R79–R87. https://doi.org/10.1093/hmg/ddaa271

Francalacci, P., Morelli, L., Underhill, P. A., Lillie, A. S., Passarino, G., Useli, A., Madeddu, R., Paoli, G., Tofanelli, S., Calò, C. M., Ghiani, M. E., Varesi, L., Memmi, M., Vona, G., Lin, A. A., Oefner, P., & Cavalli-Sforza, L. L. (2003). Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. *American journal of physical anthropology*, *121*(3), 270–279. https://doi.org/10.1002/ajpa.10265

Francis, J. Michael, "Invading Colombia: Spanish Accounts of Gonzalo Jiménez de Quesada's Expedition of Conquest." (2006). *Faculty Books*. 66. https://digitalcommons.usf.edu/books/66

Freije, D., Helms, C., Watson, M. S., & Donis-Keller, H. (1992). Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science (New York, N.Y.)*, *258*(5089), 1784–1787. https://doi.org/10.1126/science.1465614

Frudakis, T. (2010). Molecular photofitting: predicting ancestry and phenotype using DNA. *Elsevier*.

Geppert, M., Ayub, Q., Xue, Y., Santos, S., Ribeiro-dos-Santos, Â., Baeta, M., Núñez, C., Martínez-Jarreta, B., Tyler-Smith, C., & Roewer, L. (2015). Identification of new SNPs in native South American populations by resequencing the Y chromosome. *Forensic science international. Genetics*, *15*, 111–114. https://doi.org/10.1016/j.fsigen.2014.09.014

Geppert M., Baeta M., Núñez C., Martínez-Jarreta B., Zweynert S., Cruz O. W. V., et al. (2011). Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. *Forensic Science International: Genetics*, 5(2), 100–104. doi: 10.1016/j.fsigen.2010.08.016

Gibbons, A. (2014). New Sites Bring the Earliest Americans Out of the Shadows. *Science, 344(6184), 567–568.* doi:10.1126/science.344.6184.567

Goebel, T., Waters, M. R., & O'Rourke, D. H. (2008). The late Pleistocene dispersal of modern humans in the Americas. *Science (New York, N.Y.)*, *319*(5869), 1497–1502. https://doi.org/10.1126/science.1153569

Goldstein, DB & Schlotterer, C. (1999). Microsatellites: Evolution and Applications. *Oxford University Press, New York.*

Gomes, I., Pinto, N., Antão-Sousa, S., Gomes, V., Gusmão, L., & Amorim, A. (2020). Twenty Years Later: A Comprehensive Review of the X Chromosome Use in Forensic Genetics. *Frontiers in genetics*, *11*, 926. https://doi.org/10.3389/fgene.2020.00926

Gonçalves, R., Freitas, A., Branco, M., Rosa, A., Fernandes, A.T., Zhivotovsky, L.A., Underhill, P.A., Kivisild, T. and Brehm, A. (2005), Y-chromosome Lineages from Portugal, Madeira and Açores Record Elements of Sephardim and Berber Ancestry. *Annals of Human Genetics, 69: 443-454.* https://doi.org/10.1111/j.1529-8817.2005.00161.x

González, M., Gomes, V., López-Parra, A. M., Amorim, A., Carracedo, A., Sánchez-Diz, P., Arroyo-Pardo, E., & Gusmão, L. (2013). The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88. *European journal of human genetics: EJHG*, *21*(3), 324–331. https://doi.org/10.1038/ejhg.2012.167

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature reviews. Genetics*, *17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Goodwin, W., Linacre, A., & Hadi, S. (2011). An introduction to forensic genetics (Vol. 2). *John Wiley & Sons.*

Graf, K. E., & Buvit, I. (2017). Human dispersal from siberia to beringia assessing a beringian standstill in light of the archaeological evidence. *Current Anthropology*, *58*, S583–S603. https://doi.org/10.1086/693388

Graves, J. A., Wakefield, M. J., & Toder, R. (1998). The origin and evolution of the pseudoautosomal regions of human sex chromosomes. *Human molecular genetics*, *7*(13), 1991–1996. https://doi.org/10.1093/hmg/7.13.1991

Griffiths, A. D., & Tawfik, D. S. (2006). Miniaturising the laboratory in emulsion droplets. *Trends in biotechnology*, *24*(9), 395–402. https://doi.org/10.1016/j.tibtech.2006.06.009

Griffiths, A., Wessler, S., Lewontin, R. & Carroll, S. (2008). Introduction to Genetic Analysis. *New York: W. H. Freeman*

Grugni, V., Raveane, A., Ongaro, L., Battaglia, V., Trombetta, B., Colombo, G., Capodiferro, M. R., Olivieri, A., Achilli, A., Perego, U. A., Motta, J., Tribaldos, M., Woodward, S. R., Ferretti, L., Cruciani, F., Torroni, A., & Semino, O. (2019). Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. *BMC biology*, *17*(1), 3. https://doi.org/10.1186/s12915-018-0622-4

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Münsterberg, A., Vivian, N., Goodfellow, P., & Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, *346*(6281), 245–250. https://doi.org/10.1038/346245a0

Gusmão, L., Brión, M., & Gomes, I. (2008). Chapter 30 The human Y chromosome male-specific polymorphisms and forensic genetics. *Forensic Science, 969–1000.* doi:10.1016/s1567-7192(06)06034-7

Haddrill P. R. (2021). Developments in forensic DNA analysis. *Emerging topics in life sciences*, *5*(3), 381–393. https://doi.org/10.1042/ETLS20200304

Hallast, P., Kibena, L., Punab, M., Arciero, E., Rootsi, S., Grigorova, M., Flores, R., Jobling, M. A., Poolamets, O., Pomm, K., Korrovits, P., Rull, K., Xue, Y., Tyler-Smith, C., & Laan, M. (2021). A common 1.6 mb Y-chromosomal inversion predisposes to subsequent deletions and severe spermatogenic failure in humans. *eLife*, *10*, e65420. https://doi.org/10.7554/eLife.65420

Hamilton, M. J., & Buchanan, B. (2010). Archaeological support for the three-stage expansion of modern humans across northeastern Eurasia and into the Americas. *PLOS ONE, 5*(8), e12472.

Hammer, M. F., Behar, D. M., Karafet, T. M., Mendez, F. L., Hallmark, B., Erez, T., Zhivotovsky, L. A., Rosset, S., & Skorecki, K. (2009). Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. *Human genetics*, *126*(5), 707–717. https://doi.org/10.1007/s00439-009-0727-5

Hammer, M. F., Karafet, T. M., Redd, A. J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H., & Zegura, S. L. (2001). Hierarchical patterns of global human Y-chromosome diversity. *Molecular biology and evolution, 18(7), 1189–1203.* https://doi.org/10.1093/oxfordjournals.molbev.a003906

Hazzi, N. A., Valderrama-Ardila, C., Brescovit, A. D., Polotow, D., & Simó, M. (2013). New records and geographical distribution of ctenid spiders (Araneae: Ctenidae) in Colombia. *Zootaxa, 3709, 243–254.* https://doi.org/10.11646/zootaxa.3709.3.3

Head, S. R., Komori, H. K., LaMere, S. A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D. R., & Ordoukhanian, P. (2014). Library construction for next-generation sequencing: overviews and challenges. *BioTechniques*, *56*(2),https://doi.org/10.2144/000114133

Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. *Genome research*, *6*(10), 986–994. https://doi.org/10.1101/gr.6.10.986

Hodgson, J. A., & Disotell, T. R. (2010). Anthropological Genetics: Inferring the History of Our Species Through the Analysis of DNA. *Evolution: Education and Outreach*, *3*(3), 387–398. https://doi.org/10.1007/s12052-010-0262-9

Hoffecker, J. F., Elias, S. A., O'Rourke, D. H., Scott, G. R., & Bigelow, N. H. (2016). Beringia and the global dispersal of modern humans. *Evolutionary anthropology*, *25*(2), 64–78. https://doi.org/10.1002/evan.21478

Hughes, J. F., & Page, D. C. (2015). The Biology and Evolution of Mammalian Y Chromosomes. *Annual review of genetics*, *49*,507–527. https://doi.org/10.1146/annurev-genet-112414-055311

Hunter N. (2015). Meiotic Recombination: The Essence of Heredity. *Cold Spring Harbor perspectives in biology*, *7*(12), a016618. https://doi.org/10.1101/cshperspect.a016618

Ibarra, A., Freire-Aradas, A., Martínez, M., Fondevila, M., Burgos, G., Camacho, M., Ostos, H., Suarez, Z., Carracedo, A., Santos, S., & Gusmão, L. (2014). Comparison of the genetic background of different Colombian populations using the SNPforID 52plex identification panel. *International journal of legal medicine*, *128*(1), 19–25. https://doi.org/10.1007/s00414-013-0858-z

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–945. https://doi.org/10.1038/nature03001

Jannetto, P. J., Laleli-Sahin, E., & Wong, S. H. (2004). Pharmacogenomics: Methodologies for genotyping and phenotyping. *Drug Monitoring and Clinical Chemistry, 297–319.* doi:10.1016/s1567-7192(04)80013-5

Jannuzzi, J., Ribeiro, J., Alho, C., de Oliveira Lázaro e Arão, G., Cicarelli, R., Simões Dutra Corrêa, H., Ferreira, S., Fridman, C., Gomes, V., Loiola, S., da Mota, M. F., Ribeiro-dos-Santos, Â., de Souza, C. A., de Sousa Azulay, R. S., Carvalho, E. F., & Gusmão, L. (2020). Male lineages in Brazilian populations and performance of haplogroup prediction tools. *Forensic Science International: Genetics*, *44.* https://doi.org/10.1016/j.fsigen.2019.102163

Jiang, D., Klaus, S., Zhang, Y. P., Hillis, D. M., & Li, J. T. (2019). Asymmetric biotic interchange across the Bering land bridge between Eurasia and North America. *National science review*, *6*(4), 739–745. https://doi.org/10.1093/nsr/nwz035

Jobling M. A. (2012). The impact of recent events on human genetic diversity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *367*(1590), 793–799. https://doi.org/10.1098/rstb.2011.0297

Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature reviews. Genetics*, *5*(10), 739–751. https://doi.org/10.1038/nrg1455

Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nature reviews. Genetics*, *4*(8), 598–612. https://doi.org/10.1038/nrg1124

Jobling, M. A., & Tyler-Smith, C. (2017). Human Y-chromosome variation in the genome-sequencing era. *Nature reviews. Genetics*, *18*(8), 485–497. https://doi.org/10.1038/nrg.2017.36

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 M. (2017). Forensic use of Y-chromosome DNA: a general overview. *Human genetics, 136(5), 621–635.* https://doi.org/10.1007/s00439-017-1776-9

Kayser, M., Lao, O., Anslinger, K., Augustin, C., Bargel, G., Edelmann, J., Elias, S., Heinrich, M., Henke, J., Henke, L., Hohoff, C., Illing, A., Jonkisz, A., Kuzniar, P., Lebioda, A., Lessig, R., Lewicki, S., Maciejewska, A., Monies, D. M., Pawłowski, R., … Ploski, R. (2005). Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Human genetics, 117(5), 428–443.* https://doi.org/10.1007/s00439-005-1333-9

Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Krüger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., de Knijff, P., Stoneking, M., & Sajantila, A. (2000). Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son

pairs. *American journal of human genetics*, *66*(5), 1580–1588. https://doi.org/10.1086/302905

Keyeux, G., Rodas, C., Gelvez, N., & Carter, D. (2002). Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations. *Human biology*, *74*(2), 211–233. https://doi.org/10.1353/hub.2002.0022

Kim, Wonyong. (2012). Application of Metagenomic Techniques: Understanding the Unrevealed Human Microbiota and Explaining the in Clinical Infectious Diseases. *Journal of Bacteriology and Virology. 42. 263.* https://doi.org/10.4167/jbv.2012.42.4.263.

Kimura M. (1989). The neutral theory of molecular evolution and the world view of the neutralists. *Genome*, *31*(1), 24–31. https://doi.org/10.1139/g89-009

Kimura, M., & Ohta, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(6), 2868–2872. https://doi.org/10.1073/pnas.75.6.2868

King, R. & Zontini, E. (2000). The role of gender in the South European immigration model. *University of Sussex. Journal contribution.* https://hdl.handle.net/10779/uos.23446436.v1

Kivisild T. (2017). The study of human Y chromosome variation through ancient DNA. *Human genetics*, *136*(5), 529–546. https://doi.org/10.1007/s00439-017-1773-z

Kleman, J., Jansson, K., de Angelis, H., Stroeven, A., Hättestrand, C., Alm, G., & Glasser, N. (2010). North American Ice Sheet build-up during the last glacial cycle, 115-21 kyr. *Quaternary Science Reviews, 29(17–18), 2036–2051.* https://doi.org/10.1016/j.quascirev.2010.04.021

Köksal, Z., Burgos, G., Carvalho, E., Loiola, S., Parolin, M. L., Quiroz, A., Ribeiro Dos Santos, Â., Toscanini, U., Vullo, C., Børsting, C., Gusmão, L., & Pereira, V. (2022). Testing the Ion AmpliSeq™ HID Y-SNP Research Panel v1 for performance and resolution in admixed South Americans of haplogroup Q. *Forensic science international. Genetics*, *59*, 102708. https://doi.org/10.1016/j.fsigen.2022.102708

Krausz, C., and Casamonti, E. (2017). Spermatogenic Failure and the Y Chromosome. *Human Genetics 136, 637–655.* doi:10.1007/s00439-017-1793-8

Kruglyak, L. (1997). The use of a genetic map of biallelic markers in linkage studies. *Nature Genetics, 17(1), 21–24.* doi:10.1038/ng0997-21

Larmuseau, M. H. D., Otten, G. P. P. L., Decorte, R., Van Damme, P., & Moisse, M. (2017). Defining Y-SNP variation among the Flemish population (Western Europe) by full genome sequencing. *Forensic science international. Genetics*, *31*, e12–e16. https://doi.org/10.1016/j.fsigen.2017.10.008

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.*

Launching Forensic Science International Daughter Journal in 2007: Forensic Science International: *Genetics. Forensic Science International: Genetics 1:1.*

Leal Villamizar, L. M. (2022). Three Echoes of Antisemitism in Colombia from 1945 to 1948. *Contemporary Jewry*, *42*(1), 177–194. https://doi.org/10.1007/s12397-022-09423-4

Leclercq, S., Rivals, E., & Jarne, P. (2010). DNA slippage occurs at microsatellite loci without minimal threshold length in humans: a comparative genomic approach. *Genome biology and evolution*, *2*, 325–335. https://doi.org/10.1093/gbe/evq023

Liu, J., Ali, M., & Zhou, Q. (2020). Establishment and evolution of heterochromatin. *Annals of the New York Academy of Sciences*, *1476*(1), 59–77. https://doi.org/10.1111/nyas.14303

Llamas, B., Fehren-Schmitz, L., Valverde, G., Soubrier, J., Mallick, S., Rohland, N., Nordenfelt, S., Valdiosera, C., Richards, S. M., Rohrlach, A., Romero, M. I., Espinoza, I. F.Cagigao, E. T., Jiménez, L. W., Makowski, K., Reyna, I. S., Lory, J. M., Torrez, J. A., Rivera, M. A., Burger, R. L., Ceruti, M. C., Reinhard, J., Wells, R. S., Politis, G., Santoro, C. M., Standen, V. G., Smith, C., Reich, D., Ho, S. Y., Cooper, A., & Haak, W. (2016). Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances, 2(4), e1501385.*

López-Ramírez, Y. L., Aguilar-Velázquez, J. A., López-Armenta, M., RuizHernández, M., & Rangel-Villalobos, H. (2020). Paternal lineages and forensic parameters based on 23 Y-STRs (Powerplex® Y23) in Mestizo males from Mexico City. *International Journal of Legal Medicine, 134(1), 199–202.*

Lucotte, G. (2015). The major Y-chromosome haplogroup R1b-M269 in WestEurope, subdivided by the three SNPs S21/U106, S145/L21 and S28/U152, shows a clear pattern of geographic differentiation. *Advances in Anthropology, 5(01), 22.*

Manuel Fernandes Rodrigues, P. (2021). Disclosing the genetic history of male lineages present in populations from Patagonia. *(*Dissertação de Mestrado). *Disponível em Renates.(https://hdl.handle.net/10216/139409)*

Marano, L. A., & Fridman, C. (2019). DNA phenotyping: current application in forensic science. *Research and Reports in Forensic Medical Science, Volume 9, 1–8.* https://doi.org/10.2147/rrfms.s164090

Marcheco-Teruel, B., Parra, E. J., Fuentes-Smith, E., Salas, A., Buttenschøn, H. N., Demontis, D., Torres-Español, M., Marín-Padrón, L. C., Gómez-Cabezas, E. J., Alvarez-Iglesias, V., Mosquera-Miguel, A., Martínez-Fuentes, A., Carracedo, A., Børglum, A. D., & Mors, O. (2014). Cuba: exploring the history of admixture and the genetic basis of pigmentation using autosomal and uniparental markers. *PLoS genetics*, *10*(7), e1004488. https://doi.org/10.1371/journal.pgen.1004488

Martínez, B., Simão, F., Gomes, V., Nguidi, M., Amorim, A., Carvalho, E. F., Marrugo, J., & Gusmão, L. (2020). Searching for the roots of the first free African American community. *Scientific reports*, *10*(1), 20634. https://doi.org/10.1038/s41598-020-77608-8

Matisoo-Smith, E., & Ramirez, J.-M. (2009). Human Skeletal Evidence of Polynesian Presence in South America? Metric Analyses of Six Crania from Mocha Island, Chile. *Journal of Pacific Archaeology*, *1*(1), 76–88.

