Characterisation of human male lineages from Ecuador

Mariana Borges de Pinho Seguro Pereira
Mestrado em Genética Forense
Departamento de Biologia
2020

Orientador
Verónica Gomes, PhD, Instituto de Investigação e Inovação em Saúde (I3S)/ Instituto de Patologia e Imunologia Molecular (IPATIMUP)

Coorientador
Leonor Gusmão, PhD, Universidade do Estado do Rio de Janeiro (UERJ)

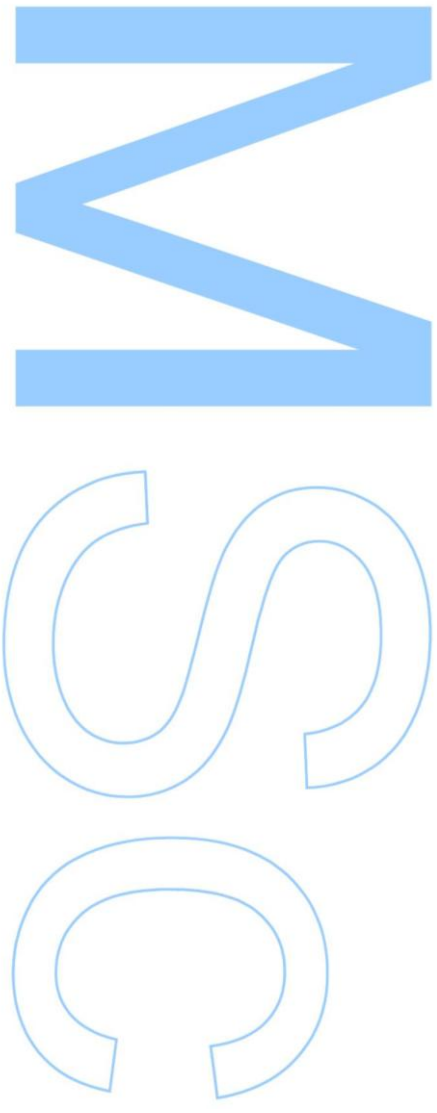
U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

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

O Presidente do Júri,

Porto, _____/_____/_____



Agradecimentos

À Verónica Gomes, minha orientadora, por todo o tempo que dedicou a ajudar-me a realizar esta tese. Por ter estado sempre disponível, por toda a ajuda desde o inicio do trabalho laboratorial até à fase final de análise de dados e escrita da tese e ainda, por todo o conhecimento que me transmitiu.

À Leonor Gusmão, por ter encontrado sempre tempo para me ajudar, principalmente na fase de análise dos resultados e revisão deste documento. Gostaria também de agradecer por tudo o que me ensinou.

Ao Germán Burgos, por me ter cedido as amostras, sem as quais este estudo não poderia ter sido realizado, e por me ter ajudado sempre que solicitei a sua ajuda.

A todos os elementos que constituíram o grupo de Genética Populacional do IPATIMUP durante este ano lectivo, por se terem disponibilizado a ajudar-me sempre que precisei e por me terem proporcionado um bom ambiente de trabalho.

Muito obrigada.

Resumo

A América foi o último continente a ser colonizado pelo Homem moderno que veio da Ásia. Através de uma ponte terrestre entrou na América do Norte e por fim chegou à América do Sul. Milhares de anos mais tarde, no final do século XV, começou o interesse Europeu pelo Novo Mundo. Espanhóis e escravos Africanos chegaram no século XVI a um pequeno país localizado na região Noroeste da América do Sul, designado de Equador. Os eventos históricos que ocorreram nesse país tornaram possível a mistura de diferentes grupos populacionais, o que levou à formação dos três principais grupos étnicos que atualmente habitam no Equador: Mestiços, Afro-Equatorianos e Nativos Americanos.

O cromossoma Y não sofre recombinação na maioria do seu comprimento e, desta forma, quase toda a informação genética deste cromossoma é transmitida de pais para filhos sem ser alterada, exceto se ocorrerem mutações. Assim, através da análise de polimorfismos encontrados na região especifica do sexo masculino do cromossoma Y é possível reconstruir a história das linhagens paternas.

Com o objetivo de investigar se dois grupos étnicos do Equador (Mestiços e Afro-Equatorianos) têm semelhanças genéticas e como ocorreu o processo de miscigenação, foram caracterizadas as linhagens masculinas de 149 Mestiços e 150 Afro-Equatorianos. Estas linhagens foram caracterizadas através da análise de 23 Y-STRs e de 59 Y-SNPs.

Os resultados mostraram que os dois grupos apresentam uma elevada diversidade haplotípica, sendo a mais elevada encontrada no grupo Mestiço. Foram encontrados 147 haplótipos únicos nos Mestiços e 121 nos Afro-Equatorianos. Relativamente aos dados dos Y-SNPs, foram identificados 32 haplogrupos diferentes na totalidade das amostras (24 haplogrupos em cada grupo étnico). Nas amostras de Mestiços, o haplogrupo com maior frequência foi o R (38,93%), seguido do Q (24,49%), enquanto que na população Afro-Equatoriana a linhagem mais frequente foi a E (76,67%). Relativamente às distancias genéticas, foram obtidos valores estatisticamente significativos (R_{ST} e F_{ST}) entre os dois grupos.

Os resultados obtidos também mostraram que as linhagens encontradas nos Mestiços são 70,47%, 23,49% e 6,04% de origem Europeia, Nativa Americana e Africana, respetivamente. Por outro lado, a população Afro-Equatoriana é caracterizada

principalmente por linhagens Africanas (76,00%), seguidas de linhagens Europeias (22,67%) e Nativas Americanas (1,33%).

Os resultados obtidos através da comparação de populações da América Latina mostraram que a população Mestiça do Equador é idêntica a populações de Nicarágua e Argentina. Considerando a ancestralidade Europeia e Africana, o presente estudo mostra que as linhagens Africanas encontradas no Equador têm múltiplas origens, principalmente de África Ocidental e Central e que a maioria das linhagens Europeias vieram da Península Ibérica. Assim, concluiu-se que os resultados obtidos são consistentes com os registos históricos relativamente à origem de populações Europeias e Africanas que migraram para o Equador.

Abstract

America was the last continent to be colonized by modern humans that came from Asia and crossed a land bridge to enter in North America and lastly get to South America. Thousands of years later, in the late 15th century, the European interest in the New World began. In the 16th century, Spaniards and Africans slaves arrived in Ecuador, which is a small country located in the Northwest region of South America. The historical events that occurred in Ecuador made possible the admixture between different populations resulting in three main ethnic groups that currently inhabit this country: Mestizos, Afro-Ecuadorians and Native Americans.

Y chromosome is devoid of recombination in most of its length. Consequently, most of the genetic information on the Y chromosome is transmitted from father to son without changing, except if mutation occurs. Therefore, it is possible to reconstruct the history of paternal lineages through the analysis of polymorphisms along the male-specific region of the Y chromosome.

To investigate if Mestizos and Afro-Ecuadorian ethnic groups from Ecuador share genetic similarities and how the admixture process occurred in these populations, the male lineages of 149 Mestizos and 150 Afro-Ecuadorians were characterised through the analysis of 23 Y-STRs and 59 Y-SNPs.

High haplotype diversities were found in the two groups (being higher in Mestizos), with 147 unique haplotypes in the Mestizos and 121 in Afro-Ecuadorians. Concerning the Y-SNPs, 32 different haplogroups were identified in the total sample (24 in each group). In the Mestizo samples, the haplogroup R (38.93%) was the most frequent, followed by haplogroup Q (24.49%), while in Afro-Ecuadorian population the most frequent lineage was E (76.67%). Significant R_{ST} and F_{ST} pairwise genetic distances were obtained between them.

The male gene pool of the Mestizo population is composed by 70.47%, 23.49% and 6.04% of European, Native American and African lineages, respectively. On the other hand, the Afro-Ecuadorian population is mainly characterised by African lineages (76.00%), followed by European (22.67%) and Native American (1.33%) lineages.

The results of a population comparison analysis between Latin American countries showed that Mestizo population from Ecuador is similar to populations from Nicaragua and Argentina. Concerning the European and African ancestries in both groups, the

results were consistent with historical records, showing that the African lineages found in Ecuador have multiple origins, mainly from Western and Central Africa, and most of the European lineages came from the Iberian Peninsula.

Keywords

Population genetics, Y chromosome, Y-STRs, Y-SNPs, Paternal lineages, South America, Ecuador, Mestizo population, Afro-descendant population, Population ancestry.

Table of Contents

Agradecimentos II
Resumo III
Abstract V
Keywords VII
List of Tables X
List of Figures XI
Abbreviations XV
Introduction 1
Population and forensic genetics 1
Uniparental markers 3
Y chromosome 4
Origin and structure 4
Y chromosome polymorphisms 6
Y-STRs 6
Y-SNPs 7
Colonization of the Americas 12
First studies 12
"Out of Beringia" model 13
Entrance in the Americas 14
Other hypotheses of the Native Americans origin 14
Genetic evidence 15
Mitochondrial DNA data 15
Y chromosomal data 16
Ecuador 17
Genetic diversity of Ecuadorian population 20
Materials and Methods 22
Population samples and DNA extraction 22
DNA analysis 22
Y-STR typing 22
Y-SNP typing 22
Data analysis 25
Results and Discussion 26

Ecuadorian genetic diversity 26
Y-STR data 26
Y-SNP data 29
Haplogroup B 30
Haplogroup E 30
Haplogroups G, I and J 32
Haplogroups K, L and T 33
Haplogroup Q 33
Haplogroup R 34
Ecuadorian origins 34
Population comparisons 36
Latin America 36
Y-STRs 36
Y-SNPs 38
Native American lineages 40
African ancestry 45
Haplogroup E 48
European ancestry 53
Haplogroup R 56
Conclusion 60
References 62
Appendix 75

List of Tables

Table 1: Selective list of Y-haplogroups, as well as their geographical distribution. [Adapted from Kayser (103)].....9

List of Figures

Figure 1: Schematic representation of inheritance patterns of both recombining autosomal markers and non-recombining genetic markers (Y chromosome and mtDNA). [Adapted from Butler (34)] 3

Figure 2: Schematic representation of the Y chromosome structure (a) and a more detailed representation of the male-specific region (MSY), focusing on the euchromatic region (b). Palindromic sequences are represented in dark blue triangles and the position of protein-coding genes are indicated by vertical lines. [Adapted from Hughes & Rozen (94)] 5

Figure 3: Map showing the frequencies of 20 major Y chromosome haplogroups in worldwide map. [Adapted from Karafet et al. (101)] 8

Figure 4: Schematic representation of the allelic discrimination through SNaPshot technology. The steps represented in this figure are: (a) annealing of the primers; (b) extension of the ddNTPs and (c) result obtained after a capillary electrophoresis. Note that the middle SNP has two alleles whereas the other SNPs have just one allele. [Adapted from González dos Santos (84)] 11

Figure 5: Representation of the peopling of the Americas according to Tamm and colleagues' proposal. Colored arrows indicate the approximated timing and the direction of the maternal gene flow events. [Adapted from Tamm et al. (170)] 13

Figure 6: Map showing the location of Ecuador (in red). 17

Figure 7: Map representing the main regions from where the African slaves were brought to America. [Adapted from Gambéla (67)] 18

Figure 8: MDS plot of the R_{ST} genetic distances between the three Ecuadorian ethnic groups analyzed, based on information from 10 Y-STR markers (Stress=0.0019540). The present study populations are highlighted with a black circle (green - Afro-Ecuadorians; blue - Mestizos and red - Native Americans).Population codes can be seen in Appendix Table S5. 28

Figure 9: Frequency of the Y-SNP haplogroups found in the two ethnic groups in study [Mestizos (a) and Afro-Ecuadorians (b)]. The arrows indicate the first haplogroup mentioned and the other ones follow the clockwise direction. Note that the haplogroups E, R and Q are represented in tones of blue, green, and orange and brown, respectively. 30

Figure 10: Network constructed with 225 European (Spain, Portugal and Italy), 17 sub-Saharan Africa (Gabon, Guinea-Bissau, Equatorial Guinea, Madagascar and Uganda) and 7 Ecuadorian samples belonging to the sub-haplogroup E-M78..... 32

Figure 11: Geographic origin of Ecuadorian lineages found in the two ethnic groups in the present study, according to Y-SNP haplogroup geographic distribution and historical facts related to the population migrations to Ecuador. 34

Figure 12: MDS plot of the R_{ST} genetic distances between the two Ecuadorian ethnic groups and other Latin American populations, based on from 15 Y-STR markers (Stress=0.0150922). The present study populations are highlighted with a black circle (Mestizos are under the Argentinean population) (green - Afro-Ecuadorians; dark green - African population blue - Mestizos/Admixed populations and red - Native Americans). (COL- Colombia, PER- Peru, BRA- Brazil, BOL- Bolivia, CHI- Chile, ARG- Argentina, PAN- Panama, CRI- Costa Rica, NIC- Nicaragua and MEX- Mexico, more details about the population codes can be seen in Appendix Tables S5-S7). 38

Figure 13: MDS plot of the F_{ST} genetic distances between the Ecuadorian population and other Latin American populations, based on Y haplogroup information (Stress=0.0567491). The present study populations are highlighted with a black circle (green - Afro-Ecuadorians; dark green - African population blue - Mestizos/Admixed populations; dark blue - European population and red - Native Americans). (COL- Colombia, PER- Peru, BRA- Brazil, BOL- Bolivia, ARG- Argentina and NIC- Nicaragua; more details about the population codes can be seen in Appendix Tables S5-S7). 39

Figure 14: MDS plot of the R_{ST} (a) and F_{ST} (b) genetic distances between the Ecuadorian population and other Latin American populations, based on information of 15 Y-STRs [Stress=0.0841538 (a) and 0.0000049 (b)]. The present study population is highlighted with a black circle. (NAM- Native American from Ecuador, COL- Colombia, PER- Peru, BRA- Brazil, BOL- Bolivia and ARG- Argentina; more details about the population codes can be seen in Appendix Table S5). 41

Figure 15: Principal component analysis, focusing on the first and second principal components (a) and on the first and third principal components (b), of Q-M242*(xM346), Q-M346*(xM3), Q-M3*(xM19) and Q-M19 haplogroup frequencies in the present study Ecuadorian population (highlighted with a black circle) and in samples from different South American populations. The first, second and third components represent 51.74%, 27.42% and 20.85% of the total variance, respectively. (NAM- Native American from Ecuador, COL- Colombia, PER- Peru, BRA- Brazil, BOL- Bolivia and ARG- Argentina; more details about the population codes can be seen in Appendix Table S5). 43

Figure 16: Relationships between Y haplotypes within Q-M3 haplogroup based on information of 15 Y-STRs. Samples from present study [Mestizos (n=31) and Afro-Ecuadorians (n=2)] and Quichua (n=54), Waorani (n=66) and Shuar (n=4) Native American populations were analysed. The arrow indicates the sample belonging to the Q-Z19483 sub-haplogroup. 44

Figure 17: MDS plot of the R_{ST} (a) and F_{ST} (b) genetic distances between the Ecuadorian population and African populations, based on 10 Y-STR loci within African lineages [Stress=0.0662166 (a) and 0.0903862 (b)]. The present study population is highlighted with a black circle. (GAB-Guinea Bissau, CIV-Ivory Coast, BEN-Benin, CMR-Cameroon, GNQ-Equatorial Guinea, GAB-Gabon, AGO-Angola and MOZ-

Mozambique; more details about the population codes can be seen in Appendix Table S6). 46

Figure 18: MDS plot of the F_{ST} genetic distances between the Ecuadorian population and African populations, based on African Y haplogroups information (Stress=0.0126493). The present study population is highlighted with a black circle. (GAB-Guinea Bissau, CIV-Ivory Coast, BEN-Benin, CMR-Cameroon, GNQ-Equatorial Guinea, GAB-Gabon and MOZ-Mozambique; more details about the population codes can be seen in Appendix Table S6). 47

Figure 19: Principal component analysis of A, B, E-M96*(xM33, M2, M35, M75), E-M33, E-M2, E-M35*(xM78, M81, M123) and E-M75 haplogroup frequencies in the present study (highlighted with a black circle) and in populations from different African countries (GAB-Guinea Bissau, CIV-Ivory Coast, BEN-Benin, CMR-Cameroon, GNQ-Equatorial Guinea, GAB-Gabon and MOZ-Mozambique; more details about the population codes can be seen in Appendix Table S6). 48

Figure 20: E-M2* median-joining network based on 10 Y-STR information. Samples from present study [Afro-Ecuadorians (n=28)], Guinea-Bissau (n=120), Ivory Coast (n=20), Benin (n=8), Equatorial Guinea (n=37) and Mozambique (n=2) were analysed. 49