Mazires, S. (2011). Towards a reconciling model about the initial peopling of America. *Comptes Rendus - Biologies*, *334*(7), 497–504. https://doi.org/10.1016/j.crvi.2011.03.007

Mehta, B., Daniel, R., Phillips, C., & McNevin, D. (2017). Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. *International journal of legal medicine*, *131*(1), 21–37. https://doi.org/10.1007/s00414-016-1490-5

Mendez, F. L., Karafet, T. M., Krahn, T., Ostrer, H., Soodyall, H., & Hammer, M. F. (2011). Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. *Human biology*, *83*(1), 39–53. https://doi.org/10.3378/027.083.0103

Merriman, B., Ion Torrent R&D Team, & Rothberg, J. M. (2012). Progress in ion torrent semiconductor chip-based sequencing. *Electrophoresis*, *33*(23), 3397–3417. https://doi.org/10.1002/elps.201200424

Millstein, Roberta L. & Skipper, Robert A. (2006). Population genetics. In David L. Hull & Michael Ruse (eds.), *The Cambridge Companion to the Philosophy of Biology. Cambridge University Press.*

Mogollón Olivares, F., Moncada Madero, J., Casas-Vargas, A., Zea Montoya, S., Suárez Medellín, D., Gusmão, L., & Usaquén, W. (2020). Contrasting the ancestry patterns of three distinct population groups from the northernmost region of South America. *American journal of physical anthropology*, *173*(3), 437–447. https://doi.org/10.1002/ajpa.24130

Moreno Mayar, J. V., Potter, B. A., Vinner, L., Steinrucken, M., Rasmussen, S., Terhorst, J., Kamm, J. A., Albrechtsen, A., Malaspinas, A-S., Sikora, M., Reuther, J. D., Irish, J. D., Malhi, R. S., Orlando, L. A. A., Song, Y. S., Nielsen, R., Meltzer, D. J., & Willerslev, E. (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature Genetics*, *553*(7687), 203-207. https://doi.org/10.1038/nature25173

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology*, *51 Pt 1*, 263–273. https://doi.org/10.1101/sqb.1986.051.01.032

Myres, N. M., Rootsi, S., Lin, A. A., Järve, M., King, R. J., Kutuev, I., … Underhill, P. A. (2011). *A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. European Journal of Human Genetics, 19(1), 95–101.* doi:10.1038/ejhg.2010.146

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., & Kumlin, E. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science (New York, N.Y.)*, *235*(4796), 1616–1622. https://doi.org/10.1126/science.3029872

National Research Council (US) Committee on DNA Forensic Science: An Update. (1996). The Evaluation of Forensic DNA Evidence. *National Academies Press (US).*

Neumann, G. (1941). German Jews in Colombia: A Study in Immigrant Adjustment. *Jewish Social Studies, 3*(4), 387–398. http://www.jstor.org/stable/4615160

Neves, W. A., & Pucciarelli, H. M. (1991). Morphological affinities of the first Americans: an exploratory analysis based on early South American human remains. *Journal of Human Evolution, 21(4), 261–273.* doi:10.1016/0047-2484(91)90107-7

Neves, W. A., Hubbe, M., Okumura, M. M., González-José, R., Figuti, L., Eggers, S., & De Blasis, P. A. (2005). A new early Holocene human skeleton from Brazil: implications for the settlement of the New World. *Journal of human evolution*, *48*(4), 403–414. https://doi.org/10.1016/j.jhevol.2004.12.001

Neves, W. A., Powell, J. F., Prous, A., Ozolins, E. G., & Blum, M. (1999). Lapa Vermelha IV hominid 1: morphological affinities of the earliest known American. *Genetics and Molecular Biology, 22(4), 461-469.*

Neves, W. A., Prous, A., González-José, R., Kipnis, R., & Powell, J. (2003). Early Holocene human skeletal remains from Santana do Riacho, Brazil: implications for the settlement of the New World. *Journal of human evolution*, *45*(1), 19–42. https://doi.org/10.1016/s0047-2484(03)00081-2

Nguyen, L., & Burnett, L. (2014). Automation of molecular-based analyses: a primer on massively parallel sequencing. *The Clinical biochemist. Reviews*, *35*(3), 169–176.

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, *541*(7637), 302–310. https://doi.org/10.1038/nature21347

Noguera, M. C., Schwegler, A., Gomes, V., Briceño, I., Alvarez, L., Uricoechea, D., Amorim, A., Benavides, E., Silvera, C., Charris, M., Bernal, J. E., & Gusmão, L. (2014). Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar. *Annals of human biology*, *41*(5), 453–459. https://doi.org/10.3109/03014460.2013.852244

Nuñez, C., Baeta, M., Fernández, M., Zarrabeitia, M., Martinez-Jarreta, B., & de Pancorbo, M. M. (2015). Highly discriminatory capacity of the PowerPlex(®) Y23 System for the study of isolated populations. *Forensic science international. Genetics*, *17*, 104–107. https://doi.org/10.1016/j.fsigen.2015.04.005

Núñez, C., Geppert, M., Baeta, M., Roewer, L., & Martínez-Jarreta, B. (2012). Y chromosome haplogroup diversity in a Mestizo population of Nicaragua. *Forensic Science International: Genetics, 6(6), e192–e195.* doi: 10.1016/j.fsigen.2012.06.011

O'Brien, M. (2018). *Setting the Stage: The Late Pleistocene Colonization of North America. Quaternary, 2(1), 1.* doi:10.3390/quat2010001

O'Brien, M. J., & Buchanan, B. (2017). Cultural learning and the Clovis colonization of North America. *Evolutionary anthropology*, *26*(6), 270–284. https://doi.org/10.1002/evan.21550

Oliveira, E. J., Pádua, J. G., Zucchi, M. I., Vencovsky, R., & Vieira, M. L. C. (2006). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology, 29(2), 294–307.* doi:10.1590/s1415-47572006000200018

Ongaro, L., Scliar, M. O., Flores, R., Raveane, A., Marnetto, D., Sarno, S., Gnecchi-Ruscone, G. A., Alarcón-Riquelme, M. E., Patin, E., Wangkumhang, P., Hellenthal, G.,

Gonzalez-Santos, M., King, R. J., Kouvatsi, A., Balanovsky, O., Balanovska, E., Atramentova, L., Turdikulova, S., Mastana, S., … Montinaro, F. (2019). The Genomic Impact of European Colonization of the Americas. *Current Biology*, *29*(23), 3974-3986.e4. https://doi.org/10.1016/j.cub.2019.09.076

O'Rourke, D. H., & Raff, J. A. (2010). The human genetic history of the Americas: the final frontier. *Current biology: CB*, *20*(4), R202–R207. https://doi.org/10.1016/j.cub.2009.11.051

Ossa, H., Aquino, J., Pereira, R., Ibarra, A., Ossa, R. H., Pérez, L. A., Granda, J. D., Lattig, M. C., Groot, H., Fagundes de Carvalho, E., & Gusmão, L. (2016). Outlining the Ancestry Landscape of Colombian Admixed Populations. *PLOS ONE,11*(10), e0164414. https://doi.org/10.1371/journal.pone.0164414

Palha, T., Gusmão, L., Ribeiro-Rodrigues, E., Guerreiro, J. F., Ribeiro-Dos-Santos, A., & Santos, S. (2012). Disclosing the genetic structure of Brazil through analysis of male lineages with highly discriminating haplotypes. *PloS one*, *7*(7), e40007. https://doi.org/10.1371/journal.pone.0040007

Panneerchelvam, S., & Norazmi, M. N. (2003). Forensic DNA profiling and database. *The Malaysian journal of medical sciences: MJMS*, *10*(2), 20–26.

Pedersen, M. W., Ruter, A. H., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., Zepeda Mendoza, M. L., Beaudoin, A. B., Zutter, C., Larsen, N. K., Potter, B. A., Nielsen, R., Rainville, R. A., Orlando, L. A. A., Meltzer, D. J., Kjær, K. H., & Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, *537*(7618), 45–49. https://doi.org/10.1038/nature19085

Pereira, V., & Gusmão, L. (2016). Types of genomes, sequences and genetic markers. In A. Amorim, & B. Budowle (Eds.), *Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation* (pp. 163-191). *World Scientific. Security Science and Technology Vol. 2.* https://doi.org/10.1142/9781786340788_0009

Pericić, M., Barać Lauc, L., Martinović Klarić, I., Janićijević, B., & Rudan, P. (2005). Review of Croatian genetic heritage as revealed by mitochondrial DNA and Y chromosomal lineages. *Croatian medical journal*, *46*(4), 502–513.

Phillips, C., Amigo, J., Tillmar, A. O., Peck, M. A., de la Puente, M., Ruiz-Ramírez, J., Bittner, F., Idrizbegović, Š., Wang, Y., Parsons, T. J., & Lareu, M. V. (2020). A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel. *Forensic science international. Genetics*, *46*, 102232. https://doi.org/10.1016/j.fsigen.2020.102232

Potter, B. A., Baichtal, J. F., Beaudoin, A. B., Fehren-Schmitz, L., Haynes, C. V., Holliday, V. T., Holmes, C. E., Ives, J. W., Kelly, R. L., Llamas, B., Malhi, R. S., Miller, D. S., Reich, D., Reuther, J. D., Schiffels, S., & Surovell, T. A. (2018). Current evidence allows multiple models for the peopling of the Americas. *Science advances*, *4*(8), eaat5473. https://doi.org/10.1126/sciadv.aat5473

Potter, Ben A.; Reuther, Joshua D.; Holliday, Vance T.; Holmes, Charles E.; Miller, D. Shane; Schmuck, Nicholas (2017). Early colonization of Beringia and Northern North

America: Chronology, routes, and adaptive strategies. *Quaternary International, S104061821631307 6–*. doi: 10.1016/j.quaint.2017.02.034

Praetorius, S. K., Alder, J. R., Condron, A., Mix, A. C., Walczak, M. H., Caissie, B. E., & Erlandson, J. M. (2023). Ice and ocean constraints on early human migrations into North America along the Pacific coast. *Proceedings of the National Academy of Sciences of the United States of America, 120*(7). https://doi.org/10.1073/pnas.2208738120

Pucciarelli, H. M., Sardi, M. L., Jimenez López, J. C., & Sanchez, C. S. (2003). Early peopling and evolutionary diversification in America. *Quaternary International, 109-110, 123–132.* doi:10.1016/s1040-6182(02)00208-2

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, 12–23.* doi: 10.1016/j.fsigen.2014.04.008

Quintana-Murci, L., & Fellous, M. (2001). The Human Y Chromosome: The Biological Role of a "Functional Wasteland". *Journal of biomedicine & biotechnology, 1*(1), 18–24. https://doi.org/10.1155/S1110724301000080

Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M., Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M. C., Malaspinas, A. S., Eriksson, A., Moltke, I., Metspalu, M., Homburger, J. R., Wall, J., Cornejo, O. E., Moreno-Mayar, J. V., Korneliussen, T. S., Pierre, T., Rasmussen, M., … Willerslev, E. (2015). POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science (New York, N.Y.), 349*(6250). https://doi.org/10.1126/science.aab3884

Ralf, A., Montiel González, D., Zhong, K., & Kayser, M. (2018). Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing

Data. *Molecular biology and evolution*, *35*(5), 1291–1294. https://doi.org/10.1093/molbev/msy032

Ralf, A., van Oven, M., Montiel González, D., de Knijff, P., van der Beek, K., Wootton, S., Lagacé, R., & Kayser, M. (2019). Forensic Y-SNP analysis beyond SNaPshot: High-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing. *Forensic science international. Genetics*, *41*, 93–106. https://doi.org/10.1016/j.fsigen.2019.04.001

Ramos-Luis, E., Blanco-Verea, A., Brión, M., Van Huffel, V., Sánchez-Diz, P., & Carracedo, A. (2014). Y-chromosomal DNA analysis in French male lineages. *Forensic science international. Genetics*, *9*, 162–168. https://doi.org/10.1016/j.fsigen.2013.12.008

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., … Comas, D. (2012). 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, 21(4), 415–422.* doi:10.1038/ejhg.2012.190

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. https://doi.org/10.1371/journal.pone.0152573

Rey-González, D., Gelabert-Besada, M., Cruz, R., Brisighelli, F., Lopez-Soto, M., Rasool, M., Naseer, M. I., Sánchez-Diz, P., & Carracedo, A. (2017). Micro and macro geographical analysis of Y-chromosome lineages in South Iberia. *Forensic science international. Genetics*, *29*, e9–e15. https://doi.org/10.1016/j.fsigen.2017.04.021

Rodrigues, P., Velázquez, I. F., Ribeiro, J., Simão, F., Amorim, A., Carvalho, E. F., Bravi, C. M., Basso, N. G., Real, L. E., Galli, C., González, A. D. C., Gamulin, A., Saldutti, R., Parolin, M. L., Gomes, V., & Gusmão, L. (2022). Tierra Del Fuego: What Is Left from the Precolonial Male Lineages? *Genes*, *13*(10), 1712. https://doi.org/10.3390/genes13101712

Roewer L. (2009). Y chromosome STR typing in crime casework. *Forensic science, medicine, and pathology*, *5*(2), 77–84. https://doi.org/10.1007/s12024-009-9089-

Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M., & de Knijff, P. (1996). Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Human molecular genetics*, 5(7), 1029–1033. https://doi.org/10.1093/hmg/5.7.1029

Roewer,L. (2016). The Y-Chromosome Haplotype Reference Database (YHRD) — Publicly Available Reference and Research Datasets for the Forensic Interpretation of Y-Chromosome STR Profiles. In A. Amorim, & B. Budowle (Eds.), Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation (pp. 231-248). *World Scientific. Security Science and Technology Vol. 2.*https://doi.org/10.1142/9781786340788_0009

Rojas, W., Parra, M. V., Campo, O., Caro, M. A., Lopera, J. G., Arias, W., Duque, C., Naranjo, A., García, J., Vergara, C., Lopera, J., Hernandez, E., Valencia, A., Caicedo, Y., Cuartas, M., Gutiérrez, J., López, S., Ruiz-Linares, A., & Bedoya, G. (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. https://doi.org/10.1002/ajpa.21270

Romero, R. E., Briceño, I., Lizarazo, R.delP., Willuweit, S., Roewer, L., & Gómez, A. (2008). A Colombian Caribbean population study of 16 Y-chromosome STR loci. *Forensic science international. Genetics*, *2*(2), e5–e8. https://doi.org/10.1016/j.fsigen.2007.09.005

Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barać, L., Pericić, M., Balanovsky, O., Pshenichnov, A., Dion, D., Grobei, M., Zhivotovsky, L. A., Battaglia, V., Achilli, A., Al-Zahery, N., Parik, J., King, R., Cinnioğlu, C., … Semino, O. (2004). Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *American journal of human genetics*, *75*(1), 128–137. https://doi.org/10.1086/422196

Rootsi, S., Myres, N. M., Lin, A. A., Järve, M., King, R. J., Kutuev, I., … Underhill, P. A. (2012). Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *European Journal of Human Genetics, 20(12), 1275–1282.* doi:10.1038/ejhg.2012.86

Rothhammer, F. and Dillehay, T.D. (2009), The Late Pleistocene Colonization of South America: An Interdisciplinary Perspective. *Annals of Human Genetics, 73: 540-549.* https://doi.org/10.1111/j.1469-1809.2009.00537.x

Ryan, W. B. F., Carbotte, S. M., Coplan, J. O., O'Hara, S., Melkonian, A., Arko, R., Weissel, R. A., Ferrini, V., Goodwillie, A., Nitsche, F., Bonczkowski, J., & Zemsky, R. (2009). Global multi-resolution topography synthesis. *Geochemistry, Geophysics, Geosystems*, *10*(3). https://doi.org/10.1029/2008GC002332

Salas, A., Acosta, A., Alvarez-Iglesias, V., Cerezo, M., Phillips, C., Lareu, M. V., & Carracedo, A. (2008). The mtDNA ancestry of admixed Colombian populations. *American journal of human biology: the official journal of the Human Biology Council*, *20*(5), 584–591. https://doi.org/10.1002/ajhb.20783

Salzano F. M. (2004). Interethnic variability and admixture in Latin America--social implications. *Revista de biologia tropical*, *52*(3), 405–415. https://doi.org/10.15517/rbt.v1i2.15273

Sanchez, G., Holliday, V. T., Gaines, E. P., Arroyo-Cabrales, J., Martinez-Taguena, N., Kowler, A., … Sanchez-Morales, I. (2014). Human (Clovis)-gomphothere (Cuvieronius sp.) association 13,390 calibrated yBP in Sonora, Mexico. *Proceedings of the National Academy of Sciences, 111(30), 10972–10977.* doi:10.1073/pnas.1404546111

Sanchez, J. J., Børsting, C., Hernandez, A., Mengel-Jørgensen, J., & Morling, N. (2004). Y chromosome SNP haplogroups in Danes, Greenlanders and Somalis. *International Congress Series, 1261, 347–349.* doi:10.1016/s0531-5131(03)01635-2

Sans M. (2000). Admixture studies in Latin America: from the 20th to the 21st century. *Human biology*, *72*(1), 155–177.

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., & Malonia, S. K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. Biology, 12(7), 997. https://doi.org/10.3390/biology12070997

Scheib, C. L., Li, H., Desai, T., Link, V., Kendall, C., Dewar, G., … Kivisild, T. (2018). Ancient human parallel lineages within North America contributed to a coastal expansion. *Science, 360(6392), 1024–1027.* doi:10.1126/science. aar6851

Seielstad, M. T., Hebert, J. M., Lin, A. A., Underhill, P. A., Ibrahim, M., Vollrath, D., & Cavalli-Sforza, L. L. (1994). Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Human molecular genetics*, *3*(12), 2159–2161. https://doi.org/10.1093/hmg/3.12.2159

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P. J., Zhivotovsky, L. A., King, R., Torroni, A., Cavalli-Sforza, L. L., Underhill, P. A., & Santachiara-Benerecetti, A. S. (2004). 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. *American journal of human genetics*, *74*(5), 1023–1034. https://doi.org/10.1086/386295

Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza, L. L., & Underhill, P. A. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science (New York, N.Y.)*, *290*(5494), 1155–1159. https://doi.org/10.1126/science.290.5494.1155

Silva NM, Pereira L, Poloni ES, Currat M (2012) Human Neutral Genetic Variation and Forensic STR Data. *PLOS ONE* 7(11): e49666. https://doi.org/10.1371/journal.pone.0049666

Sinclair, A. H., Berta, P., Palmer, M. S., Hawkins, J. R., Griffiths, B. L., Smith, M. J., Foster, J. W., Frischauf, A. M., Lovell-Badge, R., & Goodfellow, P. N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, *346*(6281), 240–244. https://doi.org/10.1038/346240a0

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature,423(6942), 825–837.* doi:10.1038/nature01722

Skoglund, P., & Reich, D. (2016). A genomic view of the peopling of the Americas. *Current opinion in genetics & development*, *41*, 27–35. https://doi.org/10.1016/j.gde.2016.06.016

Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M. L., Salzano, F. M., Patterson, N., & Reich, D. (2015). Genetic evidence for two founding populations of the Americas. Nature, 525(7567), 104–108. https://doi.org/10.1038/nature14895

Skoglund, P., Mallick, S., Bortolini, M. C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M. L., Salzano, F. M., Patterson, N., & Reich, D. (2015). Genetic evidence for two founding

populations of the Americas. *Nature*, *525*(7567), 104–108. https://doi.org/10.1038/nature14895

Smith, H. L., & Goebel, T. (2018). Origins and spread of fluted-point technology in the Canadian Ice-Free Corridor and eastern Beringia. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(16), 4116–4121. https://doi.org/10.1073/pnas.1800312115

Snell, D. M., & Turner, J. M. A. (2018). Sex Chromosome Effects on Male-Female Differences in Mammals. *Current biology*, *28*(22), . https://doi.org/10.1016/j.cub.2018.09.018

Sokal R. R. (2001). Archaeogenetics: DNA and the Population Prehistory of Europe. *American Journal of Human Genetics*, *69*(1), 243–244.