Figure 21: E-M191 phylogenetic network based on 10 Y-STRs information. Samples from present study [Mestizos (n=1) and Afro-Ecuadorians (n=44)], Guinea-Bissau (n=3), Ivory Coast (n=11). Benin (n=44), Equatorial Guinea (n=39) and Mozambique (n=19) were analysed. 50

Figure 22: Relationships between Y haplotypes within E-U209 haplogroup based on information of 10 Y-STRs. Samples from present study [Mestizos (n=5) and Afro-Ecuadorians (n=33)], Ivory Coast (n=43). Benin (n=43) and Mozambique (n=35) were analysed. 51

Figure 23: E-M75 Phylogenetic network (constructed based on 10 Y-STR loci). Samples from present study [Mestizos (n=1) and Afro-Ecuadorians (n=5)], Guinea-Bissaou (n=1), Ivory Coast (n=1). Benin (n=1), Camaroon (n=1), Equatorial Guinea (n=2), Gabon (n=49) and Mozambique (n=7) were analysed. 52

Figure 24: MDS plot of the R_{ST} genetic distances between the Ecuadorian and European populations, based on 15 Y-STR markers within European lineages (Stress=0.00369814). The present study population is highlighted with a black circle. (PRT- Portugal, ESP- Spain, FRA- France, ITA- Italy, DEU- Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix Table S7)..... 54

Figure 25: MDS plot of the F_{ST} genetic distances between the Ecuadorian and European populations, based on information from European Y-SNP haplogroups (Stress=0.0784008). The present study population is highlighted with a black circle. (PRT- Portugal, ESP- Spain, GBR- United Kingdom, FRA- France, ITA- Italy, DEU-

Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix Table S7). 55

Figure 26: Principal component analysis of E (only European lineages), G, I, J, K and R (except R-V88) haplogroup frequencies in the present study Ecuadorian population (highlighted with a black circle) and in samples from different European populations (PRT- Portugal, ESP- Spain, GBR- United Kingdom, FRA- France, ITA- Italy, DEU- Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix Table S7). Samples from the same country and study were pooled when no statistically significant differences were found between them. 56

Figure 27: MDS plot of the F_{ST} genetic distances between the Ecuadorian and European populations, based on information of haplogroup R (Stress=0.1116117). The present study population is highlighted with a black circle. (PRT- Portugal, ESP- Spain, IRL- Ireland, SCT- Scotland, ENG- England, FRA- France, ITA- Italy, NLD- Netherlands, DEU- Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix Table S7). 58

Figure 28: Principal component analysis of R-M269*(xU106, S116), R-U106, R-S116*(xU152, M529), R-U152 and R-M529 haplogroup frequencies in Ecuadorian population from the present study (highlighted with a black circle) and in different European populations (PRT- Portugal, ESP- Spain, IRL- Ireland, SCT- Scotland, ENG- England, FRA- France, ITA- Italy, NLD- Netherlands, DEU- Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix TableS7) 59

Abbreviations

AMOVA	Analysis of Molecular Variance
ANC-A	"Southern Native American" or "Ancestral A"
ANC-B	"Northern Native American" or "Ancestral B"
ddNTP	Dideoxynucleotide
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotides
Indel	Insertion/deletion
ISOGG	International Society of Genetic Genealogy
kya	Thousand years ago
LGM	Last Glacial Maximum
Mb	Million base pairs
MDS	Multi-Dimensional Scaling
MSY	Male-specific region of the Y chromosome
mtDNA	Mitochondrial DNA
NRY	Non-recombining region of the Y chromosome
PAR	Pseudo-autosomal region
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
RM	Rapidly Mutating
SBE	Single Base Extension
SMM	"Stepwise Mutation Model"
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeat
YAP	*Alu* element insertion
YCC	Y Chromosome Consortium
YHRD	Y Chromosome Haplotype Reference Database

Introduction

Population and forensic genetics

Homo sapiens is a relatively recent species that is characterised by a low genetic variation between individuals. Indeed, genetic differences between a randomly chosen pair of humans is approximately 1 per 1000 base pairs, which is a low proportion compared to other species. For example, two randomly selected chimpanzees (our closest living species) diverge from each other in approximately 1 difference in every 500 nucleotides. Considering the genetic diversity of human populations, differences between populations are lower than those found between individuals belonging to the same population (152).

The study of the genetic variation among populations is one of the main aims of population genetics. This scientific field also intends to understand how evolutionary forces such as mutation, migration, selection and genetic drift influence the genetic diversity of populations (116). Therefore, population genetics contributes to a better understanding of populations' history, which is important to other scientific fields, namely anthropological, medical and forensic genetics.

The ABO blood groups were the first markers to be used in a study of human genetic variation, published in 1919 (39). Over the years, new markers and techniques were discovered such as the Polymerase Chain Reaction (PCR) (120). PCR was a revolutionary technique that made possible the analysis of a huge number of copies of specific DNA fragments, even from small samples. The new techniques allowed the introduction of other kinds of markers present in genome such as Short Tandem Repeats (STRs) and Single Nucleotide Polymorphisms (SNPs).

The two previous mentioned polymorphisms[1] are the main source of genetic variation used in forensic genetics (94). In forensic investigations, DNA profiles obtained by genotyping these markers in casework samples can be compared with those from known reference samples, and the probability of match is calculated based on the frequency of the profiles in a reference population. The reference sample can be from a

[1]Polymorphisms or genetic markers are changes in the DNA sequence, where at least two alleles have a frequency higher than 1% in the population (130).

putative father, suspects or victims, for example. However, to report a result, the match between the tested sample and the reference sample is not enough. A random match probability (the probability that a person randomly selected from a population will have the same DNA profile in question) or a likelihood ratio (ratio between the probabilities of the observed genetic results assuming two alternative hypothesis) must be calculated (34). For that, large databases from different populations are needed (34), and these databases are obtained through population genetic studies.

Recombining and non-recombining markers

The molecular markers can be divided into two groups. One includes most of the human genome that is transmitted from both parents to offspring and undergoes recombination. This feature is very important because it allows the reshuffling of genetic information in each generation, which make possible the genetic identification of individuals (130). In this group are included the autosomal and X-chromosomal markers (130). The other cluster is constituted by genome portions, witch do not suffer recombination and are only inherited from one of the parents. The markers present in this group are known as uniparental or lineage markers (Figure 1) and the genetic information obtained from several of these markers is designated as haplotype that behaves as if it were a single allele for each uniparental marker per individual (34). Since they do not recombine, they are transmitted without changes to the next generation, except if mutation occurs (34). Therefore, these markers do not allow individual discrimination but, on the other hand, allow the identification of individuals belonging to the same paternal or maternal lineage (34). In other words, these markers provide genetic information that is shared by different individuals. Mitochondrial DNA is used to define maternal lineages, since it is only transmitted from mother to offspring, whereas the male-specific region of the Y chromosome (MSY) is used to define paternal lineages, since is only transmitted from father to sons (34).

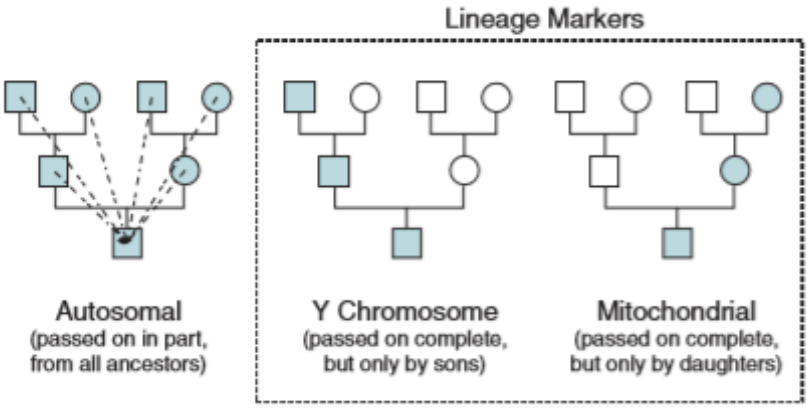


Figure 1: Schematic representation of inheritance patterns of both recombining autosomal markers and non-recombining genetic markers (Y chromosome and mtDNA). [Adapted from Butler (34)]

Uniparental markers

Uniparental markers have specific transmission patterns and also a low effective population size in comparison to autosomes (one fourth of autosomes) (130), which make them more prone to demographic events such as founder effect, bottleneck and genetic drift, and more sensitive to population substructure. The lineages defined by these markers tend to be geographical restricted (102) and, therefore, genetic differences between populations are more accentuated with uniparental markers than with autosomal ones. The aforementioned features make these markers very useful in population genetic studies. Although the autosomal markers are the most used in forensic genetics, the uniparental markers can also be very useful in some specific cases. When the genetic profiles obtained are not complete or the interpretation of the results is unclear (130), using information from different markers might help to overcome the difficulties found in a given case. For instance, since a male genetic profile can be obtained through the analysis of a male relative belonging to the same paternal lineage, the Y chromosome can be very useful in paternity cases when the genetic material of the alleged father is unavailable (130). Y-chromosomal markers can also be a crucial tool in situations of mixed samples, for instance, in cases of sexual assaults when the perpetrator is a male (34, 95). In this situation, the sample is usually constituted by a higher quantity of female than male genetic material. Nevertheless, with the genotyping of Y-chromosomal markers, it is possible to isolate the male genetic profile, helping the identification of the perpetrator (34, 130).

Y chromosome

Origin and structure

The X and Y chromosomes are responsible for sex determination in humans. These two chromosomes evolved from a pair of homologous autosomes, as proposed firstly by Ohno in 1967. The differentiation started at about 180 million years ago with the acquisition of the sex-determining gene in the proto-Y chromosome (44), followed by a series of deletions, translocations and inversions (105). These events conditioned the occurrence of recombination between the two chromosomes during the evolution process (105). Consequently, the Y chromosome got substantially smaller and lost most of its genes (40). By the end of the differentiation process, the Y chromosome in humans was approximately 90 million base pairs (Mb) smaller than the X chromosome (97). Therefore, the Y chromosome (with ~60Mb) is one of the smallest chromosomes that can be found in the human nuclear genome (34).

There are two homologous regions at the tips of both sex chromosomes that are important during the male meiosis, since allow the occurrence of recombination events between the two chromosomes (34, 97, 130). Since these regions behave as autosomes, they are known as pseudo-autosomal regions (PAR1 and PAR2). The bigger one (PAR1) is located on the short arm of the Y chromosome and the PAR2 is located on the tip of the long arm (Figure 2). Together they represent ~3Mb of the chromosome (96). Except for PAR1 and PAR2, the Y chromosome does not recombine in most of its length (approximately 95% of the chromosome) and, therefore, this region is designated as male-specific region of the Y chromosome (MSY), previously known as the non-recombining region of the Y chromosome (NRY) (162). The term NRY is no longer used, since intrachromosomal recombination was discovered in this portion of the Y chromosome (162).

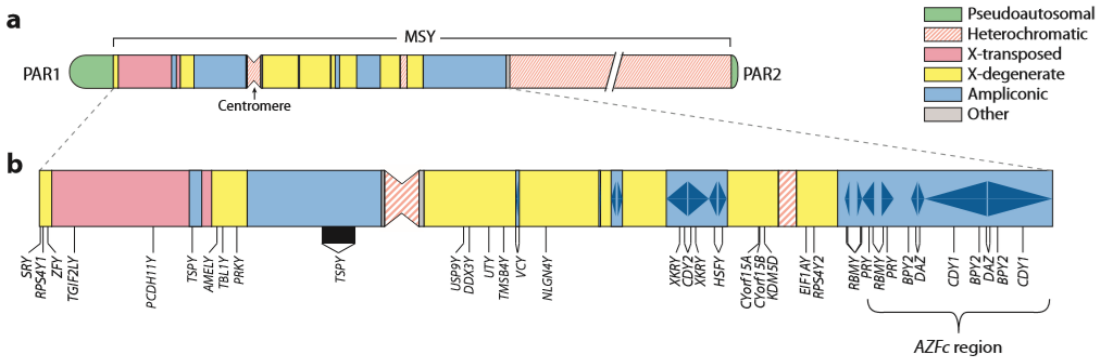


Figure 2: Schematic representation of the Y chromosome structure (a) and a more detailed representation of the male-specific region (MSY), focusing on the euchromatic region (b). Palindromic sequences are represented in dark blue triangles and the position of protein-coding genes are indicated by vertical lines. [Adapted from Hughes & Rozen (94)]

The MSY is constituted by two different regions: the heterochromatic and the euchromatic regions (Figure 2b). The heterochromatin can be mainly found in the long arm of the Y chromosome, being characterized by repetitive sequences and a variable size (34, 130). This type of chromatin can represent more than 50% of the length of the chromosome in some individuals (130). All genes identified in the Y chromosome are located in the euchromatic region (34, 162). The euchromatin has approximately 23Mb and can be divided into three main regions: the X-degenerate (8.6Mb), the X-transposed (3.4Mb) and the ampliconic region (10.2Mb) (162). The first is constituted by sequences that diverged from the ancestral autosomes (97, 162). The second one was recently transposed from the X chromosome and so, have sequences with great homology to this chromosome (162). The ampliconic region is more complex, with some duplicated and palindromic sequences, which have a large number of sequences that present similarities with other sequences in MSY (97, 162). This third region is also characterised by the occurrence of intrachromosomal recombination known as gene conversion (nonreciprocal exchange between two DNA sequences) (162).

As previously mentioned, the male-specificity and the haploidy (only one chromosome per cell) are the special features of the Y chromosome. Being haploid means that there is no recombination with another chromosome during the male meiosis, in most of its length (95). Due to the lack of recombination, the MSY is transmitted unchanged to the next generation as a block, except if any mutation occurs (95, 96). Consequently, a record of Y chromosome past is preserved and it is possible to reconstruct the histories of paternal lineages through the analysis of Y chromosome polymorphisms (95, 96). The Y chromosome genetic studies are useful to complement information from

historical, archaeological and anthropological studies, since the patterns of its diversity are related with male past behaviours (90, 97).

Y chromosome polymorphisms

The most common polymorphisms in the Y chromosome can be divided into two different categories, multi-allelic and bi-allelic markers (34, 130). Multi-allelic markers include the widely used Short Tandem Repeats (STRs) and the bi-allelic markers include Single Nucleotide Polymorphisms (SNPs), *Alu* element insertion (YAP) and other insertion/deletion (indels) markers (34). The first ones are characterised to be very polymorphic and have high mutation rates; whereas bi-allelic markers have low mutation rates and, usually, only have two possible alleles (34, 130).

Y-STRs

STRs or microsatellites are DNA sequences constituted by a variable number of repeated units with 1-6bp in length (87). The first Y-STR, currently designated as DYS19, was discovered in 1992 (149) and since then more than 400 Y-STRs were identified (103). They have high mutation rates (~10^{-4}-10^{-2} mutation per generation) (11, 34, 187) and so, show a high intra-population diversity. The markers with higher mutation rates are known as Rapidly Mutating (RM). These RM markers have a high potential to differentiate between close paternal relatives (14), due to the accumulation of mutations.

"Replication Slippage" is the most accepted mechanism used to explain the emergence of new mutation through the "Stepwise Mutation Model" (SMM) (87). According to the SMM the length of STRs is modified with the addition or removal of one repeat unit. The mutation rate of each allele is related to its size and number of uninterrupted repeats. The larger is the allele with uninterrupted repeats, the higher is the probability of mutation (30, 61).

Information from different Y-STRs can be combined in haplotypes (53), which increases the informative power of these markers. In addition, the analysis of Y-STRs can give information about the age and evolutionary history of a specific haplogroup (group of haplotypes that share a common ancestral, as determined based on less mutable biallelic markers) (53). Higher levels of haplotype diversity will be found in older haplogroups (97).

Due to the Y-STR features, these markers are very useful to disclose recent historical events (150). They can be used in forensic genetics, genealogical research, population and evolutionary genetic studies (87, 55, 102, 103, 148). For these studies, it is important to compile the results obtained in different populations, for which Y chromosome databases were created. The Y Chromosome Haplotype Reference Database (YHRD; https://yhrd.org/) (187) is considered the biggest open access online resource with information on Y-STR haplotypes from different places of the world. The information is organized in different haplotype sets depending on the number of STRs analysed. The smallest haplotype set ("minimal haplotypes") are constituted by 9 Y-STRs, whereas the biggest one have 28 Y-STRs. Most of these markers are included in commercial kits, which is important to the standardization of the obtained haplotype information. Using the same kit, it is possible to genotype the same set of Y-STRs under the same conditions. In addition, an allelic ladder is provided with each kit, which is important for an identical interpretation of the results and to obtain a consistent nomenclature between laboratories (34). Thus, these kits allow a straight comparison between Y-STR data generated by different laboratories, which is important for forensic and population genetic studies.