Solé-Morata, N., Villaescusa, P., García-Fernández, C. *et al.* Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ. *Sci Rep* **7**, 7341 (2017). https://doi.org/10.1038/s41598-017-07710-x

Staub, Nancy. (2009). Teaching Evolutionary Mechanisms: Genetic Drift and M&M's®. *BioScience. 52. 373-377*. 10.1641/0006-3568

Sun, P., Cong, Y., Xu, M., Si, H., Zhao, D., & Wu, D. (2021). An ISFET Microarray Sensor System for Detecting the DNA Base Pairing. *Micromachines*, *12*(7), 731. https://doi.org/10.3390/mi12070731

Tamm, E., Kivisild, T., Reidla, M., Metspalu, M., Smith, D. G., Mulligan, C. J., … Malhi, R. S. (2007). *Beringian Standstill and Spread of Native American Founders. PLOS ONE, 2(9), e829.* doi: 10.1371/journal.pone.0000829

Tautz, D. (1993). Notes on the definition and nomenclature of tandemly repetitive DNA sequences. In: Pena, S.D.J., Chakraborty, R., Epplen, J.T., Jeffreys, A.J. (eds) DNA Fingerprinting: State of the Science. *Progress in Systems and Control Theory*. https://doi.org/10.1007/978-3-0348-8583-6_2

Tom D. Dillehay. (2008) Peopling of New World.Encyclopedia of Archaeology. *Academic Press.*Pages 1683-1694, https://doi.org/10.1016/B978-012373962-9.00243-0

Torroni, A., Richards, M., Macaulay, V., Forster, P., Villems, R., Norby, S., Savontaus, M. L., Huoponen, K., Scozzari, R., & Bandelt, H. J. (2000). mtDNA haplogroups and frequency patterns in Europe. *American journal of human genetics*, *66*(3), 1173–1177. https://doi.org/10.1086/302789

Toscanini, U., Gusmão, L., Berardi, G., Amorim, A., Carracedo, A., Salas, A., & Raimondi, E. (2007). Testing for genetic structure in different urban Argentinian populations. *Forensic science international*, *165*(1), 35–40. https://doi.org/10.1016/j.forsciint.2006.02.042

Toscanini, U., Gusmão, L., Berardi, G., Gomes, V., Amorim, A., Salas, A., & Raimondi, E. (2011). Male lineages in South American native groups: evidence of M19 traveling south. *American journal of physical anthropology*, *146*(2), 188–196. https://doi.org/10.1002/ajpa.21562

Underhill P. A. (2003). Inferring human history: clues from Y-chromosome haplotypes. *Cold Spring Harbor symposia on quantitative biology*, *68*, 487–493. https://doi.org/10.1101/sqb.2003.68.487

Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. Q., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L., & Oefner, P. J. (1997). Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome research, 7(10), 996–1005.* https://doi.org/10.1101/gr.7.10.996

Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J., & Cavalli-Sforza, L. L. (1996). A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(1), 196–200. https://doi.org/10.1073/pnas.93.1.196

Underhill P.A., Myres N.M., Rootsi S., Metspalu M., Zhivotovsky L.A., King R.J., Lin A.A., Chow C.E., Semino O., Battaglia V., Kutuev I., Järve M., Chaubey G., Ayub Q., Mohyuddin A., Mehdi S.Q., Sengupta S., Rogaev E.I., Khusnutdinova E.K., Pshenichnov A., Balanovsky O., Balanovska E., Jeran N., Augustin D.H., Baldovic M., Herrera R.J., Thangaraj K., Singh V., Singh L., Majumder P., Rudan P., Primorac D., Villems R., Kivisild T. (2009). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *European Journal of Human Genetics, 18(4), 479–484.* doi:10.1038/ejhg.2009.194

Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazón Lahr, M., Foley, R. A., Oefner, P. J., & Cavalli-Sforza, L. L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of human genetics*, *65*(Pt 1), 43–62. https://doi.org/10.1046/j.1469-1809.2001.6510043.x

Underhill, P., Poznik, G., Rootsi, S. *et al.* The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur J Hum Genet* 23, 124–131 (2015). https://doi.org/10.1038/ejhg.2014.50

Underhill P.A., Shen P., Lin A.A., Jin L., Passarino G., Yang W.H., Kauffman E., Bonné-Tamir B., Bertranpetit J., Francalacci P., Ibrahim M., Jenkins T., Kidd J.R., Mehdi S.Q., Seielstad M.T., Wells R.S., Piazza A., Davis R.W., Feldman M.W., Cavalli-Sforza L.L., Oefner P.J. (2000). Y chromosome sequence variation and the history of human populations. *Nature genetics*, *26*(3), 358–361. https://doi.org/10.1038/81685

Valverde, L., Illescas, M., Villaescusa, P. *et al.* New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia. *Eur J Hum Genet* 24, 437–441 (2016). https://doi.org/10.1038/ejhg.2015.114

van Oven, M., van den Tempel, N., & Kayser, M. (2012). A multiplex SNP assay for the dissection of human Y-chromosome haplogroup O representing the major paternal lineage in East and Southeast Asia. *Journal of human genetics*, *57*(1), 65–69. https://doi.org/10.1038/jhg.2011.120

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. https://doi.org/10.1002/humu.22468

Villaescusa, P., Seidel, M., Nothnagel, M., Pinotti, T., González-Andrade, F., Alvarez-Gila, O., M de Pancorbo, M., & Roewer, L. (2020). *A Y-chromosomal survey of Ecuador's multi-ethnic population reveals new insights into the tri-partite population structure and supports an early Holocene age of the rare Native American founder lineage C3-MPB373. Forensic Science International: Genetics, 102427.* doi: 10.1016/j.fsigen.2020.102427

Vullo, C., Gomes, V., Romanini, C., Oliveira, A. M., Rocabado, O., Aquino, J., Amorim, A., & Gusmão, L. (2015). Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations. *International Journal of Legal Medicine*, *129*(4), 673–680. https://doi.org/10.1007/s00414-014-1025-x

Wallace D. C. (2007). Why do we still have a maternally inherited mitochondrial DNA? Insights from evolutionary medicine. *Annual review of biochemistry, 76, 781–821.* https://doi.org/10.1146/annurev.biochem.76.081205.150955

Wang, D. G. (1998). Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. *Science, 280(5366), 1077–1082.* doi:10.1126/science.280.5366.1077

Wang, Q., Fu, L., Zhang, X., Dai, X., Bai, M., Fu, G., Cong, B., & Li, S. (2016). Expansion of a SNaPshot assay to a 55-SNP multiplex: Assay enhancements, validation, and power in forensic science. *Electrophoresis*, *37*(10), 1310–1317. https://doi.org/10.1002/elps.201500353

Wang, S., Ray, N., Rojas, W., Parra, M.V., Bedoya, G., Gallo, C., Poletti, G., Mazzotti, G., Hill, K., Hurtado, A.M., Camrena, B., Nicolini, H., Klitz, W., Barrantes, R., Molina, J.A., Freimer, N.B., Bortolini, M.C., Salzano, F.M., Petzl-Erler, M.L., Tsuneto, L.T., Dipierri, J.E., Alfaro, E.L., Bailliet, G., Bianchi, N.O., Llop, E., Rothhammer, F., Excoffier, L. & Ruiz-Linares, A. (2008). Geographic patterns of genome admixture in Latin American Mestizos. *PLOS genetics*, *4*(3), e1000037. https://doi.org/10.1371/journal.pgen.1000037

Waters M. R. (2019). Late Pleistocene exploration and settlement of the Americas by modern humans. *Science (New York, N.Y.)*, *365*(6449), eaat5447. https://doi.org/10.1126/science.aat5447

Waters, M. R., Stafford, T. W., Jr, & Carlson, D. L. (2020). The age of Clovis-13,050 to 12,750 cal yr B.P. *Science advances*, *6*(43), eaaz0455. https://doi.org/10.1126/sciadv.aaz0455

Weber, J. L., & Broman, K. W. (2001). Genotyping for human whole-genome scans: past, present, and future. *Advances in genetics*, *42*, 77–96. https://doi.org/10.1016/s0065-2660(01)42016-5

Wei, W., Ayub, Q., Chen, Y., McCarthy, S., Hou, Y., Carbone, I., Xue, Y., & Tyler-Smith, C. (2013). A calibrated human Y-chromosomal phylogeny based on resequencing. *Genome research*, *23*(2), 388–395. https://doi.org/10.1101/gr.143198.112

Wheat, D. (2011). The first great waves: African provenance zones for the Transatlantic slave trade to Cartagena de Indias, 1570–1640. *The Journal of African History, 52(1), 1-22.* doi:10.1017/S0021853711000119

Wilhelm, J., & Pingoud, A. (2003). Real-time polymerase chain reaction. *Chembiochem : a European journal of chemical biology, 4(11), 1120–1128.* https://doi.org/10.1002/cbic.200300662

Willuweit, S., Roewer, L., & International Forensic Y Chromosome User Group (2007). Y chromosome haplotype reference database (YHRD): update. *Forensic science international. Genetics*, *1*(2), 83–87. https://doi.org/10.1016/j.fsigen.2007.01.0

Wright, A. E., Dean, R., Zimmer, F., & Mank, J. E. (2016). How to make a sex chromosome. *Nature communications*, *7*, 12087. https://doi.org/10.1038/ncomms12087

Wyman, A. R., & White, R. (1980). A highly polymorphic locus in human DNA. Proceedings of the National Academy of Sciences of the United States of America, 77(11), 6754–6758. https://doi.org/10.1073/pnas.77.11.6754

Xavier, C., et al. (2015). "Admixture and genetic diversity distribution patterns of nonrecombining lineages of Native American ancestry in Colombian populations." PLOS ONE 10(3): e0120155

Xu, Y., & Pang, Q. (2022). Repetitive DNA Sequences in the Human Y Chromosome and Male Infertility. In Frontiers in Cell and Developmental Biology (Vol. 10). Frontiers Media S.A. https://doi.org/10.3389/fcell.2022.831338

Xue, Y., Wang, Q., Long, Q., Ng, B. L., Swerdlow, H., Burton, J., Skuce, C., Taylor, R., Abdellah, Z., Zhao, Y., Asan, MacArthur, D. G., Quail, M. A., Carter, N. P., Yang, H., & Tyler-Smith, C. (2009). Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current biology: CB*, *19*(17), 1453–1457. https://doi.org/10.1016/j.cub.2009.07.032

Y Chromosome Consortium (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome research*, *12*(2), 339–348. https://doi.org/10.1101/gr.217602

Zalloua, P. A., Xue, Y., Khalife, J., Makhoul, N., Debiane, L., Platt, D. E., … Tyler-Smith, C. (2008). Y-Chromosomal Diversity in Lebanon Is Structured by Recent Historical Events. *The American Journal of Human Genetics, 82(4), 873–882.* doi: 10.1016/j.ajhg.2008.01.020

Zanoli, L. M., & Spoto, G. (2012). Isothermal amplification methods for the detection of nucleic acids in microfluidic devices. *Biosensors*, *3*(1), 18–43. https://doi.org/10.3390/bios3010018

Zegura, S. L., Karafet, T. M., Zhivotovsky, L. A., & Hammer, M. F. (2004). 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), 164–175. https://doi.org/10.1093/molbev/msh009

Zeng, Y., Novak, R., Shuga, J., Smith, M. T., & Mathies, R. A. (2010). High-performance single cell genetic analysis using microfluidic emulsion generator arrays. *Analytical chemistry*, *82*(8), 3183–3190. https://doi.org/10.1021/ac902683t

Zieger, M., & Utz, S. (2020). The Y-chromosomal haplotype and haplogroup distribution of modern Switzerland still reflects the alpine divide as a geographical barrier for human migration. *Forensic science international. Genetics*, *48*, 102345. https://doi.org/10.1016/j.fsigen.2020.102345

Appendix

List of Appendix Tables

Table S1: List of the 859 Y-SNPs within the kit used for the genotyping of the samples: marker name, along with the position within the chromosome and the assigned haplogroup specific to each marker.

Marker name	Hg19 position	Haplogroup
FGC26250	14301426	A00
L1086	2826312	A00
FGC25669	8210501	A00
L1144	21717306	A00
AF7	6931719	A00
FGC26856	17375117	A00
L1159	23021666	A00
L1146	21739660	A00
L987	7583448	A0a
V164	18255814	A0a1
L1080	21893300	A0a1
V159	6932829	A0a1
L1075	14125645	A0a1
L994	14902417	A0a2
L1007	17419925	A0a2
L981	6894187	A0a2
L1054	18887936	A0b
L1056	21930259	A0b
L1058	23973630	A0b
L990	14001289	A0-T
L896	21706626	A0-T
L1089	2887280	A0-T
L1085	2790726	A0-T
L1010	21043446	A0-T
V148	6788191	A0-T
L1125	15467768	A0-T
L1121	14496448	A0-T
L1120	14496439	A0-T
L1130	16661010	A0-T
L1118	14231291	A0-T
L1145	21739790	A0-T
V4	6846482	A1a
M31	21739754	A1a
L419	15204887	A1b1
V50	6845936	A1b1a
M14	21722268	A1b1a1
P36	14496441	A1b1a1
M276	22741728	A1b1a1
M277	22741821	A1b1a1

M49	21868726	A1b1a1a
M196	15030682	A1b1a1a
M6	18571026	A1b1a1a
P28	21881115	A1b1a1a1
L963	17661248	A1b1a1a2
M212	15526695	A1b1a1a2a
P262	6932148	A1b1a1a2a1a
L951	6932648	A1b1a1a2b~
M32	21740436	A1b1b
M28	21729839	A1b1b1
M144	21925500	A1b1b2
M190	14968527	A1b1b2
M220	15437050	A1b1b2
P289	8467082	A1b1b2
M51	21868863	A1b1b2a
L439	22157292	A1b1b2b~
L437	19092633	A1b1b2b~
V10	6893093	A1b1b2b~
M118	21763965	A1b1b2b1
L1053	18759708	A1b-T
V221	7589303	A1b-T
V168	17947672	A1-T
P305	2710154	A1-T
V171	4898665	A1-T
M181	14851554	B
M182	14869076	B2
V197	23120337	B2~
M150	21869519	B2a
V78	6859558	B2a1~
V75	17421349	B2a1a
M218	15437063	B2a1a1
V227	7599027	B2a1a1a~
M109	21774180	B2a1a1a1
M152	21868068	B2a1a1a1
M5838	6932674	B2a1a1a1~
Y22755	22741795	B2a1a1a1b~
Y21895	16890020	B2a1a1a1b~
G1	21932926	B2a1a1a2
M112	21767959	B2b
M192	15014262	B2b1
50f2(P)	23497067	B2b1
M8495	2736732	B2b1a~
M8497	8156900	B2b1a~
M8451	18865395	B2b1a1~

M8437	17657626	B2b1a1~
M8351	6790204	B2b1a1~
M8342	14587628	B2b1a1~
M8348	6637124	B2b1a1~
M30	21739815	B2b1a1b
M8036	7769194	B2b1a1c~
M8035	7606120	B2b1a1c~
M7715	21940347	B2b1a2~
M7618	14001964	B2b1a2~
Z38032	14001914	B2b1a2a~
Z38052	14484533	B2b1a2a~
Z37977	6932169	B2b1a2a~
M7106	6982169	B2b1b~
M7104	2711408	B2b1b~
P6	6768265	B2b1b1a
M6529	2657411	B2b1b1a
L1394	19413382	B3
L1391	14197960	B3
Z5704	21749950	B3~
Z5261	7962998	B3~
L413	6932831	BT
M42	21866840	BT
M9379	23244049	BT
M216	15437564	C
CTS2955	14587658	C
M130	2734854	C
K29	6630453	C1
F3393	23023974	C1
CTS11043	22914979	C1a
M8	7291534	C1a1
CTS7097	17287913	C1a1~
M105	21866491	C1a1~
Y11438	17880509	C1a2
V20	6845955	C1a2
Y11493	22741880	C1a2a
V3918	19219752	C1a2a1
Y10522	22187260	C1a2a2
Y10504	15576131	C1a2a2
Z44512	21645360	C1a2b
M356	2888203	C1b1a1
P92	14850853	C1b1a1a
K362	21527071	C1b1a1a1a1a
K193	14144911	C1b1a1a1a1a1a
Z31729	7606091	C1b1a1a1a1b~

Z5899	2852640	C1b1a1b
Z32648	17661323	C1b1a2a
AM00868	6932146	C1b1a2b
M38	21742158	C1b2a
M208	15576203	C1b2a1
P33_1	24987987	C1b2a1a~
Z32247	16520490	C1b2a1d1
M347	2877479	C1b2b
M210	15575780	C1b2b1
P44	14495251	C2
M217	15437333	C2
Z1453	19020366	C2
L1373	22711592	C2b
P39	14484581	C2b1a1a
M48	21749881	C2b1a2
Z32955	15030739	C2b1a2b1
F1144	8307194	C2c1
Z1338	17844001	C2c1
Z1300	7801788	C2c1a
Z31672	2815306	C2c1a2b2a~
F845	6875257	C2c1b
Z12301^	21881111	C2c1b1a
CTS8866^	18248829	C2c2
P143	14197867	CF
M294	22744945	CT
M168	14813991	CT
JST021355	2828425	D
CTS94	2726402	D
M174	14954280	D
N1	14851526	D1a1a1
Z27287	9170598	D1a1a1a
Z40608	2657404	D1a1a2b~
P47	14497168	D1a2a
PH937	14140293	D1a2a1~
Y15320	23550955	D1a2a1~
PH3836	18570919	D1a2a1~
Y15414	21827747	D1a2a1~
PH3518	17922027	D1a2a1~
M55	21872738	D1b
M179	14838700	D1b
M116	21765281	D1b1
L1375	7659530	D1b1
Z1627	18942397	D1b1a
M125	21930287	D1b1a

P42	14481731	D1b1a1~
JST022457	24464597	D1b1a2
JST006841	6736154	D1b1a2b
Z43424	6875167	D1b1a2b1a1b~
Z18561	17508034	D1b1a3
Z18552	14481729	D1b1a3
Z18518	6893070	D1b1a3
P120	15426000	D1b1b~
Z1576	17547333	D1b1c1
Z1569	6804892	D1b1c1
Z1520	14181878	D1b2
Z1516	6863631	D1b2
FGC9202	17917449	D2
FGC9230	18888218	D2
M203	15591537	DE
M145	21717208	DE
M40	2663943	E
M96	21778998	E
M5382	6631743	E
P29	14497207	E
P147	21083420	E1
CTS2031	14159750	E1a
M132	21896261	E1a
M33	21740450	E1a
L632	15018495	E1a1
M44	21752644	E1a1
Z15405	23244027	E1a2a1a1a
Z36345	22187198	E1a2b1a1a1a~
P177	14159846	E1b
P179	14060308	E1b1
P2	21610831	E1b1
CTS1978	14140277	E1b1~
V38	6818291	E1b1a
P189	14197977	E1b1a1
P1	21612167	E1b1a1
M2	14096577	E1b1a1~
M180	14850341	E1b1a1a1
L485	8852858	E1b1a1a1a1c
CTS5038	15978855	E1b1a1a1a1c1~
M191	15019613	E1b1a1a1a1c1a1
U174	16251357	E1b1a1a1a1c1a1a
Z37817	6790169	E1b1a1a1a1c1a1a3a1d2~
CTS490	6790237	E1b1a1a1a1c1a1a3c3~
L372	14869766	E1b1a1a1a1c1a1a5~

Z36721	2826363	E1b1a1a1a1c1b2a1a1~
Z6023	16374450	E1b1a1a1a1c1b2a3
Z22265	14181158	E1b1a1a1a1c1b2a3a
Z36573	8467196	E1b1a1a1a1c2b2a2~
U175	16253694	E1b1a1a1a2a
U209	17294958	E1b1a1a1a2a1~
Z37284	8736303	E1b1a1a1a2a1a3a2a1b~
U290	21646058	E1b1a1a1a2a1a3b1a
U181	16374426	E1b1a1a1a2a1a3b1a1
Z36959	6931982	E1b1a1a1a2a1a3b1a2b~
Z5994	6766753	E1b1a1b
V1150	2888212	E1b1a2a1~
L336	21903853	E1b1b
M215	15467824	E1b1b
M35	21741703	E1b1b1
CTS8131	17846754	E1b1b1
CTS11205	23009991	E1b1b1~
L539	23973415	E1b1b1a
L546	17516070	E1b1b1a
M78	21893303	E1b1b1a1
Z1902	22823155	E1b1b1a1a
V12	6823099	E1b1b1a1a1
M224	21893299	E1b1b1a1a1a~
V32	6932821	E1b1b1a1a1b
Y18637	14301470	E1b1b1a1a1b1a2a1~
BY8081	14587694	E1b1b1a1a1b1a2a1a~
BY8427	16638765	E1b1b1a1a1c1b2a~
BY8199	6931895	E1b1b1a1a1d~
V65	17287868	E1b1b1a1a2
Z1919	15499572	E1b1b1a1b
Z1920	15502485	E1b1b1a1b
L618	15339697	E1b1b1a1b1
V36	6814246	E1b1b1a1b1a
V13	6842263	E1b1b1a1b1a
Z36786	7690204	E1b1b1a1b1a1~
Z42903	14477437	E1b1b1a1b1a10a1b~
V22	6859957	E1b1b1a1b2
AM00903	21866497	E1b1b1a1b2
Z36763	6637064	E1b1b1a1b2a4a~
Y24403	2745572	E1b1b1a1b2a4c1a1a1a1~
M521	6822948	E1b1b1a1c~
V2641	14850904	E1b1b1a2~
Y21100	22191276	E1b1b1a2~
Z827	7290454	E1b1b1b

M310	22751863	E1b1b1b1
M81	21892572	E1b1b1b1a
M165	21866675	E1b1b1b1a1a~
BY9700	22711529	E1b1b1b1a1c10~
Z830	7729622	E1b1b1b2
M123	21764586	E1b1b1b2a1
M34	21740716	E1b1b1b2a1a~
FGC18329	7508107	E1b1b1b2a1a1~
Y16613	21925447	E1b1b1b2a1a1a1a1b1~
Y19892	21749887	E1b1b1b2a1a1a1a1e1b1a~
M290	21878825	E1b1b1b2a1a2
P72	21610853	E1b1b1b2b1~
M293	22744939	E1b1b1b2b2a1
Y19232	21925521	E1b1b1b2b3b~
V6	6932007	E1b1b1c
V16	6851471	E1b1b2
M281	21764500	E1b1b2
M75	21890177	E2
CTS2954	14587625	E2
M237	14923438	E2a~
CTS10698	22711493	E2b
M85	21896216	E2b1a
CTS5716	16453076	E2b1a1c
M213	15526751	F
P14	17398598	F
M89	21917313	F
M427	19092597	F2
M428	19092559	F2
M481	14496387	F3
M201	15027529	G
P257	14432928	G
L833	15541217	G1
M342	23243942	G1
M285	22741740	G1
GG182	14060297	G1a1b1
GG212	22917963	G1a1b1a
L156	17174741	G2
P287	22072097	G2
F2930	14108344	G2a
L31	14028148	G2a
P15	23244026	G2a
U5	23973594	G2a
Z35438	22823184	G2a1a3a~
PF2824	5253861	G2a2

L1259	15615340	G2a2
M286	22741799	G2a2a1a1a
L91	21645555	G2a2a1a2
Z6226^	15591548	G2a2a1a2a1b
Z6191	7290503	G2a2a1a2a1b2c~
M485	18759690	G2a2b
L30	15604899	G2a2b
U8	14692227	G2a2b
M406	2749995	G2a2b1
U16	22917995	G2a2b1a1a
L90	21628300	G2a2b1a1a
P303	21645348	G2a2b2a
Z764	18023391	G2a2b2a1
U1	15058878	G2a2b2a1a1a
M527	14871976	G2a2b2a1a1a1
U13	14698928	G2a2b2a1a1a1
Z38862	6870773	G2a2b2a1a1a1a1a1a1a~
Z738	14116041	G2a2b2a1a1b
L497	17423320	G2a2b2a1a1b
CTS5762	16471001	G2a2b2a1a1b1a~
Z41201	8307134	G2a2b2a1a1b1a1a1a1b2~
CTS4803	15833180	G2a2b2a1a1b1a1a2a
Z31323	8614265	G2a2b2a1a1b1a1a2a8~
Z31321	2887913	G2a2b2a1a1b1a1a2a8~
Y14681	14871922	G2a2b2a1a1b1a1a2b1a
Z39671	6822961	G2a2b2a1a1b1a1a2b2a
Z41155	21843875	G2a2b2a1a1b1a1a2b2a1a
CTS342	6700171	G2a2b2a1a1c
CTS8476	18043703	G2a2b2a1a1c1
Z1903	16596946	G2a2b2a1a1c1a
Z2047	14281832	G2a2b2a1a1c1a
Z43085	6740052	G2a2b2a1a1c1a1b3
FGC37729	17922058	G2a2b2a1a1c3a
V8.2	6932176	G2a2b2a1a1c3c~
PF3359	7877472	G2a2b2b
L72	23243887	G2b1
M377	15027433	G2b1
L183	8467136	G2b1
F1329	8589031	GHIJK
L901	17844304	H
M3035	21940316	H~
L902	23009978	H1
M69	21894058	H1a
M52	21753199	H1a1

M2970	18870223	H1a1a
M97	21778713	H1a1a2~
M2940	17846704	H1a1a4b
M2853	15748250	H1a1a4b2
M2972	18881280	H1a1a4b2
Z14475	14198010	H1a1a4b2b1
PR2906	18865398	H1a1a4b3
Z4507	6680599	H1a1a4b3
Z4452	17508064	H1a1b1
Z40929	19092576	H1a1b1a1a~
Y2573	3048741	H1a2
Z194	7743644	H1a2a
Apt	23496331	H1a2a
P80	6739899	H1a2a1~
Z14258	2697625	H1a2b
Z14338	14181842	H1a2b1a
Z14348	16227318	H1a2b1a
Z34749	7897013	H1a2b1b
Z34837	23243909	H1a2b1b
Z34737	6932258	H1a2b1b~
L284	15604862	H2
L286	21890111	H2
M282	21764431	H2
P96	14869743	H2
Z19055	17421318	H2a1a~
Z5857	2759285	H3
Z13595	21778710	H3
Z9469	18622326	H3a
Z13161	21733436	H3a1
Z12794	22191274	H3a2
Z5865	2667783	H3a2a
Z12831	6766724	H3a2b
Z34922	2887863	H3a2b1~
Z13721	14197992	H3b1
Z41256	21890023	H3b1b~
M578	7202703	HIJK
M170	14847792	I
M258	15023364	I
U179	16354708	I
L81	22513726	I1
L80	14640715	I1
L125	18759669	I1
L124	17766762	I1
M253	15022707	I1

L840	20834703	I1
DF29	3626279	I1a
CTS6364	16850779	I1a1
A5634	6823015	I1a10b2a~
M227	15591446	I1a1a1a1a
S4774	14477408	I1a1a1a1b1~
Y19346	2726356	I1a1a3a1a~
L22	8576009	I1a1b1
P109	15426005	I1a1b1a1
Z74	8073670	I1a1b1a4a
L813	7659777	I1a1b1a4a2
Z58	9847423	I1a2
Z59	16802360	I1a2a
Z60	18234588	I1a2a1a1~
Z140	17863355	I1a2a1a1a
Z141	7530813	I1a2a1a1a
L338	19092598	I1a2a1a1a1a1
Z44256	6766347	I1a2a1a1a1a2b1~
Y3648	15574107	I1a2a1a1a2a2a1~
A11346	15023195	I1a2a1a1a3a2b~
F13007	21739625	I1a2a1a1a3a2b~
A1616	14871952	I1a2a1a1a6~
PH4989	21868591	I1a2a1a1c1a1~
Z2539	17443720	I1a2a1a1d
Y32130	2658864	I1a2a1a1d1b2~
S18934	16890002	I1a2a1a1d1b2~
Z2041	17571016	I1a2a2
Z2040	15642141	I1a2a2a
FGC12758	18842801	I1a2a2a1a1a~
Y21492	15526759	I1a2a2a4b1~
Z139	21288655	I1a2b
Z138	15711340	I1a2b
Z63	14401486	I1a3
L1237	21742084	I1a3a1a1
Y17387	16596926	I1a3g2~
Z131	5845252	I1b
L68	18700150	I2
M438	16638804	I2
L460	7879415	I2a
CTS595	6874115	I2a1a
M26	21865821	I2a1a1
L1286	21778662	I2a1a2a
L1287	21970862	I2a1a2a1
L1294	2887401	I2a1a2b

M423	19096091	I2a1b
L161	22513718	I2a1b1
L621	18760081	I2a1b2
M436	18747493	I2a2
P217	7628484	I2a2
L37	17516123	I2a2
L700	2750000	I2a2~
M223	21717307	I2a2a
U250	18888200	I2a2a
CTS616	6906332	I2a2a1
CTS9183	18732197	I2a2a1
L369	14850314	I2a2a1a1a1
L1229	14937828	I2a2a1a2a
CTS10100	19255890	I2a2a1b
CTS10057	19232160	I2a2a1b
L702	7629205	I2a2a1b1
P78	6740387	I2a2a1b1a
Z161	2696497	I2a2a1b2
L801	21763755	I2a2a1b2a
CTS1977	14140273	I2a2a1b2a1a
CTS6433	16889964	I2a2a1b2a1b
Z78	9154908	I2a2a1b2a1b1a1
L1198	18405788	I2a2a1b2a1b1a1a1a1
CTS661	6931960	I2a2a1b2a1b1b1
L38	15668070	I2a2b1
L533	2887198	I2a2b1a1
L597	18887888	I2c
M429	14031334	IJ
P126	21225770	IJ
M522	7173143	IJK
M304	22749853	J
L134	14850935	J
P209	19179335	J
M497	17419934	J1
L827	15541333	J1
M267	22741818	J1
Z2215	16227302	J1a
Z2356	14072309	J1a2a~
L620	5274763	J1a2a~
L826	7877119	J1a2a~
PF7264	7842079	J1a2a1a1~
P58	14486667	J1a2a1a2
L817	6932171	J1a2a1a2c1
L818	19136821	J1a2a1a2c1a

Z2329	22957408	J1a2a1a2d2b2b2c~
Z1884	22797639	J1a2a1a2d2b2b2c4~
L858	9879005	J1a2a1a2d2b2b2c4~
YSC76	8432091	J1a2a1a2d2b2b2c4b~
Z640	8210480	J1a2a1a2d2b2b2c4c~
Z641	9021198	J1a2a1a2d2b2b2c4c~
Z1828	2669716	J1a2b
Z1834	8614242	J1a2b~
M172	14969634	J2
L228	7771358	J2
L559	21674327	J2a
M410	2751678	J2a
L26	22942897	J2a1
M322	15469740	J2a1a
M67	21878809	J2a1b
M92	21904023	J2a1b1
M166	21764694	J2a1b2
L210	16492197	J2a1b3
L227	6859963	J2a1b3
PF5132	17434361	J2a1b-PF5132
Z500	8529549	J2a1b-Z500
M68	21878700	J2a1c
M319	15467785	J2a1d
L24	14286528	J2a1h
M158	21716366	J2a1h1
L25	19136822	J2a1h2
L398	23243897	J2a1h2a1
M137	21764458	J2a1h2a1a
Z7700	6870824	J2a1h2-Z7700
PF5116	4065584	J2a1-PF5116
PF5125	9089648	J2a1-PF5125
PF5172	7882829	J2a1-PF5172
PF5294	14727442	J2a1-PF5294
Z1846	7372423	J2a1-Z1846
Z2220	2687198	J2a1-Z2220
Z2221	8265737	J2a1-Z2221
PF5197	19068737	J2a1-Z5197
Z6065	18182306	J2a1-Z6065
PF5008	7691561	J2a2
M340	21878699	J2a2a1
L282	5359116	J2b
M12	7583480	J2b
M205	15587509	J2b1
M241	15018459	J2b2a

L283	6932663	J2b2a1
Z2432	7508082	J2b2a2~
PF5050	22007097	J2-PF5050
M9	21730257	K
P131	15472863	K~
M526	23550924	K2
P331	16222561	K2b
P399	13563083	K2b1
P397	28759645	K2b1
M20	21733454	L
M2527	16256048	L~
M22	14790163	L1
M76	21757316	L1a1
M27	21739646	L1a1
Z5926	6766411	L1a1b
M357	2888252	L1a2
L1307^^	23496396	L1a2
Z5921^	6817129	L1a2a1a
Page116	21893864	L1b1a
L595	21739651	L2
L868	15604917	L2~
L811	21364708	LT
P326	8467290	LT
P256	8685231	M
M4	2744628	M1
M106	21866424	M1
M5	21609946	M1
SK1854	17844264	M1~
M104_1	8619592	M1a2a1
M104_2	8620324	M1a2a1
M353	17309267	M2
M177	2658869	M2a
P117	14819693	M3
M231	15469724	N
F1206	8440417	N1a
L729	19431608	N1a~
P105	15422962	N1a1
M46	14922583	N1a1
M178	21741755	N1a1a
L839	22974783	N1a1a1a
L708	7629512	N1a1a1a
VL29	14570424	N1a1a1a1a1a
L550	21286248	N1a1a1a1a1a1a
L1025	14126006	N1a1a1a1a1a1a1

Z1936	21463326	N1a1a1a1a2
Z1935	19219721	N1a1a1a1a2a1a
Z1941	8557714	N1a1a1a1a2a1a1a1
L1034	21646398	N1a1a1a1a2a1c1~
L666	7570816	N1a2
Z35091	6932855	N1a2b1b1
VL63	21843827	N1a2b2a2a~
F4201	16256046	N1b1
L732	18146985	N1b1a2b~
FGC28459	16310688	N2~
M2308	7690182	NO
F650	23617006	NO
CTS11536	23154209	NO1
M214	15471925	NO1
P186	7568568	O
F75	7410009	O1
M119	21762685	O1a
M50	21868672	O1a2
M110	21907097	O1a2
F167	8796086	O1b
M268	22739301	O1b
F1462	8910408	O1b1a1
PK4	21285962	O1b1a1
M95	21938444	O1b1a1a
M88	21900849	O1b1a1a1a1a
F1956	15472846	O1b1a2b
P49	14496407	O1b2
M176	2655180	O1b2
F3356	22752710	O1b2a1a
F3390	23021633	O1b2a1a2b
M122	21764674	O2
P200	18560005	O2a
M324	2821786	O2a
P197	19179463	O2a
F573	21286254	O2a1
KL1	18842841	O2a1
F1876	14968529	O2a1a
F964	7421428	O2a1a
M164	21757306	O2a1b~
JST002611	7546726	O2a1c
F11	2815303	O2a1c1a1a1a
F38	6740428	O2a1c1a1a1a1b
F449	17917402	O2a1c1a2
P201	2828196	O2a2

M188	14924869	O2a2a
M7	4078217	O2a2a1a2
M209	15575790	O2a2a1a2
N5	14851538	O2a2a1a2a1a2
P164	14001232	O2a2b
Page23	15999244	O2a2b1a1
Y21364	8467167	O2a2b1a1a1a1a
F869	6932135	O2a2b1a1a3a
F444	17880466	O2a2b1a2
CTS11459	23120297	O2a2b1a2a1
F3607	24358567	O2a2b1a2a1a3a
N6	21889536	O2a2b2
AM01856	6932191	O2a2b2a
P295	7963031	P
M74	21889767	P1
M45	21867787	P1
B253	2745533	P2~
M242	15018582	Q
L232	17516095	Q
L472	7014317	Q1
F1096	8087573	Q1a
F1251	8491335	Q1a
F1215	8454150	Q1a
F3243	21723218	Q1a
NWT01	2888083	Q1a1
M120	21907394	Q1a1a
F2273	16802392	Q1a1a
Y697	19431584	Q1a1a
Y558	21872679	Q1a1a1a
M143	21889818	Q1a2
M25	21866664	Q1a2
F7020	15502461	Q1a2
BZ656	18700177	Q1a2a1b~
M346	2887156	Q1b
L57	15574102	Q1b
L56	8148869	Q1b
L53	21642296	Q1b1
L331	17661226	Q1b1
L54	23292782	Q1b1a
CTS11969	23391298	Q1b1a1
CTS7779	17657530	Q1b1a1a
CTS11970	23391307	Q1b1a1a
M3	19096363	Q1b1a1a
M19	21733231	Q1b1a1a1a~

M194	15014550	Q1b1a1a1b~
P106	15422922	Q1b1a1a1c~
SA01	15014716	Q1b1a1a1d~
L891	19413414	Q1b1a1a1e1a~
Z35615	2667728	Q1b1a1a1i2~
Z35616	7271725	Q1b1a1a1i2~
BZ3401	8467139	Q1b1a1a1v~
Y18425	7536613	Q1b1a1a2b1~
L804	14028228	Q1b1a1b
L805	16699291	Q1b1a1b
BY1771	21628682	Q1b1a1b~
Z780	8510540	Q1b1a2
L569	21729827	Q1b1a2a1a1
YP891	16889973	Q1b1a2a1a1~
L565	19096384	Q1b1a2a1a1a1~
L334	14482079	Q1b1a3
L330	17766807	Q1b1a3
Y11939	15014219	Q1b1a3b
F4674	6766774	Q1b2a
L940	6846280	Q1b2b~
L529	6931700	Q1b2b1~
L527	17661303	Q1b2b1~
FGC6891	9400913	Q1b2b1a~
FGC6925	15711324	Q1b2b1a~
L314	15587574	Q2
L275	19136888	Q2
F1205	8440399	Q2
Y2056	9170505	Q2a
M378	15027507	Q2a1
L245	5675090	Q2a1a
Y18596	2888165	Q2a1a2a1a2~
BZ314	19068749	Q2a1a3b~
L301	15018672	Q2a1c1b1
Y29483	17419910	Q2a1c2a~
YP751	16453062	Q2b
SK2003	8440389	Q2b4~
M207	15581983	R
M173	15026424	R1
M306	22750583	R1
P294	7570822	R1
M420	23473201	R1a
M448	16520444	R1a1
SRY10831.2	2657176	R1a1
M198	15030752	R1a1a

M417	8533735	R1a1a1
CTS4385	15625999	R1a1a1a~
Z647	7683058	R1a1a1b
Z645	8245045	R1a1a1b
Z283	21976303	R1a1a1b1
M458	24366464	R1a1a1b1a1
L291	24366496	R1a1a1b1a1~
L260	18248830	R1a1a1b1a1a
Z91	16474793	R1a1a1b1a2
Z92	17585001	R1a1a1b1a2a
M558	15063588	R1a1a1b1a2b
CTS1211	7271722	R1a1a1b1a2b
CTS3402	14896020	R1a1a1b1a2b3
L365	21628667	R1a1a1b1a2b3a
Z284	8717196	R1a1a1b1a3
L448	6931809	R1a1a1b1a3a
Z287	17293093	R1a1a1b1a3b
Z288	18175765	R1a1a1b1a3b
Z93	7552356	R1a1a1b2
Z94	21043448	R1a1a1b2a
Z95	23956870	R1a1a1b2a~
L657	14159864	R1a1a1b2a1a
Y8	9400938	R1a1a1b2a1a1~
Y9	19383352	R1a1a1b2a1a2~
Z2124	4521678	R1a1a1b2a2
Z2123	16453077	R1a1a1b2a2a1
Z2122	14477397	R1a1a1b2a2b
CTS6	2657349	R1a1a1b2a2b1a
M343	2887824	R1b
M415	9170545	R1b1
L780	21183643	R1b1
L389	28733101	R1b1a1
P297	18656508	R1b1a1a
L502	19020340	R1b1a1a
M478	23444054	R1b1a1a1a
L1435	28733044	R1b1a1a1a
L1432	15642157	R1b1a1a1a1
Y20761	19068733	R1b1a1a1a2
M269	22739367	R1b1a1a2
L757	19054889	R1b1a1a2
M412	8502236	R1b1a1a2a1
P310	18907236	R1b1a1a2a1a
L11	17844018	R1b1a1a2a1a
P311	18248698	R1b1a1a2a1a