The PowerPlex® Y23 (Promega) and the AmpFLSTR® Yfiler® Plus (Applied Biosystems) are the commercial kits with more Y-STRs, allowing the simultaneous analysis of 23 and 27 Y-STR *loci*, respectively. Most of the markers included in these kits have mutation rates on the order of 10^{-3}, but the most recent kits also include some RM Y-STRs. In the YHRD information about the ancestry and bio-geographical origin of each profile and some Y-SNP data can also be found.

Y-SNPs

The SNPs are the most abundant class of polymorphisms, representing about 85% of the human genetic variation. These polymorphisms consist in single nucleotide substitutions whose combination is used to define haplogroups (53). These markers are characterised to have low mutation rates (on the order of ~10^{-8} mutations per generation), which means that the occurrence of two or more independent mutations in the same position (recurrent mutations) or back mutations is very unlikely (34, 87, 94, 130). Therefore, these markers can be considered as unique event polymorphisms (34). Due to their low mutation rates, SNPs do not generate new genetic information as fast as STRs and therefore, they have a lower discrimination power than STRs. Nevertheless, haplogroups have a specific geographical distribution that is very useful

for differentiating population groups (Figure 3 and Table 1). Thus, Y chromosome haplogroups allow tracing demographic events, as migrations, the inference of the evolutionary history of a particular population and also the identification of male lineages (a feature shared with Y-STRs) and their most likely geographical origin. In addition, these markers can also be useful in forensic genetics studies, though the Y-STRs are still the most used Y-chromosomal markers in this field. One advantage of the SNPs in forensic studies is that they can be analysed in shorter DNA fragments than STRs. Therefore, Y-SNPs can be a valuable tool when degraded samples or samples with low amounts of DNA are analysed (130). In addition, when the purpose of a study is the inference of paternal bio-geographic ancestry, Y-SNPs are more useful than multi-allelic markers, since they present a higher geographic specificity (102).

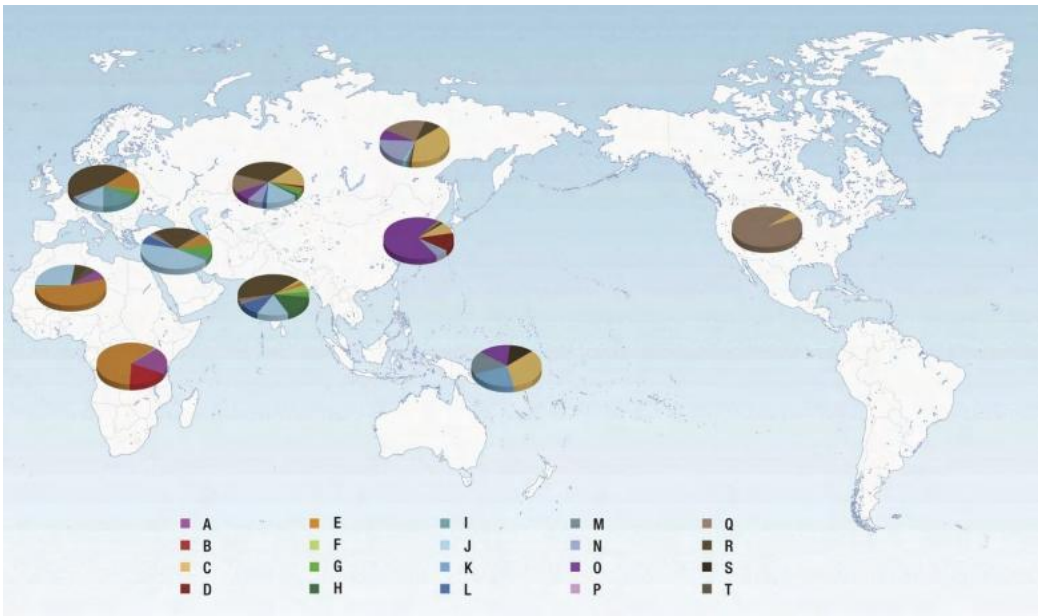


Figure 3: Map showing the frequencies of 20 major Y chromosome haplogroups in worldwide map. [Adapted from Karafet et al. (101)]

Table 1: Selective list of Y-haplogroups, as well as their geographical distribution. [Adapted from Kayser (103)]

Haplogroup	Defining bi-allelic marker(s)	Major geographic area(s) of occurrence
A00-L1086	L1086; L1159; L1284	Central Africa
A0-V148	V148; V166; L896; L991	Central Africa, West Africa
A1-M31	M31; P82; V4	West Africa, North Africa
A2-V50	V50; L602	Southern Africa, Central Africa
A3-M32	M32	East Africa, Southern Africa
B-M60	M60; M181; V244	Central Africa, Southern Africa, East Africa
D-M174	M174; CTS94; JST021355	East Asia
E-M96	M96; M40; P29	Africa, West Asia, Southern Europe
E-V13	V13; V36	Southern Europe
E-M293	M293	Southern Africa
C-M130	M130; M216	Central Asia, Northern Asia, North America, East Asia, Southeast Asia, Wallacea, Near Oceania, Remote Oceania, Australia
C-M8	M8; M105	Japan
C-V20	V20	Southern Europe
C-M356	M356	South Asia, Central Asia
C-B65	B65	Indonesia, Philippines
C-M38	M38	Wallacea, Near Oceania, Remote Oceania
C-M208	M208	Near Oceania, Remote Oceania
C-P33	P33	Remote Oceania
C-PH41	PH41; PH338; PH4186; PH4682	Australia
C-M217	M217; P44; Z1453	Central Asia, Northeast Asia, Northern Americas
C-P39	P39	Northern Americas
G-M201	M201; P257	West Asia, Europe, Central Asia
H-L901	L901; M3035	South Asia
I-M170	M170; M258; U179	Europe, West Asia
J-M304	M304; P209	West Asia, North Africa, Horn of Africa, Southern Europe, Central Asia, South Asia
L-M20	M20	South Asia, West Asia
M-P397	P397; P399; PR2099	Wallacea, Near Oceania, Remote Oceania, Australia
M-P34	P34	Near Oceania
M-M10072	M10072; FGC38729; Z33118	Australia
N-M231	M231	Northern Asia, Northern Europe
O-M175	M175; P186	East Asia, Southeast Asia, Remote Oceania
Q-M242	M242	Northern Asia, Central Asia, Americas
Q-M3	M3	Americas
Q-Z780	Z780	Americas
R-M207	M207	Europe, West Asia, Central Asia, South Asia, North Africa, Central Africa

Table 1 (continued)

Haplogroup	Defining bi-allelic marker(s)	Major geographic area(s) of occurrence
R-M458	M458	Eastern Europe, Caucasus region
R-Z284	Z284	Northwest Europe
R-Z93	Z93	South Asia, Central Asia
R-V88	V88	Africa
R-M412	M412	Western Europe
R-M479	M479	South Asia, Central Asia
S-M254	M254	Near Oceania
T-M184	M184	West Asia, Horn of Africa, North Africa, Southern Europe, South Asia

As previously mentioned, the Y-SNPs are very stable and it is possible to determine the ancestral state of each polymorphism (87,94). These important features allow the construction of a robust phylogeny with these markers (190). To standardize the haplogroup nomenclature based on these polymorphisms, the YCC published in 2002 some rules and the first consensus phylogeny of the Y chromosome. This article was followed by others updating the information of the Y chromosome phylogenetic tree (e.g. 96, 101). The phylogenetic tree of the Y chromosome can be easily found in online internet resources, such as, in the web site of the International Society of Genetic Genealogy (ISOGG; http://www.isogg.org/tree). However, this tree includes all known Y-SNPs, some of them not yet confirmed, and so, it is very complex (180). In order to simplify this phylogeny, a minimal reference tree, also available online, was created in 2014 (http://www.phylotree.org/Y/) (180).

Different methodologies can be used to type Y-SNPs, including the widely used SNaPshot technology. This technology is based on a minisequencing reaction, also known as single base extension reaction (SBE). A primer is positioned with the 3' end at the base immediately before of the polymorphism and a single fluorescently labelled ddNTP (dideoxynucleotide) is extended (167) (Figure 4). Before this reaction, a PCR amplification is required (167). The fluorescent dye of each ddNTP is detected by capillary electrophoresis (167). In order to perform a multiplex reaction, mobility modifying "tails" are added to the 5' end of the primers (167). Each primer will have a different length, which allows the spatial separation of the DNA fragments during the electrophoresis. SNaPshot is a very sensitive and robust technology that allows the analysis of different polymorphisms in a single reaction. However, the number of Y-SNPs genotyped in each reaction is limited (138). This technology uses small length DNA fragments and is efficient even when the samples have a small quantity of DNA (109). Therefore, the SNaPshot technique is considered very useful for genotyping

SNPs in forensic casework where the samples are usually small and degraded. Another advantage of this technology in forensic genetics is the fact that the detection method (capillary electrophoresis) is performed in an instrument also used for STR analysis that are the most used markers in this field (167).

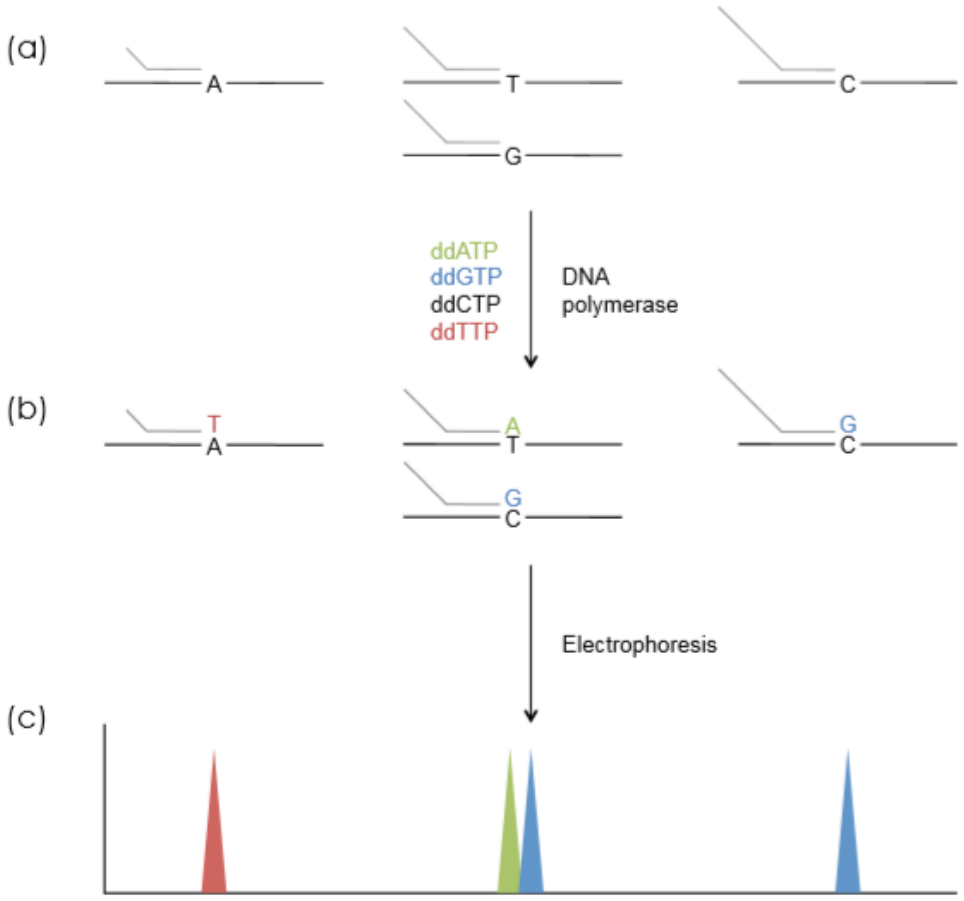


Figure 4: Schematic representation of the allelic discrimination through SNaPshot technology. The steps represented in this figure are: (a) annealing of the primers; (b) extension of the ddNTPs and (c) result obtained after a capillary electrophoresis. Note that the middle SNP has two alleles whereas the other SNPs have just one allele. [Adapted from González dos Santos (84)]

Colonization of the Americas

According to the "Out of Africa" model, the transition to the anatomically modern humans happened relatively recent, between 100 to 200 thousand years ago (kya), in Africa, followed by a dispersal across the world (173). This human dispersal culminated with the colonization of the American continent (75, 183). According to the most accepted model, known as "Out of Beringia", modern humans went from Asia to North America via Beringia and, later, colonized South America. There are a high number of studies from several scientific areas (such as linguistics, archaeology and population genetics) that focused on the questions about how and when the peopling of America occurred. However, many issues related to this process are still unclear.

First studies

The hypothesis indicating that the first American inhabitants were migrants from Asia that passed through the Bering Strait to the Americas was firstly proposed by the Spanish Jesuit José de Acosta in 1590 (51). A few centuries after, Ales Hrdliča suggested the first model for the American Colonization that was named “The Clovis First” or “The Single Origin model” (154). According to this model, the ancestors of the Amerindians entered in the New World through Beringia[2] around 11,500 years ago, being the Clovis culture the most important archaeological evidence supporting this model (57, 75, 154). However, more recent archaeological findings indicate a pre-Clovis occupation of the New World (56, 75, 184, 185). Thus, over the years, with the increase of information from several studies, new hypotheses regarding the colonization of America have been taken into consideration.

Greenberg and co-workers (86) published the first interdisciplinary study about the peopling of the Americas. They considered linguistic, dental and genetic evidence to propose the “Tripartite Model”. They believe that America was settled by three waves of migration from Asia associated with three Native American linguistic groups, Amerind, Na-Dene and Aleut-Eskimo. The first migration was probably by Amerind people, who reached the south. The Na-Dene speakers were the seconds to colonize

[2] Beringia was a land bridge that connected Northeast Asia and North America during a period when climatic conditions were much colder than today and that was submerged after the Last Glacial Maximum (LGM) (93).

America, particularly the Northwest coast, and Aleut-Eskimo speakers were possibly the last to arrive in North America.

"Out of Beringia" model

In 1997, Bonatto and Salzano (27) proposed the "Out of Beringia" model. The authors published a study based on mtDNA where they emphasized the importance of Beringia not just as a land bridge between two continents, but also as a place of settling and diversification of the Native American ancestors before the colonization of the New World. Ten years later, the “Beringian standstill hypothesis” or “Beringia incubation model” was proposed by Tamm and colleagues, as a revised version of the "Out of Beringia" model (93). According to Tamm et al. (170), a Northeast Asian descendent population was genetically isolated in Beringia, during about 15,000 years (Figure 5). With this hypothesis, also based on mtDNA data, the authors suggested that there was an initial colonization into North and South America by the Beringian population and after that, there was a migration from Northeast Asia into Beringia and then North America. They also believe that two back-migrations into Asia might have occurred.

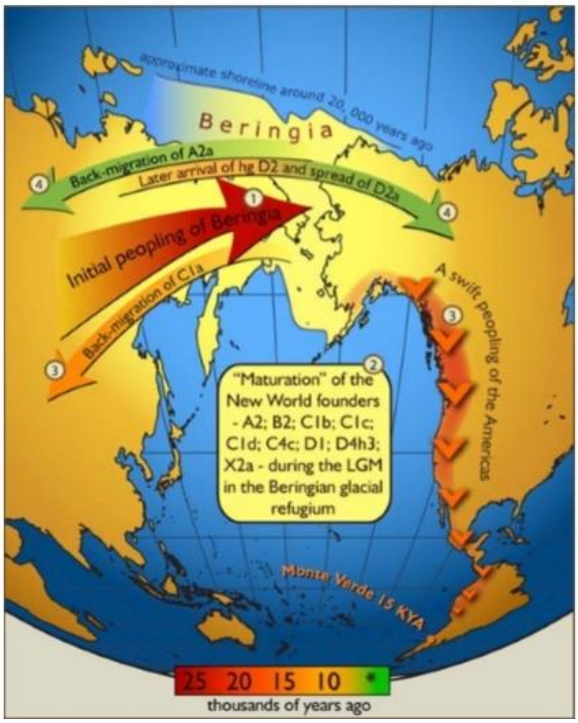


Figure 5: Representation of the peopling of the Americas according to Tamm and colleagues’ proposal. Colored arrows indicate the approximated timing and the direction of the maternal gene flow events. [Adapted from Tamm et al. (170)]

Entrance in the Americas

During the Last Glacial Maximum [LGM; ~26.5 to 19kya (43)] it is believed that humans lived in glacial refuge areas (3, 68, 127, 129), being Beringia one of these places (93, 77, 92). Some authors believe that the Native American ancestors remained isolated in Beringia during 2.4-9ky (110, 131, 137). The Laurentide and Cordilleran ice sheets covered northern North America and blocked the path between Beringia and North America until the end of the LGM (75, 165). After that, when the weather got warmer, the margins of the Laurentide and Cordilleran ice sheets started melting creating an ice-free corridor between them and a Pacific coastal corridor (75, 183). These corridors made possible the entrance into the New World that occurred approximately between 15 and 19kya (75, 110, 128, 131, 183). The finding of a pre-Colvis evidence in Southern Chile (South America) dated ~14-14.6kya [56; Dillehay, 1997 cited by Goebel et al. (75)] suggests that, after the entrance in North America, there was a rapid migration and the colonization of South America (26, 29).