L52	14641193	R1b1a1a2a1a
U106	8796078	R1b1a1a2a1a1
Z8057	15850306	R1b1a1a2a1a1a
Z18	14991735	R1b1a1a2a1a1b
L257	15668028	R1b1a1a2a1a1b1a1
Z381	7246726	R1b1a1a2a1a1c
Z156	15780341	R1b1a1a2a1a1c1
S498	19047132	R1b1a1a2a1a1c1a
Z301	14181107	R1b1a1a2a1a1c2
U198	16839499	R1b1a1a2a1a1c2a1
L48	23612197	R1b1a1a2a1a1c2b
L47	22178569	R1b1a1a2a1a1c2b1
Z9	6788390	R1b1a1a2a1a1c2b2
Z30	22971392	R1b1a1a2a1a1c2b2a
FGC930	18942433	R1b1a1a2a1a1c2b2a1b1b1a
Z331	17375104	R1b1a1a2a1a1c2b2b
Z326	23973367	R1b1a1a2a1a1c2b2b1a
L188	21868651	R1b1a1a2a1a1c2b2b1a1a1e3a1
FGC396	6997451	R1b1a1a2a1a1d
FGC15054	19383325	R1b1a1a2a1a1e1b
P312	22157311	R1b1a1a2a1a2
L1246	22974840	R1b1a1a2a1a2a~
Z195	17922066	R1b1a1a2a1a2a1
Z274	22293941	R1b1a1a2a1a2a1a
Z209	15508140	R1b1a1a2a1a2a1a1
Z220	16310705	R1b1a1a2a1a2a1a1
Z216	18021350	R1b1a1a2a1a2a1a1a1
Z214	21827794	R1b1a1a2a1a2a1a1a1a
DF17	6631746	R1b1a1a2a1a2a1a2
Z262	16320197	R1b1a1a2a1a2a1b1a
S453	19179436	R1b1a1a2a1a2a7~
U152	15333149	R1b1a1a2a1a2b
L2	5755550	R1b1a1a2a1a2b1
L408	14495249	R1b1a1a2a1a2b1~
Z367	9810794	R1b1a1a2a1a2b1a
L20	14231292	R1b1a1a2a1a2b1a1
Z34	18405864	R1b1a1a2a1a2b1a2
Z52	6893107	R1b1a1a2a1a2b1c1a1a1a
L552	2887377	R1b1a1a2a1a2b1c1b5
S23458	21716324	R1b1a1a2a1a2b1c1c
Z36	4961118	R1b1a1a2a1a2b2
Z56	18848155	R1b1a1a2a1a2b3
Z146	23627289	R1b1a1a2a1a2b3c
Z144	14544130	R1b1a1a2a1a2b3c

Z145	16454755	R1b1a1a2a1a2b3c
Z192	28688760	R1b1a1a2a1a2b-Z192
L21	15654428	R1b1a1a2a1a2c1
Z2542	17885577	R1b1a1a2a1a2c1a
DF49	22735599	R1b1a1a2a1a2c1a1a
DF23	17774409	R1b1a1a2a1a2c1a1a1a1
Z2961	7855252	R1b1a1a2a1a2c1a1a1a1a
M222	14902414	R1b1a1a2a1a2c1a1a1a1a1
FGC30115	7589247	R1b1a1a2a1a2c1a1a1a1a1a1a1a1a3
L371	18656470	R1b1a1a2a1a2c1a1c
DF41	16992602	R1b1a1a2a1a2c1a1d
Z251	8736334	R1b1a1a2a1a2c1a1e
L1335	23154201	R1b1a1a2a1a2c1a1f
Z249	5818205	R1b1a1a2a1a2c1a2
L513	7340450	R1b1a1a2a1a2c1a2
BY206	23728319	R1b1a1a2a1a2c1a2~
Z17815	22750552	R1b1a1a2a1a2c1a2a2a1b1
FGC11803	21827692	R1b1a1a2a1a2c1a2b3a1
L96	14902456	R1b1a1a2a1a2c1a3a1
CTS4466	15660495	R1b1a1a2a1a2c1a3a2
FGC29076	21767963	R1b1a1a2a1a2c1a3a2a1a2d1b1
Z255	15702775	R1b1a1a2a1a2c1a4a
Z253	7193034	R1b1a1a2a1a2c1a4b
L554	15022777	R1b1a1a2a1a2c1a4b1
Z2534	19033817	R1b1a1a2a1a2c1a4b2
Z17971	6906304	R1b1a1a2a1a2c1a4c2~
Y16775	14641164	R1b1a1a2a1a2c1a4d1
DF21	9878915	R1b1a1a2a1a2c1a5
S3058	15683936	R1b1a1a2a1a2c1a5a2a
Z246	7061380	R1b1a1a2a1a2c1a5b1a
L1447	9878916	R1b1a1a2a1a2c1a5b1a1a4a
FGC33062	21907081	R1b1a1a2a1a2c1a5c2b
BY4010	21762715	R1b1a1a2a1a2c1a5c3b1a3a
L1314	14850960	R1b1a1a2a1a2c1a6d~
DF63	2883986	R1b1a1a2a1a2c1b
BY2586	8467026	R1b1a1a2a1a2c1b2a1b
L238	21253443	R1b1a1a2a1a2d
DF19	14301499	R1b1a1a2a1a2e
Z2103	7186135	R1b1a1a2a2
Z2105	15747432	R1b1a1a2a2
L405	28731917	R1b1a1a2a2a
V88	4862861	R1b1a2
V69	18099054	R1b1a2b1b1a
PH1030	14231273	R1b1b

M335	15026395	R1b1b1
M479	20834667	R2
M124	21764501	R2a
FGC12659	8245098	R2a2b1a
Y1379	8467283	R2a2b1b2a1a
L295	4929885	R2a2b1b2b
L723	21778964	R2a2b1b2b1
Y1285	7536663	R2a2b1b2b2
Y1336	24358540	R2a2b1b2b3
Y1332	14301872	R2a2b1b2b3
FGC18152	15437015	R2a2b1b2b3b2a1
P405	8391108	S1a~
P308	15409573	S1a1a1
P60	14484471	S1a1a1~
P202	14001024	S1a1b
P79	6740377	S1a2~
P315	6740065	S1a3
Z42028	15508135	S1d
P336	9004921	S3~
M184	14898163	T
M70	21893881	T1a
PF7463	8391078	T1a
L162	16019072	T1a1
Z709	6890722	T1a1a1b2b
CTS660	6931957	T1a1a1b2b2b
L131	19372808	T1a2
S16112	14484564	T1a2a~
L1255	21926143	T1a3b1

Table S2: Libraries concentration results after re-amplification (8 cycles) due to the low concentration of the first amplification (below 80 pM). Sample's name indicates the region and the number of each sample (A-Andes).

Sample name	First concentration after qPCR (pM)	Concentration after re-amplification (pM)
A61	56	212
A30	41	147
A33	17	54
A37	55	219
A38	47	159
A41	69	287
A48	70	270
A53	55	176
A57	33	108
A60	74	267
A7	22	50
A39	21	44
A40	14	58

Table S3: Population reference Y-STR data of South American Admixed populations.

Y-STR data of South American Admixed populations		
Population	**N**	**Reference**
Caribe,Colombia	69	Present study
Andes, Colombia	98	Present study
Argentina	362	Purps et al., 2014
Peru	83	Purps et al., 2014
Panama	100	Purps et al, 2014
Costa Rica	166	Purps et al, 2014
Bolivia	44	Purps et al., 2014
Chile	978	Toscanini et al., 2016a
Mexico	302	López-Ramírez et al., 2020
North Brazil	874	Palha et al., 2012
Northeast Brazil	279	Palha et al., 2012
Central-West Brazil	143	Palha et al., 2012
Southeast Brazil	405	Palha et al., 2012
Southern Brazil	319	Palha et al., 2012
Patagonia	194	Manuel Fernandes Rodrigues, P., 2021
Ecuador	149	Borges de Pinto Seguro Pereira, M., 2020
Native American-Bolivia	56	Purps et al., 2014
European - Spain	1253	Núñez et al., 2015; Purps et al., 2014; Rey-González et al., 2017
African -Benin	50	Purps et al., 2014

Table S4: Population reference Y-SNP data of South American Admixed populations.

Y-SNP data of South American Admixed populations		
Population	**N**	**Reference**
Caribe,Colombia	69	Present study
Andes, Colombia	98	Present study
Argentina	250	Corach et al. 2011
Colombia	121	Campos Ribeiro, R., 2022
Patagonia	196	Manuel Fernandes Rodrigues, P., 2021
Bolivia	226	Cárdenas et al., 2015
Ecuador	527	Villaescusa et al., 2020
Nicaragua	164	Núñez et al., 2012
North Brazil	262	Resque et al., 2016
Northeast Brazil	243	Resque et al., 2016
Central-West Brazil	135	Resque et al., 2016
Southeast Brazil	330	Resque et al., 2016
South Brazil	290	Resque et al., 2016
Native American - Argentina	78	Toscanini et al., 2011b
European-Spain	660	Adams et al., 2008
African - Equatorial Guinea	182	González et al., 2013

Table S5: Population reference Y-STR data of Eurasian populations.

Y-STR data of Eurasian populations		
Population	**N**	**Reference**
Andes	61	Present study
Caribe	77	Present study
Belgium	105	Purps et al., 2014
Croatia	124	Purps et al., 2014
Czechia	72	Purps et al., 2014
Denmark	185	Purps et al., 2014
England	114	Purps et al., 2014
Aragon, Spain	199	Purps et al., 2014
Asturias, Spain	254	Purps et al., 2014
Barcelona, Spain	78	Purps et al., 2014
Galicia, Spain	46	Purps et al., 2014
Madrid, Spain	125	Purps et al., 2014
Germany	128	Purps et al., 2014
Greece	213	Purps et al., 2014
Brescia, Italy	124	Purps et al., 2014
Puglia, Italy	159	Purps et al., 2014
Netherlands	94	Purps et al., 2014
Poland	150	Purps et al., 2014
North of Portugal	85	Purps et al., 2014
Center of Portugal	83	Purps et al., 2014
South of Portugal	80	Purps et al., 2014
Switzerland	149	Purps et al., 2014
Ireland	30	Purps et al., 2014
Macedonia	101	Purps et al., 2014
Hungary	141	Purps et al., 2014
Finland	254	Purps et al., 2014
Lebanon	503	Purps et al., 2014

Table S6: Population reference Y-SNP data of Eurasian populations.

Y-SNP data of Eurasian populations		
Population	**N**	**Reference**
Andes	62	Present study
Caribe	77	Present study
Portugal	784	Adams et al., 2008; Beleza et al., 2006
Germany	344	Rębała et al., 2013
Poland	571	Rębała et al., 2013
Slovakia	163	Rębała et al., 2013
Croatia	104	Peričić et al., 2005
Serbia	113	Peričić et al., 2005
Denmark	191	Sanchez et al., 2004
France	541	Ramos-Luis et al., 2013
Andalusia, Spain	168	Adams et al., 2008
Basque Country, Spain	140	Adams et al., 2008
Castilla-La Manch, Spain	63	Adams et al., 2008
Catalonia, Spain	80	Adams et al., 2008
Galicia, Spain	88	Adams et al., 2008
Northern Italy	259	Boattini et al., 2013
Central Italy	198	Boattini et al., 2013
Southern Italy	337	Boattini et al., 2013
Sardinia, Italy	80	Boattini et al., 2013
Turkey	497	Cinnioğlu et al., 2004
Lebanon	884	Zalloua et al., 2008

Table S7: Population reference R-M269 Y-SNP data of European populations.

Y-SNP data of R-M269 lineages from European populations		
Population	**N**	**Reference**
Andes	41	Present study
Caribe	34	Present study
Balkan Peninsula	134	Myres et al., 2011
England	294	Busby et al., 2012; Myres et al., 2011
France	300	Myres et al., 2011
Germany	209	Busby et al., 2012; Myres et al., 2011
Ireland	427	Busby et al., 2012; Myres et al., 2011
Portugal	239	Busby et al., 2012; Myres et al., 2011
Russia	50	Myres et al., 2011
Poland	63	Busby et al., 2012; Myres et al., 2011
Scandinavia	68	Myres et al., 2011
Netherlands	1251	Altena et al., 2020; Myres et al., 2011
Wales	105	Busby et al., 2012
Andalusia, Spain	71	Myres et al., 2011
Cantabria, Spain	63	Myres et al., 2011
Castile and Léon, Spain	48	Myres et al., 2011
Valencia, Spain	52	Myres et al., 2011
Northern Italy	133	Boattini et al., 2013
Central Italy	95	Boattini et al., 2013
Southern Italy	92	Boattini et al., 2013
Sardinia	15	Boattini et al., 2013
Turkey	134	Busby et al., 2012; Myres et al., 2011

Table S8: List of amplicons that had ≤20 reads throughout all 182 samples. Y-SNPs located within each amplicon, the sub-haplogroups they define and alternative SNPs available in the kit are also mentioned.

Amplicon	Samples with reads ≤20	SNPs (sub-haplogroups)	Alternative SNPs in the Ion AmpliSeq™ HID Y-SNP Research Panel v1
SP_84.270639	182 samples (100%)	CTS616 (I2a2a1)	CTS9183
SP_85.1.1990847	182 samples (100%)	L529 (Q1b2b1~)	L527
		AF7 (A00)	FGC26856
			L1159
			L1146
			FGC26250
			L1086
			FGC25669
			L1144
		L448 (R1a1a1b1a3a)	Z287*; Z288*; Z93*; Z94*;Z95*,etc;
SP_85.1.417472	180 samples (99%)	F869 (O2a2b1a1a3a)	F444*; CTS11459*; F3607*; N6 *; AM01856*
		AM00868 (C1b1a2b)	M38*; M208*; P33_1*; Z32247*; M347*; M210*; etc;
		P262 (A1b1a1a2a1a)	M32*; M28*; M144*;M190*;M220*;etc
		L817 (J1a2a1a2c1)	L818
		Z37977 (B2b1a2a~)	Z3805
			Z38032

		V8.2 (G2a2b2a1a1c3c~)	PF3359*; L72*;M377*;L183*;F1329*
		AM01856 (O2a2b2a)	None
		Z34737 (H1a2b1b~)	L284*;L286*;M282*;P96*;etc;
SP_86.36938	178 samples (98%)	M7106 (B2b1b~)	M7104
SP_87.210087	177 samples (97%)	FGC396 (R1b1a1a2a1a1d)	FGC15054*; P312*; L1246*; Z195*; etc;
SP_122.112152	182 samples (100%)	Z194 (H1a2a)	Apt
SP_291.643377	182 samples (100%)	U175 (E1b1a1a1a2a)	U209*; Z37284*; U290*; U181*;etc;

*Alternative downstream SNPs included in the panel.

Table S9: STR profile of the final dataset Y-chromosomal samples. A1-A69 is the code name of the samples from Andean region.C1-C98 is the code name of the samples from the Caribbean region. (Table continues below until page 122).

	Haplotype																					
Sample	576	389II	389I	448	19	391	481	549	533	438	437	570	635	390	439	392	643	393	458	385	456	H4
A1	17	16	12	19	15	10	26	12	11	10	15	20	21	22	11	12	12	13	15	13, 16	14	11
A2	18	16	14	19	14	11	22	12	12	12	15	17	23	24	12	13	10	13	18	12, 14	16	12
A3	17	16	13	19	15	10	22	13	12	12	14	16	23	23	12	13	10	13	18	10, 14	15	12
A4	18	16	12	19	14	11	23	13	12	12	15	16	23	24	12	13	10	13	16	11, 14	17	12
A5	19	16	14	19	14	11	21	12	12	12	15	17	23	24	13	13	10	13	17	11, 14	15	12
A6	19	16	13	21	14	10	21	13	13	10	14	19	20	24	11	11	11	13	17	18, 19	15	12
A7	19	17	13	20	17	10	26	14	12	11	14	18	23	25	11	11	10	13	18	10, 14	16	12
A8	14	17	13	20	15	10	32	11	11	11	14	16	21	21	14	11	12	14	16	14, 16	15	12
A9	18	16	14	19	14	10	22	13	12	12	15	18	23	24	11	13	10	13	17	11, 14	15	12
A10	17	17	13	20	14	10	22	12	11	9	15	19	20	23	11	11	10	12	15	14, 18	15	11
A11	17	16	13	19	14	11	22	13	12	12	15	17	23	24	13	13	10	13	17	11, 14	18	12
A12	18	17	12	19	13	11	22	14	12	12	14	17	23	24	13	13	10	15	17	11, 14	17	11
A13	18	16	13	19	14	11	22	13	12	12	15	17	23	24	12	13	10	13	14	11, 15	15	11
A14	19	17	12	20	17	10	21	12	9	10	16	17	22	22	11	11	11	13	16	13, 14	16	11
A15	17	16	14	19	14	11	23	13	13	12	15	17	23	24	12	13	10	13	16	11, 14	16	12

A16	19	16	13	19	14	11	22	15	12	12	15	18	24	24	12	13	10	13	16	11, 14	15	12
A17	18	17	13	18	14	9	22	11	13	9	15	19	21	24	11	12	7	13	16	15, 16	16	11
A18	17	16	13	19	14	11	21	12	12	12	15	18	23	24	11	13	10	13	18	11, 14	15	12
A19	18	16	13	19	14	10	25	12	13	12	15	16	23	24	12	13	10	13	16	11, 14	15	14
A20	18	15	12	19	14	10	22	13	12	12	15	18	23	24	12	13	10	13	17	10, 15	17	12
A21	18	16	14	19	14	11	23	14	12	12	15	16	23	23	11	13	10	14	16	11, 14	16	12
A22	18	16	13	19	14	11	20	12	12	12	15	18	23	25	12	13	10	13	17	11, 14	15	12
A23	19	16	13	19	14	11	22	13	13	12	15	17	23	24	12	13	10	13	18	11, 13.2	16	11
A24	17	16	14	19	14	10	22	12	12	12	15	18	23	25	12	13	10	13	18	11, 15	15	12
A25	17	16	12	21	15	10	23	12	11	9	15	19	26	23	11	10	10	12	16	13, 15	15	11
A26	17	16	12	20	15	10	23	14	14	9	16	17	21	24	12	11	8	12	16	9.1, 14	12	11
A27	17	16	13	20	15	10	22	12	12	12	15	18	23	23	12	13	10	13	17	11, 16	14	10
A28	19	16	14	19	14	10	22	13	12	12	15	17	24	24	12	13	11	13	16	11, 14	15	11
A29	17	16	14	18	15	11	22	13	13	12	15	16	23	23	13	13	10	13	16	10, 14	15	12
A30	17	16	13	19	14	11	21	12	12	12	15	18	23	24	11	13	10	13	18	11, 14	15	12
A31	16	17	13	19	14	11	21	13	11	13	15	17	23	24	12	13	10	12	17	11, 14	16	12
A32	19	16	13	19	13	11	24	12	11	11	14	17	25	24	11	15	10	13	17	15, 18	18	11
A33	16	16	13	19	14	11	23	13	12	12	15	18	23	24	12	13	11	13	17	11, 14	15	12
A34	18	15	13	20	14	11	22	13	12	12	15	18	23	25	12	13	10	13	16	11, 14	16	12
A35	14	16	14	18	14	11	22	13	12	12	14	17	23	24	12	13	11	13	16	13, 14	16	11
A36	17	18	13	20	13	10	26	13	10	10	14	19	21	24	11	11	11	13	18	17, 17	15	11
A37	16	16	13	20	15	9	22	12	12	9	14	18	22	23	12	11	10	13	15	13, 16	16	12

A38	18	16	13	18	14	10	22	14	12	12	14	16	23	24	13	13	10	14	16	11, 14	15	11
A39	16	18	12	22	13	10	25	12	11	11	15	17	23	24	12	14	10	12	17	19, 19	15	13
A40	20	16	13	19	14	11	23	13	12	12	15	17	23	25	13	13	10	13	16	11, 13	15	12
A41	20	16	12	18	14	11	23	12	12	12	14	18	23	25	12	13	10	13	18	11, 14	16	11
A42	16	16	14	19	14	10	22	12	12	12	15	17	23	25	12	14	12	13	17	11, 14	15	12
A43	17	16	12	21	15	10	25	13	11	10	16	20	21	21	12	11	12	14	16	13, 15	14	11
A44	17	16	13	19	14	11	22	13	12	12	15	18	23	24	12	13	10	13	17	11, 14	15	12
A45	17	16	13	19	14	11	22	12	12	12	15	15	23	24	11	13	10	13	16	11, 14	15	12
A46	19	16	13	20	14	11	22	13	14	12	15	18	23	26	12	13	10	13	18	14, 14	15	12
A47	18	19	14	22	13	10	24	13	11	11	15	17	22	24	12	14	10	12	17	14, 18	15	12
A48	18	16	13	18	14	10	22	13	12	12	14	18	23	24	12	13	10	13	18	11, 15	16	12
A49	17	16	13	20	14	10	24	12	12	9	15	16	21	23	11	11	10	12	16	13, 16	15	11
A50	18	16	14	18	14	10	23	13	12	12	14	17	23	24	12	13	10	13	17	11, 14	17	11
A51	18	16	13	19	14	10	24	13	12	9	16	18	24	23	13	13	11	13	16	13, 15	14	11
A52	18	16	13	19	14	10	22	12	13	12	15	16	23	23	12	13	10	13	15	11, 11	16	12
A53	18	15	14	19	15	10	22	11	12	11	16	16	24	24	11	13	9	13	16	13, 16	15	12
A54	18	16	13	19	14	11	22	13	11	12	15	17	23	25	12	13	10	13	17	11, 14	16	11
A55	19	15	13	21	16	11	21	10	12	10	15	19	21	23	12	11	11	13	17	12, 12	13	12
A56	20	17	13	20	15	10	25	13	12	10	14	17	20	23	12	11	9	12	18.2	13, 17	14	12
A57	19	16	13	19	14	11	23	14	12	12	14	17	23	24	12	13	10	15	19	12, 15	15	12
A58	17	18	14	20	13	10	25	12	10	10	14	19	22	25	12	11	12	13	18	16, 16	14	11
A59	18	18	12	20	13	10	24	13	12	10	14	21	22	25	11	11	12	13	15	17, 18	15	11

A60	17	17	13	21	15	9	28	13	12	10	14	22	22	22	11	11	12	13	16	13, 15	13	11
A61	18	18	13	20	13	10	23	11	11	11	14	17	23	24	12	14	10	13	19	15, 15	14	12
A62	17	16	13	19	14	11	22	13	11	12	15	17	23	24	12	13	10	13	18	11, 13	16	12
A63	15	16	13	18	14	10	22	14	12	12	15	17	23	25	12	13	10	13	20	11, 13	16	12
A64	16	18	12	21	16	10	24	11	11	11	14	19	21	21	12	11	13	14	18	15, 17	14	11
A65	19	16	14	19	14	11	22	13	12	12	15	18	23	24	13	13	10	13	17	11, 15	16	12
A66	18	18	14	20	15	9	28	12	12	10	14	18	21	23	12	12	12	14	15	13, 16	13	11
A67	19	16	13	19	14	11	23	14	12	10	15	17	23	24	13	13	12	13	17	11, 14	16	12
A68	18	16	12	19	13	10	23	12	11	11	14	19	22	21	11	15	10	14	15	16, 16	16	11
A69	15	15	12	19	15	10	23	13	11	9	14	16	21	25	11	14	9	13	18	13, 16	15	11
C1	17	16	13	19	14	11	22	13	12	12	15	17	23	24	12	13	10	13	16	11, 11	15	12
C2	18	18	13	19	16	11	30	11	13	10	14	18	23	24	12	11	10	13	17	14, 15	15	11
C3	18	17	14	20	16	11	30	11	12	10	15	18	23	25	13	11	10	13	17	14, 15	15	10
C4	17	17	13	20	13	10	22	13	13	10	14	19	22	24	12	11	13	13	17	16, 18	16	13
C5	17	19	13	20	17	11	30	12	13	10	15	18	22	24	13	11	10	13	17	14, 16	15	11
C6	17	16	13	19	14	11	22	13	12	12	15	17	23	23	12	13	10	13	16	11, 11	15	12
C7	17	19	13	20	16	11	30	11	13	10	15	18	22	24	14	11	10	13	17	14, 14	15	11
C8	17	17	12	20	13	10	27	11	10	10	14	19	21	23	12	11	12	13	17	16, 18	15	9
C9	18	17	13	19	16	11	27	11	12	10	15	18	23	24	14	11	10	13	18	14, 15	15	11
C10	17	19	13	20	16	11	30	11	13	10	15	17	22	24	13	11	10	13	17	14, 14	15	11
C11	18	18	14	20	16	11	30	11	12	10	15	19	23	24	14	11	10	13	17	15, 15	15	10
C12	20	16	13	?	14	10	22	13	12	12	15	17	23	22	12	13	10	13	15	11, 14	17	12

C13	18	18	13	19	17	10	30	11	13	10	15	18	21	24	13	11	10	14	17	14, 15	15	11
C14	17	17	13	21	16	11	24	12	12	11	14	19	23	25	12	11	10	13	15	11, 14	15	12
C15	18	18	13	20	16	11	30	11	13	10	15	18	23	24	11	11	10	13	17	14, 15	15	?
C16	17	17	13	20	13	10	23	12	12	10	14	20	21	24	12	11	12	13	15	17, 18	17	11
C17	18	16	13	20	15	10	25	12	12	11	14	19	23	25	12	11	10	13	16	11, 14	17	12
C18	18	17	13	20	16	11	31	11	13	10	15	18	23	24	12	11	10	14	17	14, 15	15	11
C19	16	17	13	20	16	11	23	12	12	11	14	18	24	25	11	11	10	13	15	11, 15	16	13
C20	17	17	13	21	16	11	24	12	12	11	14	19	23	25	12	11	10	13	15	11, 14	15	12
C21	19	16	13	20	16	10	25	12	12	11	14	20	23	25	11	11	10	13	15	11, 14	17	12
C22	18	17	13	20	16	11	28	11	12	10	15	19	23	23	13	11	10	13	17	14, 14	15	11
C23	18	19	13	19	17	10	30	11	13	10	15	19	22	24	13	11	10	13	18	14, 15	15	11
C24	18	17	14	20	14	10	24	12	12	10	14	18	24	24	12	11	12	13	16	15, 17	15	11
C25	17	18	13	19	15	11	29	12	12	10	15	18	21	24	13	11	10	13	17	14, 15	15	11
C26	20	17	14	20	13	10	22	12	12	10	14	21	22	24	13	11	12	13	14	16, 19	16	11
C27	19	16	14	18	14	11	22	13	12	12	14	19	23	24	11	13	10	13	18	12, 14	15	11
C28	17	18	13	20	16	11	31	12	12	10	15	17	23	24	12	11	10	13	17	14, 15	15	11
C29	14	17	12	21	15	10	21	12	9	10	16	18	21	22	12	11	11	14	16	13, 14	15	12
C30	16	17	13	20	14	10	22	13	12	10	14	20	22	25	14	11	12	13	16	18, 19	17	13
C31	15	18	14	21	16	10	28	11	13	11	14	19	21	21	13	11	13	13	17	15, 17	15	12
C32	18	17	14	20	13	10	22	13	12	10	14	19	22	24	13	11	12	13	15	16, 17	16	11
C33	16	17	13	21	16	10	26	11	11	11	14	17	21	21	12	11	13	15	16	16, 18	17	11
C34	18	16	14	18	14	11	22	12	12	12	14	18	23	23	12	14	10	13	18	11, 14	16	11

C35	18	16	13	18	13	10	22	12	13	11	15	17	23	27	11	13	10	13	18	11, 11	17	11
C36	15	18	13	21	15	10	27	11	11	11	14	19	23	21	13	11	16	12	17	16, 17	15	12
C37	16	17	12	21	15	10	23	13	10	10	16	19	22	23	13	11	11	14	15	14, 15	15	12
C38	18	17	13	19	13	11	22	12	12	12	15	17	23	25	12	13	10	13	16	11, 15	16	11
C39	14	18	13	21	15	10	27	11	11	10	14	20	22	22	11	11	14	14	17	16, 17	15	12
C40	16	17	13	21	17	9	25	11	11	11	14	17	21	21	11	11	14	13	17	16, 19	16	11
C41	14	17	13	21	15	10	28	12	11	11	14	18	21	21	12	11	12	14	17	15, 19	15	13
C42	18	15	13	22	16	10	22	10	13	10	15	19	21	23	11	11	12	13	17	12, 12	14	12
C43	15	17	13	20	15	11	22	13	12	12	14	17	22	24	12	13	10	13	16	11, 14	15	12
C44	19	17	12	19	14	10	23	12	12	12	15	17	24	24	12	13	10	12	17	11, 14	15	13
C45	17	18	13	19	13	11	21	11	12	10	14	20	24	24	12	11	13	14	15	17, 19	17	11
C46	19	16	12	18	14	11	23	13	12	12	14	17	23	24	12	14	12	13	18	11, 14	15	11
C47	18	16	14	19	14	11	22	13	12	12	14	17	23	24	11	13	10	13	18	11, 15	16	13
C48	18	18	14	20	13	11	23	12	11	11	14	17	22	23	12	14	10	13	17	14, 17	15	11
C49	16	17	13	19	15	10	23	12	12	7	16	16	22	24	13	11	9	12	19	14, 14	15	12
C50	19	16	13	19	15	11	22	13	12	11	15	17	23	24	12	13	10	13	17	11, 14	15	12
C51	18	16	13	20	15	11	26	13	13	10	15	18	20	23	11	12	13	14	16	15, 16	16	11
C52	17	16	13	19	15	11	24	13	12	12	15	17	24	24	12	13	9	13	16	11, 14	15	12
C53	18	16	13	18	14	11	22	12	12	12	14	17	23	23	12	13	10	13	17	11, 14	16	12
C54	16	16	13	19	14	11	22	13	12	12	15	17	23	24	13	13	10	13	18	11, 15	15	12
C55	18	16	13	19	14	11	24	13	12	12	15	18	23	24	11	13	10	13	16	12, 15	15	12
C56	18	16	13	19	15	10	21	13	14	12	15	18	23	22	11	13	10	12	17	11, 14	16	12

C57	18	18	13	19	15	11	22	13	12	12	14	17	23	23	11	13	10	13	19	11, 15	16	12
C58	16	17	12	20	15	10	28	11	11	11	14	18	21	21	12	11	13	13	17	14, 17	15	12
C59	18	16	13	19	14	9	18	12	11	7	16	15	24	23	12	11	10	12	17	14, 16	15	11
C60	16	18	13	21	15	10	28	13	11	11	14	21	21	21	13	10	13	13	17	16, 18	15	12
C61	17	15	13	21	17	10	23	12	12	10	15	19	22	23	12	11	12	13	15	12, 12	14	12
C62	19	18	13	21	14	10	24	13	11	10	14	17	21	23	11	11	9	12	17.2	13, 15	15	11
C63	16	16	12	20	15	10	25	12	11	10	16	17	22	22	12	11	12	13	14	13, 14	14	11
C64	16	17	12	19	14	11	22	13	12	12	15	16	23	24	12	13	10	13	16	11, 14	16	13
C65	15	18	13	20	13	10	22	12	12	10	14	18	22	24	11	11	11	13	15	15, 18	15	12
C66	16	17	13	20	13	10	22	12	11	10	14	21	22	24	12	11	12	13	15	16, 18	16	11
C67	15	16	13	21	16	10	27	12	11	11	14	18	21	21	11	11	12	13	18	17, 17	16	12
C68	18	15	14	18	14	11	22	13	13	12	14	17	23	24	11	13	10	13	18	11, 14	16	11
C69	18	16	15	18	14	11	22	12	12	12	15	16	23	24	12	13	10	13	17	11, 14	15	11
C70	18	17	13	21	16	9	25	11	12	11	14	17	22	21	13	11	13	14	16	16, 17	15	11
C71	18	15	14	18	14	11	22	13	13	12	14	17	23	24	11	13	10	13	18	11, 14	16	11
C72	15	16	13	21	16	10	27	12	11	11	14	18	21	21	11	11	12	13	18	17, 17	16	12
C73	17	17	14	21	17	10	25	11	11	11	14	15	21	21	12	11	13	15	16	17, 17	16	11
C74	16	16	13	21	17	10	25	11	11	11	14	17	21	21	13	11	13	15	16	19, 2	15	11
C75	19	16	13	20	14	10	22	14	13	12	15	16	23	24	12	13	11	13	17	11, 14	15	11
C76	16	17	13	21	17	10	22	11	11	11	14	16	21	21	12	11	13	14	16	17, 19	16	11
C77	18	16	13	19	14	10	22	13	12	12	15	19	23	24	12	13	10	13	17	11, 14	16	12
C78	18	16	14	19	13	10	27	12	11	11	14	18	22	22	12	14	10	13	16	14, 17	15	14

C79	20	17	13	19	13	10	25	12	11	11	14	16	22	24	12	14	10	11	13	14, 14	16	13
C80	16	17	13	20	13	9	28	12	12	10	14	15	21	23	13	11	11	14	16	15, 16	15	12
C81	20	14	13	19	14	11	23	14	11	13	15	17	24	24	12	13	10	13	17	11, 14	17	11
C82	18	17	13	19	14	10	22	12	12	12	15	15	23	24	12	13	10	13	17	11, 14	15	12
C83	18	16	12	20	14	11	22	13	12	12	15	18	23	24	13	13	9	12	17	11, 15	16	12
C84	19	18	13	21	14	10	22	11	12	9	15	13	21	23	11	11	10	12	14	12, 17	15	11
C85	19	17	14	19	14	11	22	13	12	12	15	15	23	24	12	13	10	13	18	11, 14	15	11
C86	17	16	14	19	14	10	22	13	13	12	14	17	23	24	12	13	10	13	17	11, 14	17	11
C87	18	17	14	19	14	10	24	12	12	9	15	18	20	25	12	11	9	12	15	13, 15	15	11
C88	18	17	13	19	14	11	22	13	11	12	15	17	24	24	12	13	10	13	17	11, 15	16	12
C89	15	17	13	21	15	11	28	11	11	11	14	19	21	21	11	11	14	13	15	16, 17	15	12
C90	17	18	13	20	13	10	24	13	11	11	14	18	23	25	13	14	10	13	17	14, 18	15	12
C91	18	18	13	20	16	10	27	12	12	10	14	19	21	23	11	12	14	14	15	15, 15	15	11
C92	18	16	13	18	14	11	22	13	13	12	15	20	23	24	11	13	10	14	18	11, 15	16	12
C93	18	17	13	20	14	8	22	12	12	9	14	17	21	23	11	11	11	12	15	14, 19	16	11
C94	18	17	14	18	14	11	24	13	12	12	14	17	23	24	11	13	10	13	17	11, 13	16	11
C95	19	17	13	20	13	10	24	13	11	11	14	18	23	24	11	15	10	13	18	14, 17	14	12
C96	19	16	13	18	14	11	22	13	12	12	14	17	23	24	11	13	10	13	18	11, 14	16	11
C97	18	17	14	21	17	10	27	11	11	11	14	16	20	21	12	11	13	15	17	17, 18	15	10
C98	18	16	13	19	14	10	22	12	12	12	15	16	23	24	12	13	10	12	16	11, 14	15	13

Table S 10: Haplogroup prediction of the final dataset samples. Haplogroups were predicted with both NevGen, based on the STR profiles, and Y-leaf, based on the SNP allelic state, software's. A1-A69 is the code name of the samples from Andean region.C1-C98 is the code name of the samples from the Caribbean region. (Table continues below until page 129).