Other hypotheses of the Native Americans origin

Evidences found in the New World led some researchers to suggest other ancestry populations to American Indigenous. A European origin of Native Americans was proposed with the Solutrean hypothesis (28). According to this hypothesis, mainly based on archaeological data, the Upper Palaeolithic Solutrean technologies of Southwestern Europe originated the pre-Clovis and Clovis technologies (28). Thus, the Clovis ancestors would have crossed the Atlantic instead of the Bering land bridge (28). Some Y chromosome data seems to support this hypothesis (186), however other genetic studies show that the first Americans did not have a European ancestry (136, 139). Another contribution to the South American gene pool was proposed due to the discoveries of individuals from Lagoa Santa and some present-day Amazonian tribes that have some genetic similarities with native people from New Guinea, Australia and Andaman Islands (120, 137, 164, 165). This Australasian genomic signature found in some Native Americans was not found in ancient DNA samples from Siberia or Beringia (162). However, according to Skoglund et al. (164), this similarity is not due to a migration from Australasian to America but derived from an ancestral population that no longer exists and contributed for both Australasian and Native American ancestors.

Genetic evidence

Molecular data supports that American Indigenous people from the north and south of the continent derive from the Beringian population (87, 119). It is also widely agreed that this population splits into two branches named "Southern Native American" or "Ancestral A" (ANC-A) and "Northern Native American" or "Ancestral B" (ANC-B) (1, 119, 132, 137, 160). Some authors believed that this division happened at 17.5-14.6kya (119). However, this idea as well as the place of the split are not clear yet (1, 119, 120, 160, 183). It might have occurred at the south of Laurentide and Cordilleran ice sheets (119, 160), or south of eastern Beringia (119, 120). The ANC-B branch only peopled the northern region of North America whereas the ANC-A branch reached the south, which means that different groups of this branch colonized North, Central and South America (120, 132, 183). Genetic evidences also indicate that the nowadays Indigenous might be descendent from three different migration waves, two of them occurred more recently and only to North America (140, 143). Some authors defend that, firstly, humans followed a southward path along the Pacific coastal line (26, 63) and reached the northern latitudes of South America, where they split and followed two main distinct routes (the Pacific and Atlantic coasts) (77, 120, 132). However, according to Scheib et al. (160) the humans that reached Central and South America are a result of one or multiple admixture events between the two branches (ANC-A and ANC-B).

Mitochondrial DNA data

Studies focus on the mtDNA data revealed that present-day Native Americans are characterised by five mtDNA haplogroups (A, B, C, D and X) of Asian origin, except the X lineage (2, 17, 129). The haplogroups A, B, C and D are ubiquitous in the continent whereas the X is mainly found in North America (58, 129). Some authors believe that these maternal lineages entered in the New World approximately 16kya after a period of isolation in Beringia (110). The entrance in the double continent might have occurred through both the Pacific coastal corridor, marked by D4h3a sub-haplogroup, and the ice-free corridor, marked by X2a and C4c sub-haplogroups (102, 128). Mitogenome based evidences also indicate a rapid migration through the Americas that possibly took about 1.4ky (110).

Y chromosomal data

The genetic diversity of the paternal lineages in Native Americans is also low and decreases from north to south of the continent, as happens with mtDNA diversity. Relatively to South America, the West coast presents a higher level of paternal genetic diversity than the East coast (171).

The Y chromosome founding lineages (C and Q) of Native Americans have also an Asian origin (186, 194). The haplogroup C is mainly present in North America whereas the haplogroup Q can be found all over the American continent (21, 59, 194). Concerning the haplogroup C, there are two main sub-lineages that can be found in the American continent. The C-P39 is commonly present in North American native populations (194) whereas the C-M217 can be found in South America, more specifically in Kichwa and Waorani native groups from Ecuador (74, 151) and in the Wayuu from Colombia (194). The South American lineage, although more basal than C-P39, possibly came to the New World with the first American colonizers and was lost by genetic drift in almost all American populations (115). Relatively to the haplogroup Q, there are two main paternal founding lineages, the Q-M3 and the Q-Z780 (87). After entering in the American continent, the Q-M3 lineage splits into Q-M848 and Q-Y4276 (87). Together with Q-B34 and Q-Y4300, the Q-Y4276 lineage possibly entered North America at the same time and through the same route of the X2a and C4c mtDNA haplogroups, constituting the ANC-B group (87). The Q-M848 and Q-Z780 probably entered in the Americas with the first settlers and followed the Pacific coastal route to reach South America, representing the ANC-A component (87). Thus, Y chromosome data seems to indicate that the differentiation between ANC-A and ANC-B components occurred in Eastern Beringia and then they followed the coastal and inland corridors to enter in the Americas. Paternal data also support an isolation period in Beringia before the entrance in the New World at about 19.5-15.2kya (131). Paternal and maternal evidence also suggests an admixture event (before ~9kya) in North America between the two ancestral components, since the Y chromosome DNA of Kennewick man (Q-M848) belongs to ANC-A component, but according with mtDNA analyses the same individual belongs to ANC-B (X2a) (87, 120, 141).

Ecuador

The Republic of Ecuador is a small country located in the Northwest region of South America, which is composed by a continental land surrounded by Colombia (North), Peru (East and South) and the Pacific Ocean, and the Galápagos Islands (Figure 6). The mainland territory can be divided into 23 different provinces distributed in three ecological regions: Coast, Andes (*Sierra*) and Amazonia. A huge variety of habitats with great biological diversity can be found in Ecuador that is considered the smallest megadiverse country in the world (80). Ecuador is the most biodiverse country in comparison to its area and have also a huge diversity of human ethnic groups (80).

Figure 6: Map showing the location of Ecuador (in red).

The pre-Colombian times in Ecuador can be divided into two main periods related to the pre-Inca and Inca cultures. Before the Inca conquest, the area that nowadays is known as Ecuador was populated by a variety of tribes who have different cultures and languages (89). The Inca conquest of Ecuador started in 1463 and during the following years some aspects in agriculture and social organization, for example, suffered some modifications (8, 89). However, other aspects, such as religious beliefs remained unchanged during the Inca period (8, 89). In 1492, Christopher Columbus arrived for the first time to the New World, beginning the European interest in that continent (8, 172).

The Spanish people arrived in Ecuador in the early sixteenth century (first expedition was in 1526). With them, some diseases (such as, smallpox and measles) also arrived that were responsible for a dramatic decline of the indigenous population mainly in the Coast region (8, 89, 172). During the Spanish occupation of Ecuador, the Amerindians were enslaved and converted to Christianity (89). Africans were also enslaved, firstly, they were brought from different African countries (Figure 7) to the Caribbean region

and then distributed to other American countries (9, 70). Thus, the Africans did not come directly from Africa to Ecuador but from another American region. Different regions of the African coast, between Senegambia and Angola, were mentioned to be the source of some African groups in Ecuador (60, 80, 156). Since Africans were more expensive than indigenous, they were mainly used in the plantations along the coast where some natives proved not to be able to work (9, 89). Africans were also brought to work in plantations in the Valle de Chota located in the Northern Andean region (9, 60). By the end of the colonial Era, approximately 60,000 African slaves inhabited in Ecuador (89).

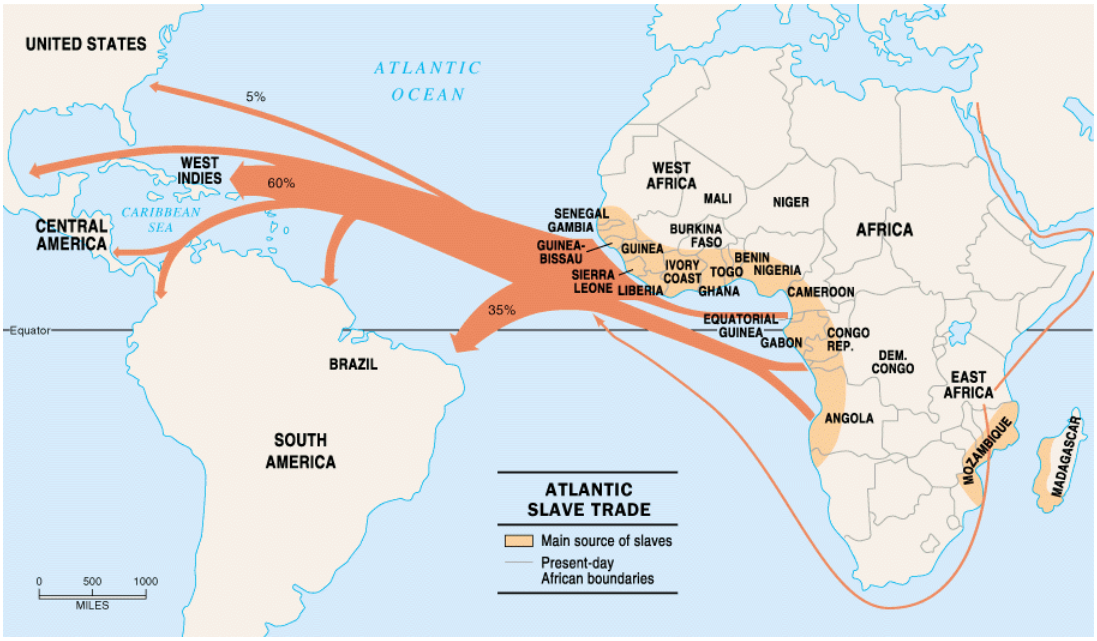


Figure 7: Map representing the main regions from where the African slaves were brought to America. [Adapted from Gambéla (67)]

During the colonial period, some regions in Amazonia and Coast were not conquered by the Spaniards and therefore were perfect refuges to Native Americans (89). In 1553, a boat that was transporting African slaves from Panama to Peru, shipwrecked nearby the Ecuadorian coast (9, 36, 70, 172). The African people, who survived to the shipwreck (17 men and 6 women), went to Esmeraldas, where they overpowered the native populations (60, 70, 172). A few years after, this region was mainly populated by Africans and Zambos (descendants of Africans and Indigenous) that lived independent from the Spanish colonizers (89, 105, 147). Through the years this region was a refuge for other survivors of shipwrecks in the Esmeraldas coast and in the 18th century, some Colombian slaves escaped to Esmeraldas that increased the African population in that area (70, 105). Until these days there are some Afro-descendant communities in

America that live relatively isolated, they are known as *Palenques* or *Quilombos* (in Brazil).

More recently, other waves of migration from Europe to Latin America occurred to Ecuador, mainly from Great Britain, Italy and Germany (9). Immigration from China to the Americas in the 19th century was also reported, and some of these Chinese immigrants went to Ecuador, even though it was not their main destination (9). In the beginning of the 20th century, people from the Middle East (mainly from Lebanon, as well as from Syria and Palestine) arrived in Ecuador (9). This country also received Jewish immigrants, mainly from Germany (9).

Thus, the historical events that occurred in this country make possible the admixture between different populations (Native Americans, Europeans and Africans) resulting in three main ethnic groups that currently inhabit the Ecuador: Mestizos, Afro-Ecuadorians and Native Americans. The Ecuadorian population is mainly constituted by Mestizos (72%), followed by a mestizo group called Montubios (7%), Afro-Ecuadorians (7%) and Native Amerindians (7%) (104). Ecuador is also inhabited by White Ecuadorians (mainly descended from the Spaniards), which represents 6% of the population, and some other ethnic groups can be found at 1% (159). These values were based on self-determination of the ethnic group by each individual, which means that probably the values would be different if the genetic data was taken into consideration.

The Mestizos are mainly descendent from European men, in particular Spaniards, and Indigenous women (8, 82). They inhabit principally the urban regions of the country (7). The Montubios is also an admixed group, originated in the coastal region, that was recognized as a distinct ethnic group by the Ecuadorian government in 2001 (159). This group has cultural characteristics that are different from the other ethnic groups, even from Mestizos (191). Afro-Ecuadorians are mostly descended from African slaves and they can be found mainly in the province of Carchi and Esmeraldas (79). The Native American population is composed of 21 indigenous groups (108). The Quichua (Kichwa) is the most numerous native ethnic group that can be found in both Amazonia and Andean region (82, 176). They speak Quichua (Kichwa), which is the second most spoken idiom in the country (79). Ecuador is also home of indigenous groups that live in volunteer isolation (64). They are known as Tagaeri, Taromane, Oñamenane and Huiñatare, all belong to Waorani (Huaorani) ethnic group, which is considered to be the last hunter-gatherer nomad group from Ecuador (79, 80, 178). This group can be found in Amazonia rainforest, in the provinces of Pastaza, Napo and Orellana (80,178).

Genetic diversity of Ecuadorian population

Previous genetic studies demonstrated that the nowadays genetic diversity of Ecuador is in agreement with the historical facts about the country. Concerning autosomal data, Native Americans, Europeans and Africans contributed to the Mestizo population (82, 133, 158, 192, 193). The Native American population seems to be the main contributor to the Mestizos, followed by the Europeans (Spaniards) and by a smaller percentage the Africans (82, 133, 158, 192, 193). The Afro-Ecuadorians are also the result of admixture between three ancestral populations, with the greatest contribution from Africans followed by Amerindians and lastly Europeans (82,158). The present-day Native American groups have a small proportion of European ancestry (158). The mtDNA data showed a predominant Amerindian contribution to the Mestizo and Native American populations. A prevailing African and Native American maternal ancestry is present in the Afro-Ecuadorian population (10, 13, 146).

On the other hand, the Y chromosome data tells us a slightly different history about the Ecuadorian population. Both Native American Y haplogroups (Q and C) can be found in Ecuadorian Amerindians, being the Q-M3 lineage the most frequent (13, 74, 151, 157). Some European lineages can also be found in Amerindian population (82). The major ancestry in Mestizos males is European, followed by Amerindian and African lineages (13, 81, 82). Afro-Ecuadorian Y chromosome lineages are mainly from an African origin, followed by European and Amerindian origins (81, 82). Looking at all Ecuadorian ethnic groups together, we can see that the predominant lineages have a European origin, but Native American lineages are also important (71, 176). The Amazon rainforest is the region with more balanced proportions of Amerindian and European ancestry and is the only region where African ancestry seems to be absent (176). The African ancestry is higher in the Coast region, but, as in the Andean region, the African contribution is small compared with the other two ancestral contributions (176). Therefore, the Y chromosome lineages that can be found in Ecuadorians are predominantly within the major haplogroups Q, R and E (13, 71, 74, 82, 151, 157). However, since most studies about Mestizos and Afro-Ecuadorians are focused only in Y-STR analysis, further studies are still required to increase the genetic diversity characterization of the Y chromosome lineages in the Ecuadorian population.

Aims

The study of the genetic diversity in human populations is important to unravel past and recent demographic events. In particular, the knowledge of the genetic background of South American populations is crucial to a better understanding of the colonization of this continent by the modern humans, and the more recent human migrations during the colonial Era. There are already some genetic studies focused on South American populations. However, due to the high heterogeneity of its populations, more genetic data is still required for a comprehensive view of the genetic diversity and its stratification in this sub-continent.

The main objective of this work was to contribute to a better knowledge of the paternal genetic diversity in Ecuador and its stratification across different population groups. To this end, it was intended to characterise the Ecuadorian paternal lineages, in particular from the Mestizo and Afro-Ecuadorian ethnic groups. Comparing data from the present study with data from the literature for Central and South American populations, we aimed to contribute to a better understanding of the admixture process behind the formation of these two ethnic groups.

To achieve the main aim of this study, the following goals were proposed:

- Characterise the paternal lineages in two population samples from Mestizos and Afro-Ecuadorians living in Ecuador, through the analysis of Y-chromosomal specific markers (23 Y-STRs and 59 Y-SNPs);
- Compare the two studied groups through R_{ST} and F_{ST} analysis and investigate if they share genetic similarities;
- Disclose the diversity of the Ecuadorian samples in the context of Latin American populations;
- Determine the most likely origin of the European and African haplogroups found in the two Ecuadorian ethnic groups, based on Y chromosome phylogeographic information;
- Interpret the genetic results obtained taking into account the paternal history of Ecuador.