Haplogroup prediction		
Sample	NevGen	Y-leaf
A1	I1 Z58>Z138>S2293>S6277> Y15902	I-Z63*(xL1237,Y17387)
A2	R1b	R-Z18*(xL257)
A3	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A4	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A5	R1b	R-Z262
A6	E1b1b V22	E-V22*(xZ36763,Y24403)
A7	R1a	R-L260
A8	E1b1a V38>> L485	E-U209*(xZ37284,U290)
A9	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A10	J2a1 PF5191>> PF5177	J-Z2221
A11	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A12	R1b	R-Z2542*(xDF49,L371,DF41,Z251,L1335,L513,CTS4466,Z255,Z253,DF21,L1314)
A13	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A14	G2a2 > U1> L13	G-Z764*(xU1,Z738,CTS342)
A15	R1b	R-P310/etc*(xU106,P312)
A16	R1b	R-Z56*(xZ144)
A17	J2a2 PF5008	J-PF5008*(xM340)
A18	R1b	R-Z253*(xL554,Z2534)

A19	R1b	R-Z2542*(xDF49,L371,DF41,Z251,L1335,L513,CTS4466,Z255,Z253,DF21,L1314)
A20	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A21	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A22	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A23	R1b	R-Z262
A24	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A25	J2a1 Z7700> SK1382	J-L26*(xPF5116,PF5125,PF5172,PF5294,Z1846,Z2220,Z2221,PF5197,Z6065,M322,M67,M68,M319,L24)
A26	J2b2a M241>>CTS3617	J-L283
A27	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A28	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A29	R1b	R-DF17
A30	R1b	R-Z253*(xL554,Z2534)
A31	R1b	R-Z2103*(xL405)
A32	Q M346>> Z780	Q-M3*(xM19,M194,P106,SA01,L891,Z35615,BZ3401,Y18425)
A33	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A34	R1b	R-Z156(xS498)
A35	R1b	R-Z216*(xZ214)
A36	E1b1b M123>M34> M84	E-FGC18329*(xY16613,Y19892)
A37	J2a1 Z387>L70> Z2177	J-L398*(xM137)
A38	R1b	R-Z209*(xZ216)
A39	Q M346>> M3> M902	Q-M3*(xM19,M194,P106,SA01,L891,Z35615,BZ3401,Y18425)
A40	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A41	R1b	R-Z209*(xZ216)
A42	R1b	R-Z301*(xU198,L48)
A43	I1 DF29> BY13858	I-DF29*(xCTS6364,Z58,Z63)
A44	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A45	R1b	R-P312*(xL1246,U152,L21,L238,DF19)

A46	R1b	R-Z251
A47	Q M346>> M3> M902	Q-M3*(xM19, M194,P106,SA01,L891,Z35615,BZ3401,Y18425)
A48	R1b	R-Z214
A49	J2a1 Z6065> Y7708>M47	J2a1a-M322
A50	R1b	R-Z214
A51	T L131 > CTS933	T1a2-L131(xS16112)
A52	R1b	R-P312*(xL1246, U152,L21,L238,DF19)
A53	R1b V88	R-L780*(xL389, V88,PH1030)
A54	R1b	R-P312*(xL1246, U152,L21,L238,DF19)
A55	I2a1a Sardinian M26>PF4088	I-M26
A56	J1a >> P58>> FGC11>> S20171	J-Z640
A57	R1b	R-Z195*(xZ274,Z262)
A58	E1b1b M123>M34> M84	E-FGC18329*(xY16613,Y19892)
A59	E1b1b L67	E-V65
A60	E1b1b V257> PF2431	E-M310*(xM81)
A61	Q M346>> M3> M902	Q-M3*(xM19, M194,P106,SA01,L891,Z35615,BZ3401,Y18425)
A62	R1b	R-P312*(xL1246 ,U152,L21,L238,DF19)
A63	R1b	R-P312*(xL1246, U152,L21,L238,DF19)
A64	E1b1a V38>> L485	E-L485*(xCTS5038)
A65	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A66	I2a2a M223>Z161>L801>CTS6433> S2364>ZS20	I-L702*(xP78)
A67	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
A68	Q M346>> Z780	Q-M3*(xM19,M194,P106,SA01,L891,Z35615,BZ3401,Y18425)
A69	T >> CTS11451> PF7455	T1a1-L162(xZ709,CTS660)
C1	R1b	R-Z2103*(xL405)
C2	I2a1b3a Slavic-Carpathian > S17250> BY37513	I-L621

C3	I2a1b3a Slavic-Carpathian> S17250> Z16971> A815	I-L621
C4	E1b1b > V13	E-V13*(xZ36786,Z42903)
C5	I2a1b3a Slavic-Carpathian> S17250> Y4882> A1328	I-L621
C6	R1b	R-Z2103*(xL405)
C7	I2a1b3a Slavic-Carpathian> Z17855> A16413	I-L621
C8	E1b1b M123>M34> M84	E-FGC18329*(xY16613,Y19892)
C9	I2a1b3a Slavic-Carpathian > S17250	I-L621
C10	I2a1b3a Slavic-Carpathian> Z17855	I-L621
C11	I2a1b3a Slavic-Carpathian> S17250> Y4882> A1328	I-L621
C12	R1b	R-FGC930
C13	I2a1b3a Slavic-Carpathian > S17250> A356	I-L621
C14	R1a	R-M558(xCTS3402)
C15	I2a1b3a Slavic-Carpathian> Z17855	I-L621
C16	E1b1b > V13	E-V13*(xZ36786,Z42903)
C17	R1a	R-M458*(xL260)
C18	I2a1b3a Slavic-Carpathian> S17250> Y4882	I-L621
C19	R1a	R-CTS3402*(xL365)
C20	R1a	R-M558(xCTS3402)
C21	R1a	R-M458*(xL260)
C22	I2a1b3a Slavic-Carpathian> S17250> Z16971> A815	I-L621
C23	I2a1b3a Slavic-Carpathian > S17250> A5913	I-L621
C24	E1b1b PF1975	E-M34*(xFGC18329,M290)
C25	I2a1b3a Slavic-Carpathian > S17250	I-L621
C26	E1b1b > V13	E-V13*(xZ36786,Z42903)
C27	R1b	R-Z214
C28	I2a1b3a Slavic-Carpathian> Y4460	I-L621

C29	G2a2 > L497>> Z725>> CTS4803> S2808> S23438	G-CTS4803*(xZ31323)
C30	E1b1b V22	E-V22*(xZ36763,Y24403)
C31	E1b1a V38>> M4231	E-U209*(xZ37284,U290)
C32	E1b1b > V13	E-V13*(xZ36786,Z42903)
C33	E1b1a V38>> L485	E-U174*(xZ37817,CTS490,L372)
C34	R1b	R-Z214
C35	R1b	R-DF41
C36	E1b1a V38>> M4231	E-U181
C37	G2a2 > L497>> Z725>> CTS4803> S2808> S23438	G-Z43085
C38	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
C39	E1b1a V38>> M4231	E-U290*(xU181)
C40	E1b1a V38>> L485	E-U174*(xZ37817,CTS490,L372)
C41	E1b1a V38>> M4231	E-U209*(xZ37284,U290)
C42	I2a1a Sardinian M26>PF4088	I-M26
C43	R1b	R-L2*(xZ367,Z52,L552,S23458)
C44	R1b	R-L2*(xZ367,Z52,L552,S23458)
C45	E1b1b > V13	E-V13*(xZ36786,Z42903)
C46	R1b	R-Z209*(xZ216)
C47	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
C48	Q M346>> M3> M902	Q-M3*(xM19,M194,P106,SA01,L891,Z35615,BZ3401,Y18425)
C49	J2a1 Z6046> SK1313	J-L559(xL26,PF5008)
C50	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
C51	I2a2a M223>Y4450> L1229>> BY524	I-L1229
C52	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
C53	R1b	R-Z214
C54	R1b	R-P312*(xL1246,U152,L21,L238,DF19)

C55	R1b	R-Z56*(xZ144)
C56	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
C57	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
C58	E1b1a V38>> M4231	E-U290*(xU181)
C59	J2a1 Z7671> CTS900> CTS6804	J2a1b-M67(xM92,M166,L210,L227)
C60	E1b1a V38>> M4231	E-U209*(xZ37284,U290)
C61	I2a1a Sardinian M26>PF4088	I-M26
C62	J1a >> P58>> ZS241>> Z18271> Z18290	J-Z2329*(xZ1884)
C63	I1 Z58>Z59>> CTS7362>> BY266> Z2900	I-Z139
C64	R1b	R-P312*(xL1246,U152,L21,L238,DF19)
C65	E1b1b > V12	E-V12*(xM224,V32,BY8427,BY8199)
C66	E1b1b > V13	E-V13*(xZ36786,Z42903)
C67	E1b1a V38>> M4231	E-U209*(xZ37284,U290)
C68	R1b	R-Z214
C69	R1b	R-Z214
C70	E1b1a V38>> L485	E-U174*(xZ37817,CTS490,L372)
C71	R1b	R-Z214
C72	E1b1a V38>> M4231	E-U209*(xZ37284,U290)
C73	E1b1a V38>> L485	E-U174*(xZ37817,CTS490,L372)
C74	E1b1a V38>> L485	E-U174*(xZ37817,CTS490,L372)
C75	R1b V88 >> V1589	R-P312*(xL1246,U152,L21,L238,DF19)
C76	E1b1a V38>> L485	E-U174*(xZ37817,CTS490,L372)
C77	R1b	R-Z262
C78	Q M346> YP4004	Q-M3*(xM19,M194,P106,SA01,L891,Z35615,BZ3401,Y18425)
C79	Q M346>> M3> M902	Q-M3*(xM19,M194,P106,SA01,L891,Z35615,BZ3401,Y18425)
C80	E1b1b M123>M34> Z841	E-M34*(xFGC18329,M290)
C81	R1b	R-Z251

C82	R1	R-Z195*(xZ274,Z262)
C83	R1b	R-L405
C84	J2a1 Z6063>> BY759	J-L26*(xPF5116,PF5125,PF5172,PF5294,Z1846,Z2220,Z2221,PF5197,Z6065,M322,M67,M68,M319,L24)
C85	R1b	R-Z2534
C86	R1b	R-Z214
C87	J2a2 PF5008 >> Y29673	J-PF5008*(xM340)
C88	R1b	R-Z2542*(xDF49,L371,DF41,Z251,L1335,L513,CTS4466,Z255,Z253,DF21,L1314)
C89	E1b1a V38>> M4231	E-U209*(xZ37284,U290)
C90	Q M346>> M3> M902	Q-M3*(xM19,M194,P106,SA01,L891,Z35615,BZ3401,Y18425)
C91	I2a2a M223>Z161>L801>CTS6433> L1425	I-CTS6433*(xZ78,CTS661)
C92	R1b	R-Z195*(xZ274,Z262)
C93	J2a1 PF5087> Z7430	J2a1b-M67(xM92,M166,L210,L227)
C94	R1b	R-Z214
C95	Q M346>> M3> M902	Q-M3*(xM19, M194 ,P106,SA01,L891,Z35615,BZ3401,Y18425)
C96	R1b	R-Z209*(xZ216)
C97	E1b1a V38>> L485	E-U174*(xZ37817, CTS490, L372)
C98	R1b	R-Z2103*(xL405)

Table S11: Pairwise R_{ST} genetic distances and, respectively, *p*-values of present study Colombian samples and other South American populations. South American populations are represented in green; Colombia regions in blue; a European reference population from Spain in orange, one Native American population from Bolivia is coloured in light orange as well as one sub-Saharan population from Benin. R_{ST} genetic distances are below the diagonal and the respectively *p*-values above it. Non-significant values are highlighted in red ($p \geq 0.00029$) (Bonferroni-*p*-corrected). Population names are represented in the following order: AND-Andes; CAR-Caribe; COL-Colombia; ARG-Argentina; ECU-Ecuador; BOL-Bolivia; PAT-Patagonia; NBR-Norh, Brazil; CBR-Central, Brazil; NEBR-North East, Brazil; SBR-South, Brazil; SEBR-South East, Brazil; NIC-Nicaragua; NAM-Native American Bolivia; SUB-Sub-Saharan, Benin; SPA-Spain.

	CHI	BOL	PER	ARG	PAN	COR	MEX	SPA	ECU	PAT	NBR	NOBR	CWBR	SEBR	SBR	NAM	SUB	AND	CAR
CHI	-	0.00000	0.00000	0.01257	0.00000	0.00267	0.00000	0.00000	0.00594	0.00723	0.00000	0.00000	0.00109	0.00010	0.00010	0.00000	0.00000	0.29363	0.00000
BOL	0.13464	-	0.00604	0.00000	0.00000	0.00000	0.01247	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.05485	0.00000	0.00000	0.00000
PER	0.06217	0.03218	-	0.00000	0.00000	0.00000	0.00099	0.00000	0.00139	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000
ARG	0.00324	0.13503	0.06079	-	0.00010	0.02208	0.00000	0.00000	0.01693	0.76775	0.00000	0.00109	0.01544	0.00000	0.00030	0.00000	0.00000	0.10593	0.00000
PAN	0.04306	0.12819	0.07020	0.02723	-	0.00050	0.00000	0.00000	0.00030	0.00149	0.00000	0.00030	0.00020	0.00000	0.00000	0.00000	0.00000	0.00020	0.00000
COR	0.00928	0.13302	0.05477	0.00610	0.03123	-	0.00000	0.00010	0.04148	0.07059	0.00059	0.00267	0.01544	0.00465	0.20453	0.00000	0.00000	0.28255	0.00198
MEX	0.08622	0.01871	0.02141	0.08899	0.10995	0.09257	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
SPA	0.00749	0.15719	0.08784	0.01934	0.07260	0.01656	0.11128	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.52034	0.00000
ECU	0.00876	0.08969	0.02424	0.00752	0.02503	0.00734	0.05263	0.02467	-	0.00990	0.00000	0.00000	0.00089	0.00030	0.00030	0.00000	0.00000	0.12850	0.00010
PAT	0.00683	0.16015	0.07588	-0.00158	0.02151	0.00480	0.11051	0.02388	0.01083	-	0.00059	0.19523	0.16820	0.00446	0.02663	0.00000	0.00000	0.05792	0.00050
NBR	0.00645	0.13305	0.07533	0.01129	0.03840	0.01121	0.10309	0.00803	0.01886	0.01056	-	0.00000	0.00921	0.00059	0.00475	0.00000	0.00000	0.17137	0.00000
NOBR	0.02110	0.17768	0.09025	0.00906	0.02326	0.01225	0.13589	0.03819	0.02167	0.00153	0.01920	-	0.02079	0.00040	0.00139	0.00000	0.00000	0.00178	0.00426
CWBR	0.01304	0.15377	0.08680	0.00760	0.02974	0.01066	0.11897	0.02231	0.02011	0.00264	0.00744	0.00782	-	0.28106	0.08217	0.00000	0.00000	0.05435	0.00079
SEBR	0.01125	0.15020	0.07768	0.01279	0.03904	0.00950	0.11381	0.01475	0.01705	0.00873	0.00550	0.01117	0.00087	-	0.11494	0.00000	0.00000	0.07059	0.00020
SBR	0.01059	0.15816	0.08014	0.00985	0.04130	0.00159	0.11831	0.01418	0.01744	0.00540	0.00506	0.00954	0.00389	0.00175	-	0.00000	0.00000	0.14583	0.00050
NAM	0.21973	0.01868	0.08914	0.21055	0.20636	0.21453	0.05568	0.25669	0.16889	0.24432	0.21970	0.25526	0.23689	0.23464	0.24443	-	0.00000	0.00000	0.00000
SUB	0.38246	0.51625	0.42967	0.33658	0.27213	0.33565	0.45859	0.43031	0.34911	0.33715	0.34613	0.26966	0.32221	0.32302	0.33305	0.56298	-	0.00000	0.00000
AND	0.00119	0.11097	0.05208	0.00549	0.04392	0.00189	0.07982	-0.00128	0.00584	0.00990	0.00328	0.02352	0.01044	0.00722	0.00439	0.21500	0.41980	-	0.00030
CAR	0.04913	0.21343	0.10803	0.03378	0.04654	0.02505	0.16555	0.07360	0.04034	0.02371	0.04379	0.01391	0.02737	0.02876	0.02252	0.29295	0.26424	0.05342	-

Table S12: Pairwise F_{ST} genetic distances and, respectively, p-values of present study Colombian samples and other South American populations. South American populations are represented in green; Colombia regions in blue; a European reference population from Spain in orange; one Native American from Argentina is represented in light orange as well as sub-Saharan population. F_{ST} genetic distances are below the diagonal and the respectively p-values above it. Non-significant values are highlighted in red ($p \geq 0.00042$) (Bonferroni-p-corrected). Population names are represented in the following order: AND-Andes; CAR-Caribe; COL-Colombia; ARG-Argentina; ECU-Ecuador; BOL-Bolivia; PAT-Patagonia; NBR-North, Brazil; CBBR-Center, Brazil; NEBR-Northeast Brazil; SBR-South, Brazil; SEBR- Southeast, Brazil; NIC-Nicaragua; NAM-Native American, Argentina; SU-Sub-Saharan Equatorial Guinea; SPA-Spain.

	AND	CAR	COL	ARG	ECU	BOL	PAT	NBR	CBR	NEBR	SBR	SEBR	NIC	NAM	SU	SPA
AND	-	0.00020	0.00000	0.15216	0.00000	0.04168	0.03970	0.00059	0.00010	0.00000	0.00000	0.00000	0.04406	0.00000	0.00000	0.00198
CAR	0.06040	-	0.00000	0.00020	0.00000	0.00000	0.00030	0.01267	0.00119	0.00000	0.00000	0.00307	0.00119	0.00000	0.00000	0.00000
COL	0.10979	0.09622	-	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
ARG	0.00566	0.03625	0.09392	-	0.00000	0.00000	0.63647	0.00010	0.00079	0.00000	0.00000	0.00000	0.05435	0.00000	0.00000	0.00000
ECU	0.16677	0.11953	0.03616	0.14766	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
BOL	0.01475	0.07959	0.06205	0.03633	0.12589	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00109	0.00000	0.00000	0.00000
PAT	0.01463	0.03148	0.08414	-0.00156	0.13095	0.04203	-	0.00020	0.00198	0.00000	0.00000	0.00000	0.06732	0.00000	0.00000	0.00000
NBR	0.04210	0.01335	0.06747	0.02039	0.11497	0.05067	0.01908	-	0.14504	0.00000	0.00000	0.00990	0.00287	0.00000	0.00000	0.00000
CBR	0.05184	0.02967	0.09789	0.02088	0.14444	0.07845	0.02097	0.00319	-	0.00000	0.00030	0.05158	0.00149	0.00000	0.00000	0.00000
NEBR	0.14866	0.04833	0.13561	0.10301	0.16009	0.16400	0.09359	0.03681	0.03349	-	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000
SBR	0.10567	0.04769	0.11450	0.06811	0.14814	0.11642	0.05541	0.02377	0.02426	0.02018	-	0.00000	0.00000	0.00000	0.00000	0.00000
SEBR	0.07711	0.01814	0.10661	0.04809	0.14145	0.09560	0.04651	0.00715	0.00632	0.01458	0.02161	-	0.00000	0.00000	0.00000	0.00000
NIC	0.01417	0.02742	0.05113	0.00614	0.09809	0.02333	0.00576	0.01354	0.02176	0.08937	0.06302	0.03874	-	0.00000	0.00000	0.00000
NAM	0.33910	0.27190	0.07592	0.28605	0.09923	0.23791	0.27356	0.22813	0.28225	0.27473	0.26958	0.26290	0.22639	-	0.00000	0.00000
SU	0.30381	0.11843	0.28734	0.25193	0.27839	0.30611	0.24291	0.18063	0.21735	0.14280	0.19172	0.15188	0.23198	0.41012	-	0.00000
SPA	0.03797	0.18604	0.27491	0.08958	0.29422	0.08779	0.11168	0.15284	0.17443	0.29799	0.23034	0.19211	0.12383	0.49585	0.44535	-

Table S13: Pairwise R_{ST} genetic distances and, respectively, p-values of present study Colombian samples and other Eurasian populations. South American populations are represented in green; Colombia regions in blue and one population from Lebanon is representing Middle Eastern coloured in light orange. R_{ST} genetic distances are below the diagonal and the respectively p-values above it. The statistically significant values are highlighted in red ($p \geq 0.00014$) (Bonferroni-p-corrected). Population names are represented in the following order AND-Andes; CAR-Caribe; BEL-Belgium; CRO-Croatia; CZC-Czechia; DNK-Denmark; ENG-England; ARA-Aragon, Spain; AST-Asturias, Spain; BAR-Barcelona, Spain; GAL-Galicia, Spain; MAD-Madrid, Spain; GRM-Germany; BRE-Brescia; PUG-Puglia; NTH-Netherlands; POL-Poland; NPT-North, Portugal; CPT-Center, Portugal; SPT- South, Portugal; SWT-Switzerland; IRE-Ireland; MAC-Macedonia; HUNG-Hungary; LEB-Lebanon; FIN-Finland.

	AND	CAR	BEL	CRO	CZC	DNK	ENG	ARA	AST	BAR	GAL	MAD	GRM	GRE	BRE	PUG	NTH	POL	NPT	CPT	SPT	SWT	IRE	MAC	HUNG	LEB	FIN
AND	•	0.50559	0.79408	0.19850	0.50302	0.06287	0.03495	0.53480	0.74834	0.91694	0.22730	0.47471	0.00000	0.25403	0.68399	0.18246	0.00069	0.00000	0.57955	0.39501	0.59073	0.49837	0.37264	0.25017	0.38046	0.00000	0.00000
CAR	-0.00953	•	0.34878	0.44253	0.71924	0.09405	0.00426	0.22503	0.59578	0.24542	0.68765	0.23166	0.00604	0.13553	0.52252	0.10138	0.00505	0.00000	0.50579	0.25027	0.24938	0.56034	0.23770	0.17097	0.87912	0.00000	0.00000
BEL	-0.00889	-0.00330	•	0.08356	0.33700	0.01713	0.06217	0.70706	0.32145	0.75854	0.29918	0.82992	0.00406	0.36204	0.21196	0.25245	0.11741	0.00000	0.55985	0.84348	0.97980	0.20394	0.28453	0.33759	0.17830	0.00000	0.00000
CRO	0.00478	-0.00041	0.00929	•	0.34541	0.00000	0.00000	0.01148	0.00881	0.10890	0.26423	0.02930	0.00000	0.01485	0.00000	0.00832	0.00020	0.00000	0.13256	0.05623	0.10821	0.00010	0.21612	0.05079	0.48668	0.00000	0.00000
CZC	-0.00994	-0.01091	-0.00388	0.00104	•	-0.00294	0.52421	0.29344	0.21612	0.60746	0.26017	0.71547	0.33492	0.22057	0.10009	0.52945	0.07197	0.27086	0.00000	0.58489	0.28542	0.25245	0.89209	0.37046	0.09950	0.73052	0.00000
DNK	0.01064	0.00311	0.01784	0.01505	-0.00294	•	0.51995	0.00782	0.13276	0.02604	0.30235	0.00970	0.70904	0.00079	0.11682	0.00020	0.80546	0.00000	0.09425	0.00861	0.02663	0.51094	0.05247	0.00158	0.10197	0.00000	0.00000
ENG	0.01388	0.01163	0.01636	0.01705	0.00266	-0.00044	•	0.01040	0.15454	0.01594	0.00545	0.04673	0.00000	0.01129	0.00119	0.00465	0.24720	0.00000	0.02594	0.00158	0.00059	0.53628	0.05643	0.00475	0.00089	0.00000	0.00000
ARA	-0.00461	0.00328	-0.00581	0.01807	0.00351	0.02658	0.02362	•	0.15444	0.53757	0.10702	0.77081	0.01445	0.26641	0.12256	0.21790	0.03257	0.00307	0.33294	0.75824	0.74567	0.05594	0.34838	0.04455	0.12830	0.00000	0.00000
AST	-0.00951	-0.00494	-0.00099	0.01610	-0.00524	0.00686	0.00490	0.00395	•	0.61974	0.16236	0.35036	0.03802	0.01445	0.57351	0.00634	0.06564	0.02069	0.88823	0.23720	0.40154	0.39927	0.10870	0.10049	0.12830	0.00000	0.00000
BAR	-0.01297	-0.00401	-0.00919	0.01010	-0.00454	0.02151	0.02293	-0.00668	-0.00577	•	0.20354	0.73725	0.00000	0.22532	0.53391	0.21107	0.00040	0.00000	0.55965	0.54371	0.66340	0.21533	0.12623	0.24691	0.21523	0.00000	0.00000
GAL	0.00292	-0.00576	0.00496	-0.00016	-0.00391	0.00648	0.02669	0.01457	0.00826	0.00682	•	0.00000	0.08484	0.02723	0.10722	0.13989	0.12652	0.00842	0.00188	0.39590	0.11464	0.17662	0.32007	0.10623	0.34541	0.00059	0.00000
MAD	-0.00683	0.00078	-0.00728	0.01482	0.00033	0.02470	0.02292	-0.00568	0.00052	-0.00849	0.01103	•	0.00000	0.28997	0.23404	0.25839	0.02891	0.00000	0.37452	0.79418	0.98327	0.06772	0.20384	0.32135	0.07742	0.00000	0.00000
GRM	0.03071	0.01230	0.02122	0.01107	0.00458	-0.00136	0.05746	0.03010	0.01174	0.03633	0.02644	0.03078	•	0.00812	0.00000	0.00168	0.00010	0.00000	0.00000	0.00000	0.00000	0.00743	0.00010	0.00307	0.17800	0.00000	0.00000
GRE	0.00610	0.01229	0.00071	0.02167	0.01259	0.03775	0.03384	0.00093	0.01888	0.00290	0.01818	0.03695	0.03633	•	0.03633	0.80942	0.01277	0.00089	0.10801	0.53450	0.35086	0.01564	0.10900	0.59687	0.02455	0.00000	0.00000
BRE	-0.00806	-0.00114	0.00297	0.01631	-0.00325	0.00902	0.00730	0.00860	-0.00367	-0.00189	0.00954	0.00577	0.02800	0.02162	•	0.01505	0.00000	0.00000	0.53341	0.20770	0.29472	0.48807	0.20414	0.05990	0.02515	0.00000	0.00000
PUG	0.01023	0.01764	0.00327	0.02704	0.01846	0.05008	0.04484	0.00317	0.02594	0.00586	0.02105	0.00311	0.04971	-0.00477	0.02853	•	0.00604	0.00020	0.07514	0.43817	0.22968	0.00604	0.11801	0.46104	0.00921	0.00000	0.00000
NTH	0.01673	0.00940	0.01570	0.01340	0.00191	-0.00381	0.00215	0.02406	0.00623	0.02434	0.02379	0.02259	0.03075	0.03242	0.01236	0.04289	•	0.11090	0.00000	0.00000	0.00000	0.43431	0.01257	0.06445	0.00475	0.00000	0.00000
POL	0.05359	0.02833	0.03119	0.01002	0.02029	0.02150	0.14143	0.03968	0.02210	0.05428	0.03436	0.04285	0.03981	0.04416	0.05142	0.05888	0.11090	•	0.00000	0.00000	0.00000	0.00000	0.10289	0.00000	0.00475	0.00000	0.00475
NPT	-0.01222	-0.00903	-0.00393	0.00770	-0.00979	0.00426	0.00694	0.00166	-0.00706	-0.00760	0.00262	-0.00122	0.01860	0.01272	-0.00762	0.01805	0.00942	0.03907	•	0.30977	0.30809	0.54173	0.29680	0.19256	0.44639	0.00000	0.00000
CPT	-0.00546	-0.00283	-0.00880	0.01309	0.00223	0.03527	0.03524	-0.00805	-0.00384	-0.00859	0.01096	-0.00880	0.04286	-0.00476	0.01164	-0.00337	0.00129	0.05892	0.00137	•	0.79467	0.04118	0.19107	0.56410	0.15305	0.00000	0.00000
SPT	-0.01119	-0.00353	-0.00996	0.00996	-0.00385	0.02333	0.02589	-0.00752	-0.00433	-0.01214	0.00714	-0.00934	0.03644	0.00044	0.00149	0.00305	0.02596	0.05342	-0.00601	-0.01021	•	0.08128	0.09771	0.41649	0.26987	0.00000	0.00000
SWT	-0.00315	-0.00451	0.00636	0.01217	-0.00718	-0.00286	-0.00061	0.01398	-0.00105	0.00544	0.00368	0.01108	0.00542	0.02610	-0.00135	0.03496	-0.00053	0.02879	-0.00538	0.01791	0.00729	•	0.16602	0.03752	0.24691	0.00000	0.00000
IRE	0.00471	0.01168	0.00260	0.00603	0.00491	0.04075	0.05335	0.01224	0.00721	0.00526	0.00883	0.08575	0.01679	0.01959	0.02133	0.04951	0.10289	0.00958	0.01051	0.00729	0.02756	0.02756	•	0.22087	0.00366	0.00000	0.00000
MAC	0.00323	0.00848	-0.00274	0.01442	0.00933	0.04375	0.04373	-0.00229	0.01411	-0.00052	0.01672	-0.00321	0.04744	-0.00478	0.02293	-0.00361	0.04024	0.05896	0.00995	-0.00814	-0.00383	0.02733	0.01717	•	0.05445	0.00000	0.00000
HUNG	0.00044	-0.00694	0.00426	0.00037	-0.00494	0.00487	0.01013	0.01236	0.00626	0.00558	-0.00057	0.01003	0.00301	0.01916	0.00977	0.02609	0.00670	0.00693	0.00004	0.00960	0.00467	0.00234	0.00907	0.01265	•	0.00000	0.00000
LEB	0.11554	0.12536	0.10995	0.13671	0.12524	0.16157	0.14871	0.11802	0.15264	0.11175	0.12625	0.11138	0.15334	0.09211	0.13879	0.08317	0.14560	0.16080	0.12608	0.09781	0.10666	0.14715	0.12021	0.09548	0.13931	•	0.00000
FIN	0.08624	0.09718	0.05264	0.05012	0.05222	0.05414	0.22984	0.05260	0.03058	0.07717	0.09390	0.05720	0.24430	0.05911	0.06747	0.07872	0.21805	0.28812	0.07128	0.08215	0.07992	0.05655	0.22015	0.09209	0.05169	0.18071	•

Table S14: Pairwise F_{ST} genetic distances and, respectively, p-values of present study Colombian samples and other European populations. European populations are represented in green; Colombia regions in blue and two populations from the Middle Eastern (Turkey and Lebanon) are coloured in light orange. F_{ST} genetic distances are below the diagonal and the respectively p-values above it. The non-significant values are highlighted in red ($p \geq 0.00024$) (Bonferroni-p-corrected). (Bonferroni-p-corrected). AND-Andes; CAR-Caribe; PT-Portugal; CAT-Catalunya, Spain; GAL-Galicia, Spain; CLM-Castilla-La Mancha; Spain; ANDA-Andalusia, Spain; BAS-Basque Country, Spain; NIT-North, Italy; CIT-Center, Italy; SDIT-Sardinia, Italy; GRM-Germany; POL-Poland; SLV-Slovakia; DNK-Denmark; CRO-Croatia; SRB-Serbia; FRA-France; TUR-Turkey; LEB-Lebanon.

	AND	CAR	PT	CAT	GAL	CLM	ANDA	BAS	NIT	CIT	SIT	SDIT	GRM	POL	SLV	DNK	CRO	SRB	FRA	TUR	LEB
AND	-	0.00089	0.24374	0.07326	0.16790	0.42412	0.88408	0.00030	0.02574	0.01000	0.00000	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.83311	0.00000	0.00000
CAR	0.05903	-	0.00030	0.00000	0.01388	0.00000	0.00010	0.00000	0.01802	0.00178	0.00000	0.00089	0.44857	0.00000	0.00000	0.01218	0.00000	0.00000	0.00000	0.00000	0.00000
PT	0.00243	0.04353	-	0.00030	0.87370	0.04010	0.22938	0.00000	0.30267	0.00119	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.03976	0.00000	0.00000
CAT	0.01755	0.15110	0.03951	-	0.00399	0.43243	0.00822	0.00916	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00257	0.00000	0.00000
GAL	0.00735	0.02840	-0.00252	0.05585	-	0.03763	0.19643	0.00000	0.13653	0.02148	0.00000	0.00000	0.00138	0.00000	0.00000	0.00000	0.00000	0.00000	0.27987	0.00000	0.00000
CLM	-0.00143	0.10121	0.01338	-0.00176	0.02213	-	0.16493	0.00130	0.00198	0.00119	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.11278	0.00000	0.00000
ANDA	-0.00497	0.06314	0.00119	0.02791	0.00377	0.00613	-	0.00000	0.00208	0.00148	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.19414	0.00000	0.00000
BAS	0.07614	0.24298	0.08333	0.01604	0.12991	0.05911	0.08107	-	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
NIT	0.01848	0.01805	0.00880	0.07451	0.00490	0.03764	0.02046	0.13699	-	0.31552	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CIT	0.02750	0.03026	0.01411	0.08997	0.01597	0.04610	0.02524	0.16693	0.00178	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
SIT	0.10666	0.05314	0.08456	0.18352	0.07428	0.13106	0.10233	0.26092	0.04482	0.02740	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
SDIT	0.20379	0.05395	0.17124	0.31722	0.14437	0.24999	0.20591	0.42916	0.12276	0.12642	0.08680	-	0.00000	0.00000	0.00000	0.00096	0.00000	0.00000	0.00000	0.00000	0.00000
GRM	0.05332	-0.00098	0.04277	0.11918	0.03175	0.08400	0.05608	0.16717	0.02643	0.03731	0.07160	0.07930	-	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
POL	0.36805	0.25292	0.32691	0.44130	0.33470	0.39644	0.35978	0.49346	0.29953	0.28584	0.23740	0.27645	0.21822	-	0.00168	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
SLV	0.26744	0.13764	0.24195	0.36266	0.22628	0.30267	0.26841	0.44900	0.19763	0.18490	0.14038	0.14832	0.12293	0.01973	-	0.00000	0.01988	0.00000	0.00000	0.00000	0.00000
DNK	0.14707	0.02582	0.12858	0.23444	0.10870	0.19043	0.15015	0.29811	0.09906	0.11173	0.11589	0.04317	0.02542	0.19779	0.09717	-	0.00020	0.00000	0.00000	0.00000	0.00000
CRO	0.27224	0.10766	0.24128	0.38254	0.22012	0.31878	0.27386	0.47869	0.19279	0.19116	0.14719	0.09094	0.10231	0.08072	0.01828	0.04725	-	0.00138	0.00000	0.00000	0.00000
SRB	0.24514	0.08112	0.22285	0.35899	0.18837	0.29509	0.25326	0.45668	0.15711	0.15916	0.09791	0.05207	0.10251	0.17853	0.08005	0.06551	0.03714	-	0.00000	0.00000	0.00000
FRA	-0.00325	0.04894	0.00258	0.02727	0.00145	0.00694	0.00175	0.06567	0.01575	0.02836	0.11145	0.19230	0.04336	0.34682	0.26209	0.13279	0.25941	0.24572	-	0.00000	0.00000
TUR	0.17148	0.10986	0.15350	0.25614	0.14454	0.20294	0.16656	0.32652	0.10876	0.07915	0.02061	0.11475	0.12777	0.23632	0.15153	0.15720	0.16593	0.11600	0.18418	-	0.00000
LEB	0.25341	0.18469	0.22911	0.33891	0.22411	0.28973	0.24376	0.39890	0.18168	0.14929	0.06607	0.17707	0.20666	0.29909	0.22360	0.23360	0.23474	0.16089	0.26287	0.01821	-

Table S15: Pairwise F_{ST} genetic distances and, respectively, p-values of the present study Colombian R-M269 samples and European R-M269 populations. European populations are represented in green; Colombia regions in blue and one population from Turkey is representing Middle Eastern coloured in light orange. R_{ST} genetic distances are below the diagonal and the respectively p-values above it. The non-significant values are highlighted in red ($p\geq0.00022$) (Bonferroni-p-corrected).Population names are represented in the following order AND-Andes; CAR-Caribe; PT-Portugal; CAT-Cantabria; CL-Castile and Leon, Spain; VAL-Valencia, Spain; ANDA-Andalusia, Spain; NIT-Northern, Italy; CIT-Central, Italy; SIT-Southern, Italy; FRA-France; IRL-Ireland; WAL-Wales; ENG-England; SCA-Scandinavian regions; NTH-Netherlands; GRM-Germany; POL-Poland; RUS-Russia; BALC-Balkan Peninsula; TUR-Turkey.

	AND	CAR	PT	CAT	CL	VAL	ANDA	NIT	CIT	SIT	SDIT	FRA	IRL	WAL	ENG	SCA	NTH	GRM	POL	RUS	BALC	TUR
AND	-	0.26502	0.21513	0.90852	0.83853	0.36858	0.00851	0.00000	0.00000	0.00000	0.00000	0.00079	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CAR	0.00877	-	0.20364	0.21285	0.04307	0.08356	0.00000	0.00000	0.00000	0.00000	0.00000	0.02594	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PT	0.00584	0.00735	-	0.03158	0.02515	0.12573	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	0.00000	0.00000
CAT	-0.01562	0.01065	0.02029	-	0.61063	0.13127	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	0.00000	0.00000
CL	-0.01544	0.04725	0.02678	-0.00725	-	0.17929	0.08455	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	0.00000
VAL	0.00029	0.02781	0.00917	0.01526	0.01255	-	0.00089	0.00000	0.00000	0.00000	0.00000	0.00109	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ANDA	0.06123	0.18016	0.09654	0.05218	0.02286	0.09836	-	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	0.00000
NIT	0.31587	0.25079	0.24721	0.34534	0.36571	0.27826	0.47595	-	0.21978	0.00000	0.21028	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CIT	0.41345	0.34505	0.32526	0.44388	0.46523	0.37150	0.58404	0.00341	-	0.00000	0.28354	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
SIT	0.27775	0.21668	0.21005	0.31311	0.32710	0.25957	0.45096	0.07960	0.10241	-	0.19048	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00149	0.00000	0.00000
SDIT	0.41645	0.32143	0.29178	0.45833	0.49282	0.36300	0.69559	0.01600	0.01135	0.02094	-	0.00030	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00149	0.00059	0.00000
FRA	0.07226	0.03454	0.03978	0.09234	0.10906	0.05885	0.19450	0.11546	0.18009	0.12299	0.16922	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
IRL	0.55795	0.45741	0.47319	0.55424	0.59317	0.56538	0.65837	0.50002	0.54875	0.43364	0.55917	0.39003	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
WAL	0.31479	0.20165	0.25725	0.32823	0.36887	0.32609	0.49140	0.28486	0.34299	0.20855	0.32269	0.17446	0.07421	-	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ENG	0.21059	0.15236	0.16596	0.23539	0.25268	0.21745	0.34787	0.17349	0.21925	0.10019	0.20712	0.08823	0.26794	0.08531	-	0.10375	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
SCA	0.28589	0.21906	0.22942	0.32574	0.33885	0.28246	0.48498	0.22431	0.27258	0.12542	0.24537	0.13452	0.32821	0.11514	0.00870	-	0.00000	0.00079	0.00198	0.00158	0.00000	0.00000
NTH	0.33677	0.31423	0.28562	0.36646	0.36925	0.34388	0.44482	0.28672	0.32370	0.20353	0.33725	0.21019	0.42731	0.27464	0.07134	0.06468	-	0.00000	0.00000	0.00000	0.00000	0.00000
GRM	0.28850	0.25257	0.23533	0.32563	0.32968	0.27921	0.43242	0.15743	0.18968	0.09686	0.19105	0.13394	0.44475	0.23130	0.04765	0.04575	0.03044	-	0.00327	0.00000	0.00000	0.00000
POL	0.31626	0.26435	0.24808	0.36249	0.36659	0.29223	0.51100	0.15787	0.18588	0.04617	0.12838	0.15199	0.46147	0.22782	0.07252	0.04899	0.10999	0.03433	-	0.00644	0.00050	0.00000
RUS	0.32883	0.27808	0.25670	0.37300	0.37648	0.29120	0.53259	0.26361	0.30910	0.15904	0.25796	0.17933	0.46523	0.24278	0.11889	0.06478	0.19032	0.12136	0.04523	-	0.00287	0.00000
BALC	0.33986	0.29325	0.27275	0.37140	0.37662	0.29623	0.48452	0.20527	0.22884	0.10328	0.14823	0.20098	0.46886	0.27421	0.18099	0.15087	0.28693	0.17610	0.05559	0.04527	-	0.00693
TUR	0.45923	0.42178	0.37268	0.48426	0.48887	0.41350	0.58499	0.35459	0.38369	0.22293	0.30251	0.31835	0.54591	0.38623	0.29069	0.26623	0.38465	0.29809	0.15811	0.10626	0.02986	-