Materials and Methods

Population samples and DNA extraction

In this work a total of 299 male samples from Ecuador was collected (149 from Mestizo and 150 from Afro-Ecuadorian ethnic groups). These samples were obtained from unrelated males, with at least two preceding generations (parents and grandparents) from Ecuador. Samples were collected through mouth swabs in Afro-Ecuadorian individuals, and in Mestizo people through mouth swabs or blood spots on Whatman™ FTA™ cards. The DNA was extracted using the Phenol-chloroform and Chelex-100® (Sigma-Aldrich) standard methods. Information about ethnic group and region of birth for each sample donor is presented in Appendix (Table S1).

DNA analysis

Y-STR typing

The Y-STR genotyping was performed in the Universidad de Las Américas, Quito, Ecuador, with the PowerPlex® Y23 System (Promega). This kit allows the simultaneously amplification of 23 Y-STR loci (DYS576, DYS389 I, DYS448, DYS389 II, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385 a/b, DYS456, Y GATA H4). The Y-STR haplotype data allowed to predict the haplogroup to which sample belong, using an online available tool (Y-DNA Haplogroup Predictor - NEVGEN; https://www.nevgen.org/). In accordance with the results from NEVGEN, it was possible to select which multiplex of Y-SNPs (see Appendix Figure S1) would be genotyped in each sample, reducing costs and laboratory time.

Y-SNP typing

The analysis of 59 Y-SNPs (SRY10831.1, M213, M9, Tat, 92R7, M173, SRY10831.2, P25, M70, M60, M182, M150, M109, M112, M30, M168, M201, M170, M26, 12f2, M62, M172, M242, P36.2, M346, M3, M19, Z19319, SA01, Z19483, M557, SA05, M96, M33, P2, M2, U209, M154, U290, M191, U174, M35, M75, M85, M78, M81, M123, V6,

M293, M207, V88, M269, L23, U106, S116, U152, M529, M153, M167) allowed the identification of 59 different Y chromosome haplogroups. These Y-SNPs were included in eight multiplexes (see Appendix Figure S1), except V88 that was genotyped in singleplex. More details about the multiplexes 1 and 2 can be found in Brion et al. (31). Concerning multiplexes E2 and B, details can be found in Gomes et al. (76). Details on multiplex R1 and R2 can be found in Resque et al. (144). The V88 primer sequence can be seen in González et al. (83). Information about multiplexes E1 and Q as well as some optimisation details in concentrations and sequences of the primers of the other multiplexes are presented in Appendix (Tables S2-S4).

As previously mentioned, the multiplexes were selected in accordance with the haplogroup prediction results. For some samples we had to type more than one multiplex, since the prediction was incorrect and/or the multiplex did not include all Y-SNPs needed. The Y-SNP V88 was only typed in two samples, which haplogroup predictor identified to belong to haplogroup R-V88.

The DNA samples were first amplified by multiplex PCR. The PCR mix contained 2.5µL of 2x QIAGEN® Multiplex PCR Kit [which includes an enzyme, minerals and deoxynucleotides (dNTPs)], 0.5µL of a primer mix specific for each multiplex (each primer at 2µM). The volume of DNA added was 0.5µL for Mestizos samples (≥0.5µM) and 1µL for Afro-Ecuadorians samples (<0.5µM), and water was added to complete 5µL of final volume. In some cases, the volume of sample had to be optimized (depending on DNA concentration) and ranged from 1µL to 2µL. A negative control was used in each PCR reaction in order to confirm the absence of PCR contamination.

The amplification reaction started with an initial denaturation temperature of 95ºC for 15 minutes, followed by 35 cycles at 94ºC for 30 seconds (denaturation step), 60ºC for 90 seconds (annealing step) and 72ºC for 60 seconds (extension step). A final extension step was performed at 72ºC for 10 minutes. The annealing temperature in multiplex Q was optimized to 62ºC, to increase the specificity of the PCR products. The amplification products were confirmed on polyacrylamide gel (T9%, C5%) by silver staining method as described by Budowle et al. (31), but with few timing optimisations (10 minutes in the ethanol solution instead of 5 and 5 minutes in the nitric acid solution instead of 3). Examples of polyacrylamide gel of each multiplex are presented in Appendix (Figures S2-S7).

After confirming the successful amplification of all target products, a purification step was performed to remove any unconsumed nucleotides and primers, avoiding interference in the next reaction. The PCR products were purified adding 0.5µL of

mixture of Exonuclease I (25µL) and FastAP (100µL) enzymes (Thermo Scientific™) to 1µL of each amplification product. In cases where PCR products were weak, 1.5µL of PCR products were added and the enzyme mix volume was adjusted to 0.75uL. The conditions used in this purification step were incubation at 37ºC for 15 minutes, followed by 15 minutes at 85ºC for the inactivation of the enzyme.

The Y-SNP alleles were sequenced through a single-base extension (SBE) reaction using the SNaPshot™ Kit (Applied Biosystems™). The SBE reaction was performed for each multiplex in a final volume of 5µL, using 1µL of SNaPshot™ Multiplex Ready Reaction Mix (Applied Biosystems™), all volume of the purified PCR product and 1µl of the SBE mix composed by a specific concentration of each SBE primers as described in Tables S3 and S4. The remaining volume was completed with water. The samples were then submitted to 25 cycles at 96ºC for 10 seconds (denaturation step), 50ºC for 5 seconds (annealing step) and 60ºC for 30 seconds (extension step).

A final purification was performed immediately after the end of the SBE reaction by adding 1µL of FastAP (Thermo Scientific™) to the products. Incubation time was 1 hour at 37ºC followed by an enzyme inactivation at 85ºC for 15 minutes.

Finally, 0.5µL of the purified product was combined with 0.05µL of Size Standard Liz120 (Applied Biosystems™) and 9.45µL of Formamide Hi-Di, and run in an ABI 3500 Genetic Analyser (Applied Biosystems™). The results were analysed with GeneMapper™ Software v5.0 (Applied Biosystem™). Examples of the electropherograms of each multiplex are presented in Appendix (Figures S8-S13).

As already mentioned, the Y-SNP V88 was not included in any of the multiplexes used, but genotyped through direct DNA sequencing reaction. Prior to the sequencing reaction the DNA samples were amplified in a PCR . The PCR mix contained 2.5µL of 2x QIAGEN® Multiplex PCR Kit, 0.5µL of each primer (each primer at 2µM), 1µL of DNA and water to complete a 5µL of final volume. A negative control was also used.

The PCR amplification conditions were the same as for the aforementioned multiplexes, except the annealing temperature that was increased to 63ºC. This optimization allowed to increase the specificity of the primers regarding the Y chromosome since V88 is in a region where the Y chromosome has high homology with the X chromosome. The amplification of products was also confirmed by an electrophoretic run in polyacrylamide gel.

The amplified products were purified with Exonuclease I plus FastAP (Thermo Scientific™) in the same conditions as previously described for the multiplexes. The

sequencing reaction was performed in a total volume of 5µL, including 0.8µL of BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems™), 1µL of BigDye™ Terminator v1.1, v3.15x Sequencing Buffer (Applied Biosystems™) and 0.5µL of the sequencing primer at a concentration of 10µM. The volume of purified product for sequencing was 1.5µL and water was added to complete the 5µL final volume.

After the sequencing reaction, a final purification step was performed in order to remove salts, dNTPs and ddNTPs, and other impurities that could interfere with the results. This purification step was done through columns with Sephadex™ G-50 (GE Healthcare) at 10%. The entire volume of sequencing products was purified, mixed with 8µL of Formamide Hi-Di and submitted to capillary electrophoresis in an ABI 3500 Genetic Analyser (Applied Biosystems™). The results were analysed through Sequencing Analysis Software v6.0 (Applied Biosystems™).

Data analysis

Y-SNP haplogroup frequencies were determined by direct counting. The Arlequin 3.5.2.2 Software (62) was used to calculate genetic diversities and pairwise R_{ST} and F_{ST} genetic distances between our samples. Pairwise R_{ST} and F_{ST} genetic distances were also calculated between our samples and samples from other Latin American, European and African countries (see Appendix Tables S5-S7).

In the majority of the pairwise genetic distances analyses, the results were visualized in a two-dimensional graphic through the Multi-Dimensional Scaling (MDS) analysis included in the software STATISTICA 13.3 (www.statsoft.de). This software was also used to understand which lineages were influencing the relationship between some populations analysed through the Principal Component Analysis (PCA).

The Network 10.0 software (Fluxus Technology Ltd.) was used to investigate genetic relationships within specific haplogroups. Median-Joining networks (15) were constructed after applying the Reduce Median method (16) to reduce the reticulation. Differential microsatellite weighting (inversely proportional to the variance) was also applied in order to obtain the most parsimonious network in accordance with Qamar et al. (135). The Y-STR DYS385 was not used for the construction of the networks and the allele of DYS389 II was obtained after subtracting the number of repeats obtained for DYS389 I, which was also done for the analyses with Arlequin.

Results and Discussion

Ecuadorian genetic diversity

The haplotypes and haplogroups obtained for each of the 299 samples are present in Appendix (Table S1). The two Ecuadorian ethnic groups of the present study were compared using both classical (F_{ST}) and the stepwise-based (R_{ST}) genetic distance methods. The F_{ST} method evaluates the frequency of different alleles between each pair of populations, not considering the mutational mechanisms (35). Therefore, it is more appropriated than the R_{ST} method when genetic drift is the cause for the differences between populations (35). On the other hand, the R_{ST} method assumes the single stepwise model for the formation of new alleles and takes into account both genetic drift and mutation, evaluating the number of differences observed at each locus (35). The R_{ST} method was developed to perform comparisons focused on microsatellite data (166).

Y-STR data

Through the genotyping of the 23 Y-STR markers included in the PowerPlex® Y23 System (Promega), a total of 147 unique haplotypes were identified in the Mestizos and 121 in Afro-Ecuadorians. No founder effects were detected, since high haplotype diversities were obtained in the two groups, being the higher value found in Mestizos (0.9998±0.0008 and 0.9961±0.0015, respectively). These high values are expected since the Ecuadorian population is a result of admixture between well genetically differentiated populations of Amerindians, Europeans and Africans. Furthermore, the high haplotype diversities obtained in this study are in accordance to the previously reported for Ecuadorian population (12, 71, 176) and specifically for these two ethnic groups (82). Nevertheless, according to the values obtained in González-Andrade et al. (82) it was expected a higher value of haplotype diversity for the Afro-Ecuadorian group.

Looking to the pairwise genetic distances between the two ethnic groups, a significant R_{ST} value (R_{ST}=0.24073; P=0.00000) was obtained, which is consistent with the different histories underlying the origin of these two ethnic groups.

In order to understand if the genetic composition of the two ethnic groups of this study showed significant differences between regions of the country further pairwise genetic distances analyses and an AMOVA (Analysis of Molecular Variance) were done (data not shown). The samples from Mestizos were divided into the three ecological regions (82 from Coast, 59 from Andes and 7 from Amazonia) and the Afro-Ecuadorian samples were divided into two regions (117 from Coast and 33 from Andes). These divisions were made according to the birthplace of each individual sampled. The R_{ST} values were concordant with the previously shown in this study since significant differences were obtained between different ethnic groups belonging to the same region. When considering the same ethnic group in different regions, no significant R_{ST} values were found. AMOVA results also showed non-significant differences between regions within the same ethnic group.

Population comparisons were also performed between the two ethnic groups of the present study and data available from other studies from Ecuador (13, 74, 81), using 10 Y-STR markers common between all studies (Appendix Table S8). Considering the two Ecuadorian ethnic groups of the present study, a different R_{ST} value was observed between them, since the number of markers used in the analysis influences this value. All Native American samples show significant R_{ST} values (P<0.05) when compared with the Mestizos and Afro-Ecuadorians, which is an expected result since the main population underlying the formation of these ethnic groups is different, as previously mentioned in this thesis. Between the Native American populations, significant R_{ST} values were also obtained. These results can be explained by the small sample size (n=27) in one case. For the remaining samples, results can be due to differences in populations' composition. The indigenous Quichua group (81) mainly composes one of them, whereas Waorani group (74) mostly composes the other one. Indeed, a strong effect of genetic drift was reported in the Waorani group and that this group presents less haplotype diversity than the Quichua population (81). Moreover, a Y-SNP was found that seems to be specific of the Quichua group, not being present in any of the Waorani samples analysed in the study of Geppert et al. (73). Non-significant differences were detected between the Mestizos samples from all studies, but a significant R_{ST} value (R_{ST}=0.06079; P=0.00000) was obtained between the Afro-Ecuadorian samples from this and the González-Andrade et al. (81) studies. Effectively is notable the difference between haplotypes found in these two studies. One possible reason for this difference is the fact that in González-Andrade et al. (81) all samples are from Esmeraldas province whereas in the present study 33 samples are from the Andean region and the other 117 are from the Coast, mostly from Esmeraldas.

Moreover, it should be taken into consideration that 27 of the samples from Esmeraldas belong to an isolated group of Afro-descendants (a *Palenque* in Playa de Oro) that has never been reported. According to historical information, Africans in both regions have multiple origins. Therefore, the differences found between the Afro-Ecuadorians might be due to the presence of a specific population group in the foundation of the Afro-descendant group in one of the Ecuadorian regions and its absence in the other region. Another reason to justify these differences could be due to different histories underlying population admixture in both Coastal and Andean regions from Ecuador.

A MDS plot was constructed in order to make easier the visualization of the R_{ST} genetic distances (Figure 8). The plot shows a clear proximity between all Mestizos and the high distances between all Native American samples. The two Native America populations (NAM1 and NAM2) presented a smaller R_{ST} genetic distance between each other (R_{ST}=0.08175; P=0.00000) than with the other population sample (R_{ST}=0.12102; P=0.00000 and R_{ST}=0.24811; P=0.00000). In the MDS plot, it is also notable the distance between Mestizos and Afro-Ecuadorians from this study.

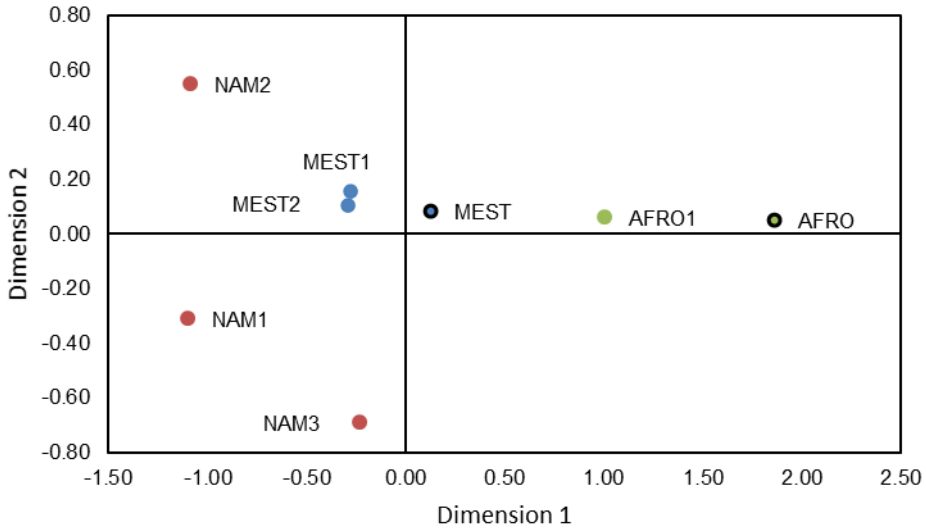
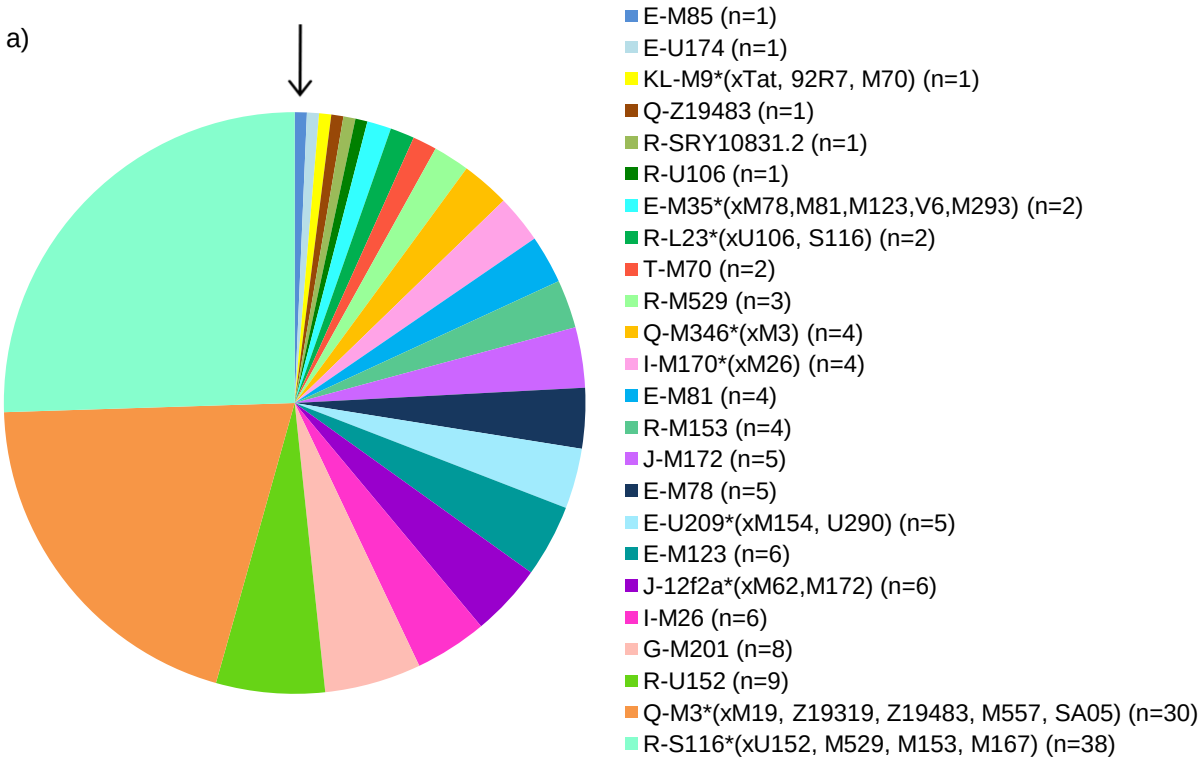


Figure 8: MDS plot of the R_{ST} genetic distances between the three Ecuadorian ethnic groups analyzed, based on information from 10 Y-STR markers (Stress=0.0019540). The present study populations are highlighted with a black circle (green - Afro-Ecuadorians; blue - Mestizos and red - Native Americans).Population codes can be seen in Appendix Table S5.

Y-SNP data

Concerning the Y-SNP data, 32 different haplogroups were identified in the total sample (24 in Mestizos and 24 in Afro-Ecuadorians), 16 of which are common between the two groups. High haplogroup diversities were obtained in the two groups (0.8815±0.0167 in Mestizos and 0.8565±0.0169 in Afro-Ecuadorians). Moreover, a significant pairwise F_{ST} genetic distance was obtained between the two ethnic groups (F_{ST}=0.10846; P=0.00000). This is in accordance with the high differences in Y-haplogroup frequencies within each population group (Figure 9).

a)



(Continues on the next page)

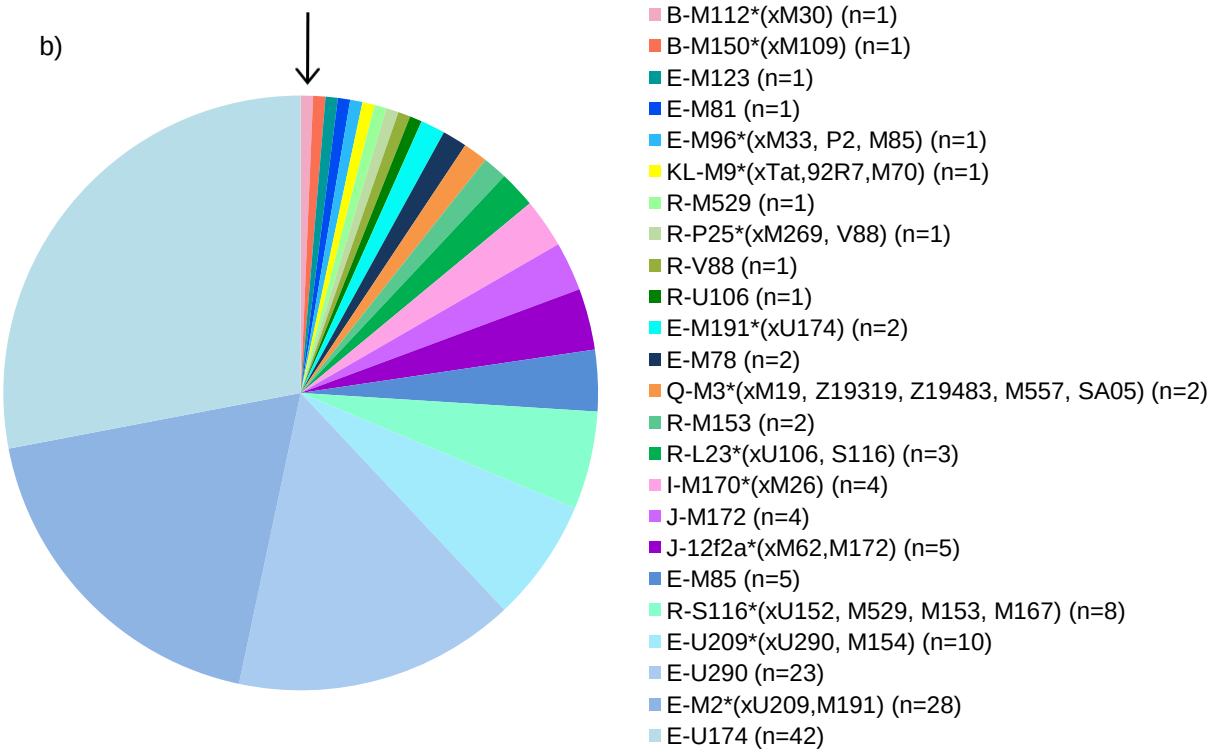


Figure 9: Frequency of the Y-SNP haplogroups found in the two ethnic groups in study [Mestizos (a) and Afro-Ecuadorians (b)]. The arrows indicate the first haplogroup mentioned and the other ones follow the clockwise direction. Note that the haplogroups E, R and Q are represented in tones of blue, green, and orange and brown, respectively.

Haplogroup B

The haplogroup B represents 1.33% of the Afro-Ecuadorian lineages. This haplogroup is one of the most basal on the Y chromosome phylogenetic tree and it is present almost exclusively in African populations, being present in other continents due to recent migrations. The sub-haplogroup B-M150 can be found throughout sub-Saharan Africa (19) and the B-M112 can be found mainly in Pygmy and Khoisan individuals (20).

Haplogroup E

The majority of the Y-lineages found in the Afro-Ecuadorians belong to major haplogroup E (76.67%), being E-U174 the most common sub-haplogroup (36.52% of the E lineages). Contrarily, the haplogroup E only represents 16.11% of the total lineages in the Mestizo group. Haplogroup E is present at high frequencies in the African continent, but can also be found in the Middle East, Southern Europe and Central and South Asia (101). The sub-haplogroup E-M2 and its sub-lineages, which are present at high frequencies in most of Bantu sub-Saharan populations (52, 155),

are the majority of E lineages found in this study (84% and 25% in Afro-Ecuadorian and Mestizos, respectively).

The haplogroup E-M35 is distributed for different continents, being found in Africa, West Asia and Europe (46, 47, 52, 177). Within E-M35 haplogroup, the E-M81 and E-M123 sub-lineages are present in North Africa (being absent in sub-Saharan Africa) and in Europe (46, 112). In present study, a European origin was assigned to these two sub-lineages since North Africans have not been reported as a possible source of African lineages in Ecuador. E-M78 sub-lineages (downstream E-M35) were also found in the two ethnic groups studied. This Y-haplogroup can be found in Europe and in sub-Saharan Africa. Therefore, due to the colonization history of Ecuador, the E-M78 lineages have two probably origins. In order to clarify its origin in Ecuador a network was constructed. This analysis was based on information of 10 Y-STR markers common between all studies [Ecuadorian samples, Europeans (23, 39, 76, 83, 152, 174) and sub-Saharan Africans (4, 25, 124, 145, 168)] (see Figure 10). In the network, it is not possible to see a clear separation between the European and sub-Saharan African lineages. Nevertheless, two Ecuadorian samples share haplotypes with Europeans (indicated with arrows in Figure 3) and the remaining is closer to the European than to the African lineages. For that reason, the E-M78 lineages found in our Ecuadorian population are more likely to have come from Europe.

In the Mestizo group these three lineages within E-M35 (E-M78, E-M81 and E-M123) represent the majority of E Y chromosomes found (62.5%).

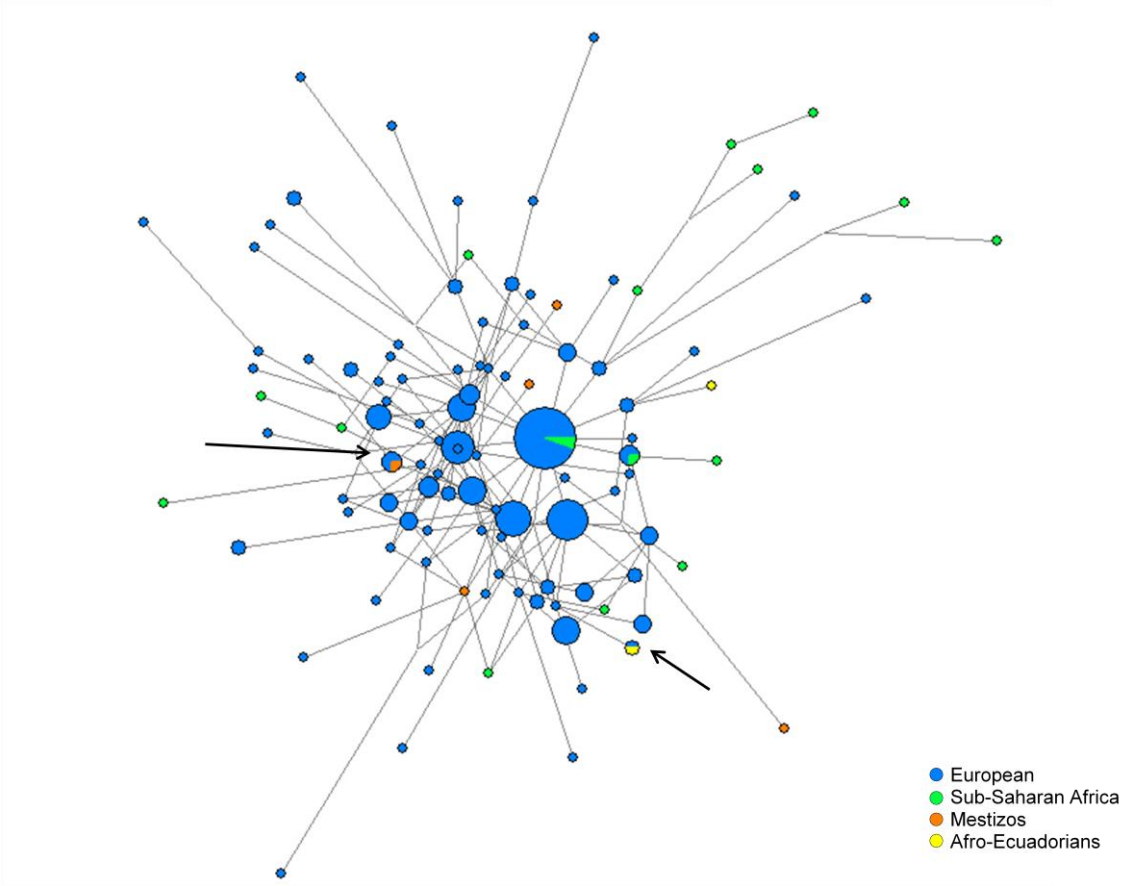


Figure 10: Network constructed with 225 European (Spain, Portugal and Italy), 17 sub-Saharan Africa (Gabon, Guinea-Bissau, Equatorial Guinea, Madagascar and Uganda) and 7 Ecuadorian samples belonging to the sub-haplogroup E-M78.

Haplogroups G, I and J

The haplogroup G-M201 was found in 5.37% of the Mestizos chromosomes and is absent in Afro-Ecuadorian samples. This haplogroup is mainly present in the Middle East, the Mediterranean, and the Caucasus Mountains (101).

The haplogroup I is one of the most frequent European Y-lineages (101) and although detected in both Ecuadorian ethnic groups, it was more frequent in the Mestizo group (6.71% in Mestizos and 2.67% in Afro-Ecuadorians).

The J clade can be mainly found in the Middle East, North Africa, Europe, Central Asia, Pakistan, and India (101), being the sub-haplogroup J-M172 the most common in European populations (161). In our Ecuadorian sample, 7.38% of the Mestizos and 6% of the Afro-Ecuadorian lineages belong to haplogroup J. The majority of these lineages belong to the sub-haplogroup J-12f2a*(xM62, M172). However, the number of J-12f2a*(xM62, M172) and J-M172 sub-haplogroups differed between the two groups by just one sample.

Considering the history of the country, all these lineages are most probably of European origin. Nevertheless, other origins cannot be completely discarded for some of these lineages, since a Jewish ancestry in Lojanos (individuals from Loja province in Ecuador) coming from Europe has been reported (181).

Haplogroups K, L and T

One sample of each ethnic group was classified as KL-M9*(xTat, 92R7, M70) which means that none of these samples belong to haplogroups N, P (Q and R) or T. However, there are more lineages downstream K-M9 that were not tested in this study. The Afro-Ecuadorian KL-M9*(xTat, 92R7, M70) was predicted to belong to haplogroup L (with a prediction score of 100%). This haplogroup is mainly present in India and can also be found in the Middle East, Central Asia, Northern Africa, and Europe along the Mediterranean coast (101). The haplogroup prediction of Mestizo sample presented a different result with a very low prediction score. Nevertheless, both individuals [Afro-Ecuadorian and Mestizo classified as KL-M9*(xTat, 92R7, M70)] were considered to have a European ancestry.

The haplogroup T is present in the Middle East, Africa and Europe (101, 114, 124), and it was only observed in the Mestizo group (1.34%). Although it is one of the most widely dispersed paternal lineages in the world, this haplogroup was reported in other Latin American studies as having a European origin (5, 50, 126). For this reason, we also considered a European origin of the T-lineages found in our study.

Haplogroup Q

The haplogroup Q represents 24.49% and 1.33% of the Mestizos and Afro-Ecuadorian lineages, respectively. The higher frequency of this haplogroup in Mestizos is expected due to the admixture between Europeans and Native Americans. The haplogroup Q can be mainly found in North Eurasia, Siberia and in the Americas and, at lower frequencies, in Europe, East Asia and the Middle East (101). The sub-haplogroup Q-M3 is the major Y-lineage in South Native American populations (151), representing 88.57% and 100% of the Mestizos and Afro-Ecuadorian Q-lineages, respectively.

Haplogroup R

The majority of the Y-lineages found in Mestizo group belong to haplogroup R (38.93%), mainly represented by the sub-haplogroup R-S116*(xU152, M529, M153, M167) (65.52% of the R-lineages). The haplogroup R was also found in 11.33% of the Afro-Ecuadorians chromosomes. This haplogroup represents the majority of the Y-lineages in Europe, but it can also be found in Asia and Africa. In particular, the sub-haplogroup R-V88 can be found in Northern and West-central African populations (48, 49, 83) and represents 5.88% of the Afro-Ecuadorian R-lineages. This means that although most of the R-lineages came from Europe to Ecuador, the particular lineage R-V88 came from Africa.

Ecuadorian origins

Relatively to ancestry of the lineages found (Figure 11), the Mestizo ethnic group is composed by lineages mainly from European origin [haplogroups G, I, J, KL, R (except R-V88), T, E-M81, E-M123 and E-M78 (ancestry inferred with the network presented on Figure 10)], followed by Native American origin (haplogroup Q). In Afro-Ecuadorian population, the results show that most of the lineages have an African origin [haplogroups B, E (except the previously mentioned haplogroups) and R-V88].

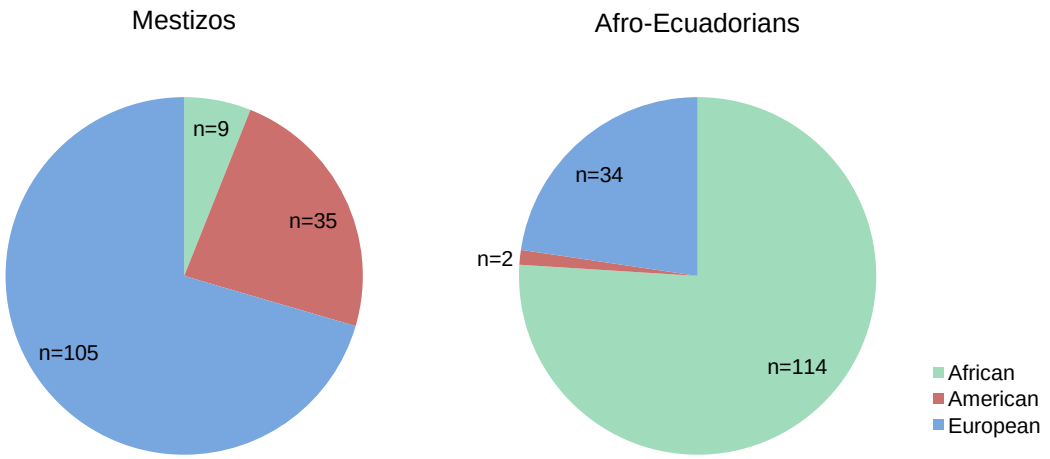


Figure 11: Geographic origin of Ecuadorian lineages found in the two ethnic groups in the present study, according to Y-SNP haplogroup geographic distribution and historical facts related to the population migrations to Ecuador.

Joining both ethnic groups, the Ecuadorian population is characterised by 46.49%, 41.14% and 12.37% of European, African and Native American ancestries, respectively. This result is a little different from that obtained in study of Toscanini et al. (176) (61% European, 34% Native American and 5% African ancestries), which is probably due to an unbalance representation or even an absence of at least one of the three main ethnic groups from Ecuador in the samples of both studies. Nonetheless, it is possible to consider that most Ecuadorians have a European ancestry.

Based in this study, the male gene pool of the Mestizo population is constituted by 70.47% European, 23.49% Native American and 6.04% African lineages. When considering the autosomal data, the Mestizo group is genetically characterised by a higher Native American (more than 59%) (133, 158, 192, 193). According to mtDNA data the Mestizos are almost only characterised by Native American lineages (10, 29). These differences between molecular markers results can be explained by a gender bias in the foundation of the Mestizo group from Ecuador, which is in agreement with historical records of the country (9) and has also been described for other South American populations (37, 38, 50, 69).

The male gene pool of the Afro-Ecuadorian population is mainly constituted by African lineages (76%), followed by European (22.67%) and Native American (1.33%) lineages. The previous autosomal data also reveal a high African ancestry (58.8%) in this group, but a Native American (28%) ancestry higher than the European (13%) (158). Looking to these values, the present study reveals an unexpected high frequency of paternal African ancestry, considering that in most Afro-American populations there was a sex biased matting between European males and African females, responsible for an African biparental ancestry higher than the paternal one. A sex biased matting between African and European men and Native American women is, therefore, necessary to the conciliate our results with those from the autosomal data described by Santangelo et al. (158). However, there are other explanations to these results. In the study from Santangelo et al. (158) the sample size is low and, unlike the present study, it does not include samples from a *Palenque*, which may justify the unexpected difference in the African contribution in both studies. Thus, this results might also be explained due to differences in the population groups sampled. The samples from Santangelo et al. (158) were possibly collected from individuals that were born in a more admixed community than the majority of the samples collected in the present study. In order to better understand if the data presented in this study reveals the above mentioned gender matting bias, with a higher contribution of African and European men and Native American women to Afro-Ecuadorians, it would be important

to also analyse autosomal markers in the Afro-descendant sample from the present study.

Population comparisons

The data obtained in this work for both Y-SNPs and Y-STRs were compared with available results from Latin American, African and European populations (Appendix Tables S5-S7). The analyses performed require a common set of Y-STRs and/or the same resolution of Y-SNP haplogroups, which led to loss some genetic information obtained in the present study. The Y-STRs were reduced to 10 (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437) or 15 (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y GATA H4) depending on the studies used. Moreover, since not all previous studies used to perform comparisons have results for both markers, the populations and/or studies used are different in both comparisons.

Latin America

Y-STRs

Pairwise genetic distances were calculated based on 15 Y-STRs common between our samples and other Latin American populations. The populations used are Colombia (118; data submitted for publication), Peru (18, 134), Brazil (134), Bolivia (134), Chile (175), Argentina (134), Panama (134), Costa Rica (134), Nicaragua (125) and Mexico (111). A European population from Spain (145) and an African population from Equatorial Guinea (83) were also included as external.

Statistical significant R_{ST} values (P<0.05) were obtained for most of the comparisons performed (see Appendix Table S9). All populations present a significant pairwise R_{ST} value with Afro-Ecuadorian group of the present study, being the smallest value obtained between this group and the *Palenque* from Colombia (R_{ST}=0.03269; P=0.00000). These results are expected since the population from *Palenque* was the only one used in the analysis that belongs to the Afro-descendent group. However, has a different degree of admixture (61.05% of the African, 35.79% of European and 3.16% of Native American lineages) (data submitted for publication). The Mestizo group

showed non-significant R_{ST} values with Argentina (R_{ST}=0.00129; P=0.28175) and Nicaragua (R_{ST}=0.00306; P=0.14474).The Argentinean samples used are from a European ancestry and, therefore, is expected that are composed mainly by European lineages as the Mestizo group from the present study. The population from Nicaragua used is also a Mestizo ethnic group with a high European component (~65.4%) and a similar amount of African and Native American components (~15.2% in each group).

In order to simplify the genetic relationships between all Latin American populations analysed, a two dimensional MDS plot was constructed (Figure 12). There is a clear separation between all Latin American populations and the African sample. Indeed, the R_{ST} values between this population and all other populations used were the highest values found in this analysis (see Appendix Table S9). This result was expected with Mestizos and Native American groups since most of their Y chromosomes do not have an African ancestry. However, being the Afro-descendant populations composed mainly by African lineages the huge distance with African sample was not expected. Moreover, it can also be seen in the MDS plot that the African population is closer to the Mestizo group than to the Afro-Ecuadorian group, being the R_{ST} value lower between the African population and Mestizos (R_{ST}=0.92856; P=0.00000) than between Africans and Afro-descendants (R_{ST}=0.93320; P=0.00000). Looking to the left side of the MDS plot is likewise possible to identify a central cluster composed by most of the Latin American populations and the European population and another cluster composed by the Afro-descendants populations. Just one Native American population is separated from the central cluster, with the other two Native populations being aggregated in the cluster with European sample. The high Native American component (~68%) can be the reason of the distance of Peru Native American population (18) on the MDS plot.

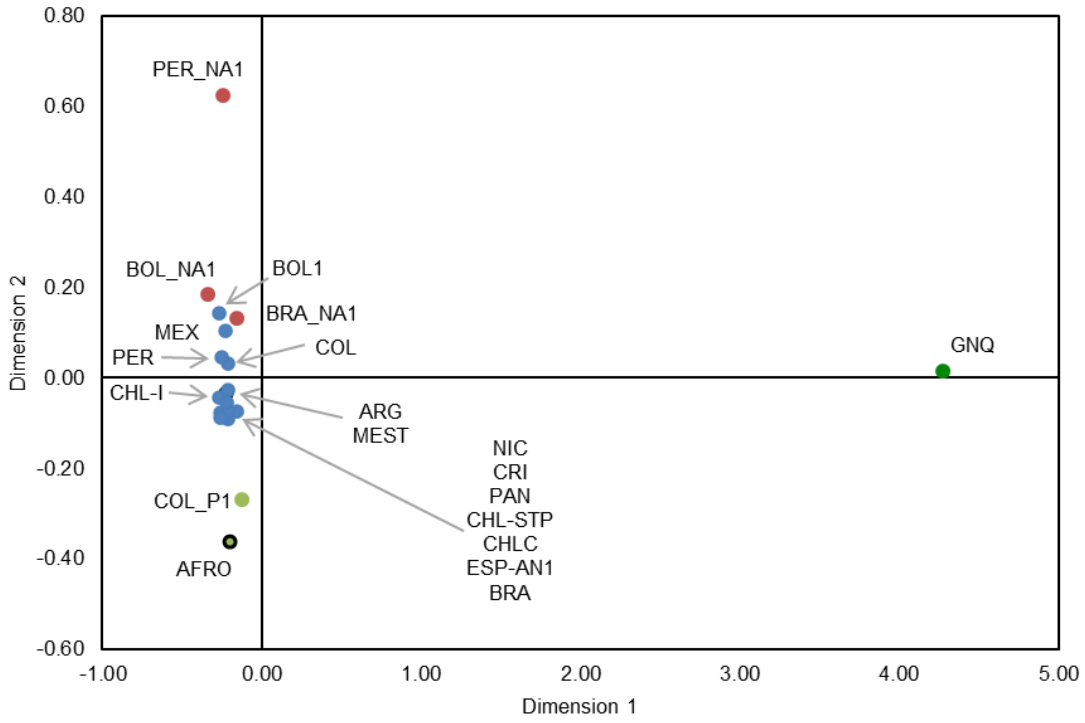


Figure 12: MDS plot of the R_{ST} genetic distances between the two Ecuadorian ethnic groups and other Latin American populations, based on from 15 Y-STR markers (Stress=0.0150922). The present study populations are highlighted with a black circle (Mestizos are under the Argentinean population) (green - Afro-Ecuadorians; dark green - African population blue - Mestizos/Admixed populations and red - Native Americans). (COL- Colombia, PER- Peru, BRA- Brazil, BOL- Bolivia, CHI- Chile, ARG- Argentina, PAN- Panama, CRI- Costa Rica, NIC- Nicaragua and MEX- Mexico, more details about the population codes can be seen in Appendix Tables S5-S7).

Y-SNPs

To perform genetic comparisons at Y haplogroup level there was the need to reduce the high resolution of the present study. 14 different haplogroups (A, DE-YAP*(xM96), E-M96*(xM35), E-M35, F-M213*(xM9), G-M201, I-M170, J-12f2a*(xM172), KLT-M9, Q-M242*(xM346), Q-M346*(xM3), Q-M3*(xM19), R-M207) were compared between three Ecuadorian ethnic groups (present study samples, 13, 74) and other Latina American populations from Colombia (5,123, 189, data submitted for publication), Peru (18), Brazil (126), Argentina (24) and Nicaragua (125). In addition, as in the Y-STRs analysis, a Spanish (31) and Equatorial Guinean (83) populations were included in the pairwise genetic distances analysis.

Most of the F_{ST} genetic distances values obtained were significant (P<0.05) (see Appendix Table S10). Non-significant values were obtained between Mestizos from the present study and Ecuadorian Mestizos from Bafalluy, study (13) (F_{ST}=0.01025; P=0.16969) and also with Nicaragua (F_{ST}=0.00680; P=0.06643), being in accordance with the previous results obtained for Y-STRs. A non-significant F_{ST} genetic distance was also obtained between the Afro-Ecuadorian group and the African population

(F_{ST}=-0.00022; P=0.35185). Therefore, although these two populations have a different haplotype diversity when concerning the haplogroup diversity this difference is not noticed, since both populations are mainly composed by lineages belonging to haplogroup E. This proximity between these two populations is easily observed in the MDS plot presented in Figure 13.

In the visual representation of the pairwise F_{ST} genetic distances through MDS analysis, is possible to observe that the Afro-Ecuadorian and African populations are more distant from the other populations than from the Afro-descendants from Colombia (COL_P1 and COL_P2). Furthermore, is possible to detect one cluster composed by Native American populations that seems to be a very diverse group. Another cluster is composed by the Mestizos, Mixed and European populations and also the Native Americans from Argentina, being this last group very close to Mestizos from Ecuador. Indeed, the Argentinean population, even being Native American populations, is composed by a higher European component than Native American or African components (24, 69), which can explain this proximity.

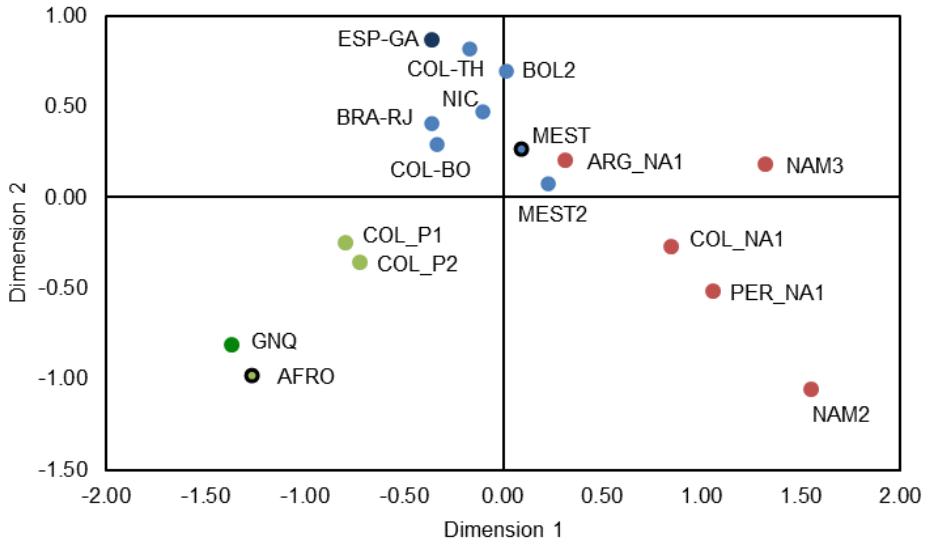


Figure 13: MDS plot of the F_{ST} genetic distances between the Ecuadorian population and other Latin American populations, based on Y haplogroup information (Stress=0.0567491). The present study populations are highlighted with a black circle (green - Afro-Ecuadorians; dark green - African population blue - Mestizos/Admixed populations; dark blue - European population and red - Native Americans). (COL- Colombia, PER- Peru, BRA- Brazil, BOL- Bolivia, ARG- Argentina and NIC- Nicaragua; more details about the population codes can be seen in Appendix Tables S5-S7).

Native American lineages

Exploration on the haplogroup Q was performed in order to understand if there are similarities of the Q lineages between the Ecuadorian population of the present study and other American populations (Native Americans and admixed populations). The Q lineages from previously studied Native American Ecuadorian populations were also used in order to see if there are a specific Native Ecuadorian group closest to populations from the present study. For these analyses our Mestizos and Afro-Ecuadorians Q lineages were pooled together.

R_{ST} pairwise genetic distances were calculated based on 15 Y-STRs common to all populations under analysis [present study; Native American populations from Ecuador (74, 151), Colombia (151), Peru (18, 151), Brazil (151), Bolivia (72, 151) and Argentina (151); and also admixed populations from Colombia (5), Bolivia (37) and Brazil (126)]. The results (Appendix Table S11) show non-significant genetic distances between the Native Y chromosome lineages in Ecuador and those in Native American populations from Colombia (R_{ST}=0.00838; P=0.18404) and Bolivia (R_{ST}=0.00917; P=0.15573), or in an admixed population from Brazil (R_{ST}=0.01479; P=0.20345). Since genetic drift was an important evolution force acting in Native American populations, the F_{ST} method was also used to calculate the genetic distances based on Y-STRs (Appendix Table S12). With this method, the results show non-significant genetic distances between Ecuadorian Native lineages and those in admixed populations from Brazil (F_{ST}=0.00000; P=0.99990) and Bolivia (F_{ST}=0.00037; P=0.51520). Analysing the MDS result, based on R_{ST} pairwise genetic distances, it is possible to observe a proximity between Ecuador and other Native American and admixed populations from Colombia, Brazil and Bolivia (Figure 14a). The native lineages from Ecuador seems also to be closer to those in Native Americans from Peru than from Ecuador. These distances are also evident in the MDS based on F_{ST} pairwise genetic distances (Figure 14b). Although this MDS plot is very different from the one obtained with R_{ST} values, it can be noticed that the populations from Colombia, Brazil, Bolivia and Peru are the closest populations to the present study population. The distance to Native American Ecuadorian populations might be due to the high haplotype diversity found in our admixed Ecuadorian samples. The fact that the Native Ecuadorian groups analysed are more isolated than the other Natives is also a reason for the results obtained.

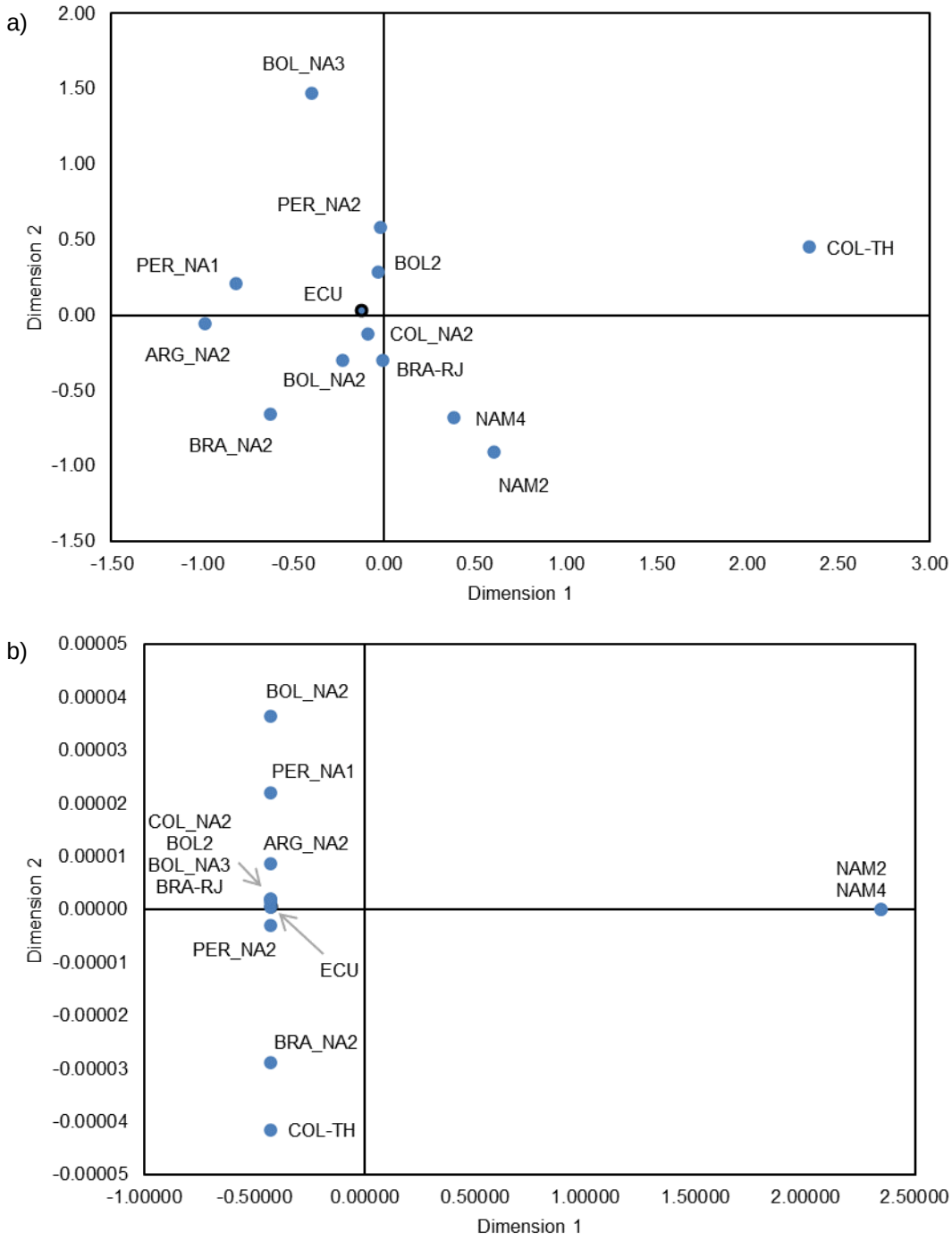
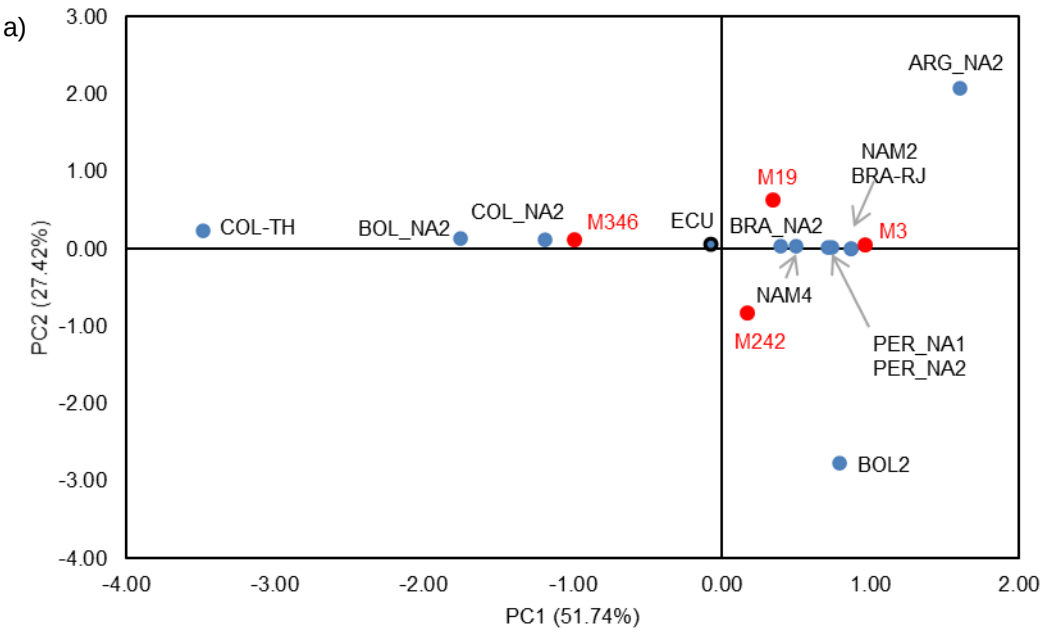


Figure 14: MDS plot of the R_{ST} (a) and F_{ST} (b) genetic distances between the Ecuadorian population and other Latin American populations, based on information of 15 Y-STRs [Stress=0.0841538 (a) and 0.0000049 (b)]. The present study population is highlighted with a black circle. (NAM- Native American from Ecuador, COL- Colombia, PER- Peru, BRA- Brazil, BOL- Bolivia and ARG- Argentina; more details about the population codes can be seen in Appendix Table S5).

F_{ST} genetic distances (Appendix Table S12) and a PC analysis (Figure 15) were performed based on 4 haplogroups [Q-M242*(xM346), Q-M346*(xM3), Q-M3*(xM19) and Q-M19] common between the present study samples and samples from American populations already used in the previous analysis (5, 37, 74, 126, 151). The F_{ST} results show non-significant distances between present study Ecuadorian population and

Native American populations from Ecuador, Colombia, Peru and Brazil and also with admixed populations from Bolivia and Brazil. Analysing the PCA results (Figure 15a) it can be seen that most of the populations, including the present study Ecuadorian population, are in the same line with Q-M346*(xM3) and Q-M3*(xM19) haplogroups. When considering only the PC1 the Ecuadorian population is closer to the haplogroup Q-M242*(xM346) than to the other two previous mentioned, but looking to PC2 and PC3 (Figure 15b) the Q-M242*(xM346) lineage is more distant. The frequency of Q-M19 haplogroup is responsible for the distant position of Native American Argentinean population and the Bolivian population is distant from the others due to the frequency of haplogroup Q-M242*(xM346). Looking to the PCA focused on the first and third principal components (Figure 15b) is more notable the distance between the haplogroups Q-M3*(M19) and Q-M346*(M3), and also that the haploroups Q-M242*(xM346) and Q-M19 are more close to each other as well as the populations that have higher frequencies of these two haplogroups.

a)



(Continues on the next page)

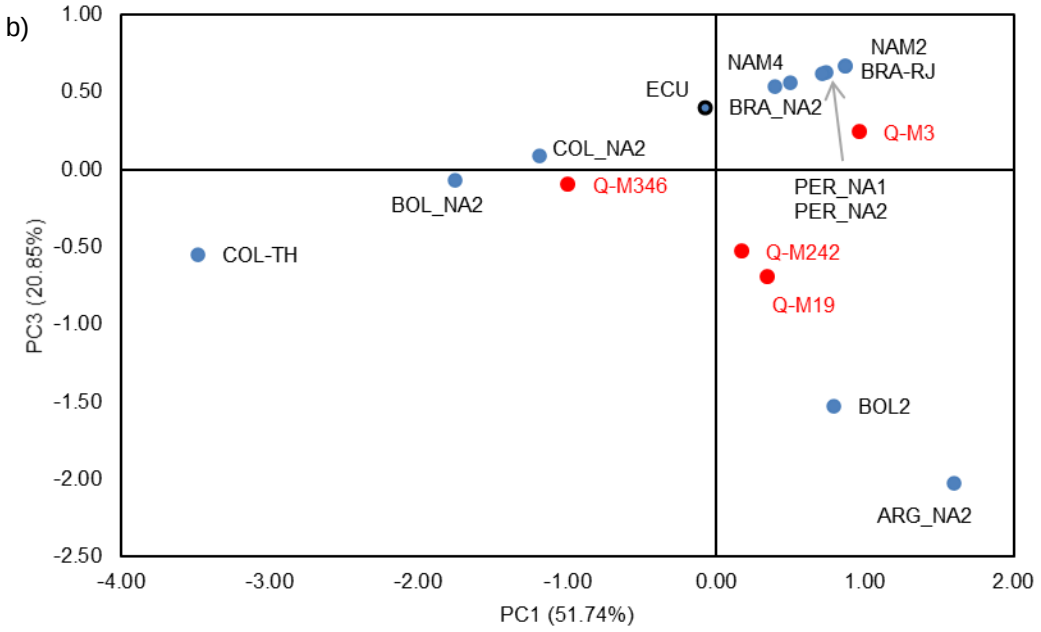


Figure 15: Principal component analysis, focusing on the first and second principal components (a) and on the first and third principal components (b), of Q-M242*(xM346), Q-M346*(xM3), Q-M3*(xM19) and Q-M19 haplogroup frequencies in the present study Ecuadorian population (highlighted with a black circle) and in samples from different South American populations. The first, second and third components represent 51.74%, 27.42% and 20.85% of the total variance, respectively. (NAM- Native American from Ecuador, COL- Colombia, PER- Peru, BRA- Brazil, BOL- Bolivia and ARG- Argentina; more details about the population codes can be seen in Appendix Table S5).

A network focused on Q-M3 lineages from the present study population and Native Ecuadorian populations (13, 74, 151) was built based on 15 Y-STRs (Figure 16). There are other Native American groups living in Ecuador that were not analysed in this network, since was not found genetic data from those populations.

The network result shows that there are many Waorani individuals sharing the same haplotype, being the founder effect very visible. This is expected since this Native ethnic group lives relatively isolated from the remaining Ecuadorian populations. The Quichua group present more diversity of haplotypes than the Waorani group, but there are two haplotypes that are shared between the two ethnic groups. In the network it is possible to highlight that our Ecuadorian population do not share haplotypes with the other Native Ecuadorian populations as well as between themselves, being the haplotypes very dispersed in the network. It can be seen one haplotype that is only one step different from a Quichua haplotype, which might suggest a contribution from the Quichua group to the Mestizo group from Ecuador. The unique haplotype found in the Ecuadorian population from the present study that corresponds to the haplogroup Q-Z19483 (downstream to Q-M3) is highlighted in Figure 16 with an arrow. Since there is only one individual with this haplogroup and the other studies did not have analysed this Y-SNP, more studies are needed in order to understand if this haplogroup is

specific of the Mestizo group and/or any Native Ecuadorian group. The Shuar samples used are not enough to conclude if this group present any relationship with other ethnic groups.

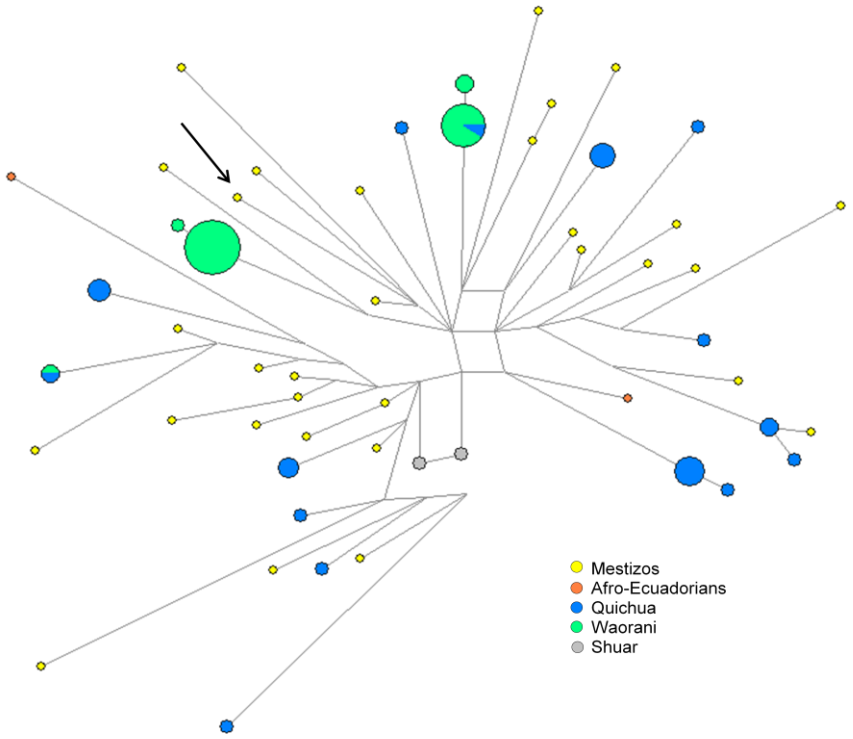


Figure 16: Relationships between Y haplotypes within Q-M3 haplogroup based on information of 15 Y-STRs. Samples from present study [Mestizos (n=31) and Afro-Ecuadorians (n=2)] and Quichua (n=54), Waorani (n=66) and Shuar (n=4) Native American populations were analysed. The arrow indicates the sample belonging to the Q-Z19483 sub-haplogroup.

Taking into consideration the results from previous analyses, the distribution of Q-lineages in the Ecuadorian population seems to be more similar to those in South American populations, such as Brazil and Bolivia, than in Native America Ecuadorian populations. Moreover, the high diversity of haplotypes present in our sample (illustrated in the network) suggests that there are more Y-SNPs downstream to M3 that were not analysed and should be explored. These other Y-SNPs could allow to a better differentiation of the Q-lineages, which would be important to increase the knowledge of the relationship between South American populations. Indeed, it had already been reported markers downstream to M3 (e.g. SA01, M557, MG2) in some Native American populations, that can help to differentiate Q-M3 lineages (21, 73, 99). Two of these markers were also analysed in the present study and the other marker (Q-MG2) would also be important to be analysed in a future study since it is possibly restricted to a Native American Ecuadorian population (lowland Kitchwa; 73).

African ancestry

Genetic comparisons between the African lineages present in Ecuador and in African populations were performed in order to infer the origin of the African lineages present in Ecuador. The data of both ethnic groups from Ecuador were pooled together since they probably have the same African origin. However, it was not possible to confirm if there were no significant differences between the two groups due to the small sample size of the African lineages in the Mestizos.

It was previously mentioned that R_{ST} was developed to perform calculations with microsatellite data and so F_{ST} is normally used only with SNPs data. However, since the R_{ST} method undervalues the importance of genetic drift in sub-Saharan populations, it has been proposed that the weight given to mutations by this method is not in agreement with the demographic events involved in Y-STR evolution in these populations (35). Due to that, both methods were used in this section for calculations with Y-STR data.

R_{ST} and F_{ST} pairwise genetic distances were calculated based on 10 Y-STRs common between Ecuadorian samples and African populations from Guinea Bissau (152), Ivory Coast (66), Benin (66, 107), Cameroon (23), Equatorial Guinea (83), Gabon (23), Angola (44) and Mozambique (155). These populations were selected to represent the African regions involved in the trans-Atlantic slave trade.

The R_{ST} results show non-significant differences between Ecuadorian population and populations from Benin (107) (R_{ST}=0.00319; P=0.20681), Equatorial Guinea (R_{ST}=0.00796; P=0.09276) and Gabon (R_{ST}=-0.00183; P=0.70825) (Appendix Table S13). All distances obtained with the F_{ST} method between the Ecuadorian population and Africans were significant (Appendix Table S14). However, the results of both methods seem to be concordant, since F_{ST} smallest distances found were between the present study Ecuador and the same three African populations that present non-significant differences in the R_{ST} analysis.

Observing the genetic distances values it can also be concluded that the African lineages in the Ecuadorian population did not suffer genetic drift, since the values of distances between Ecuadorian population and all other African populations are in a similar range of values of the distances found only between the African populations. A representation of these results can be seen in Figure 17. In the MDS from R_{ST} results is clear the proximity between Ecuador, Benin, Equatorial Guinea and Gabon, but it can also be observed that Guinea-Bissau is also close to Ecuador. In the MDS from F_{ST}

genetic distances the populations are more dispersed in the plot, but populations from Benin, Equatorial Guinea and Gabon seem also to be the closest populations to Ecuador.

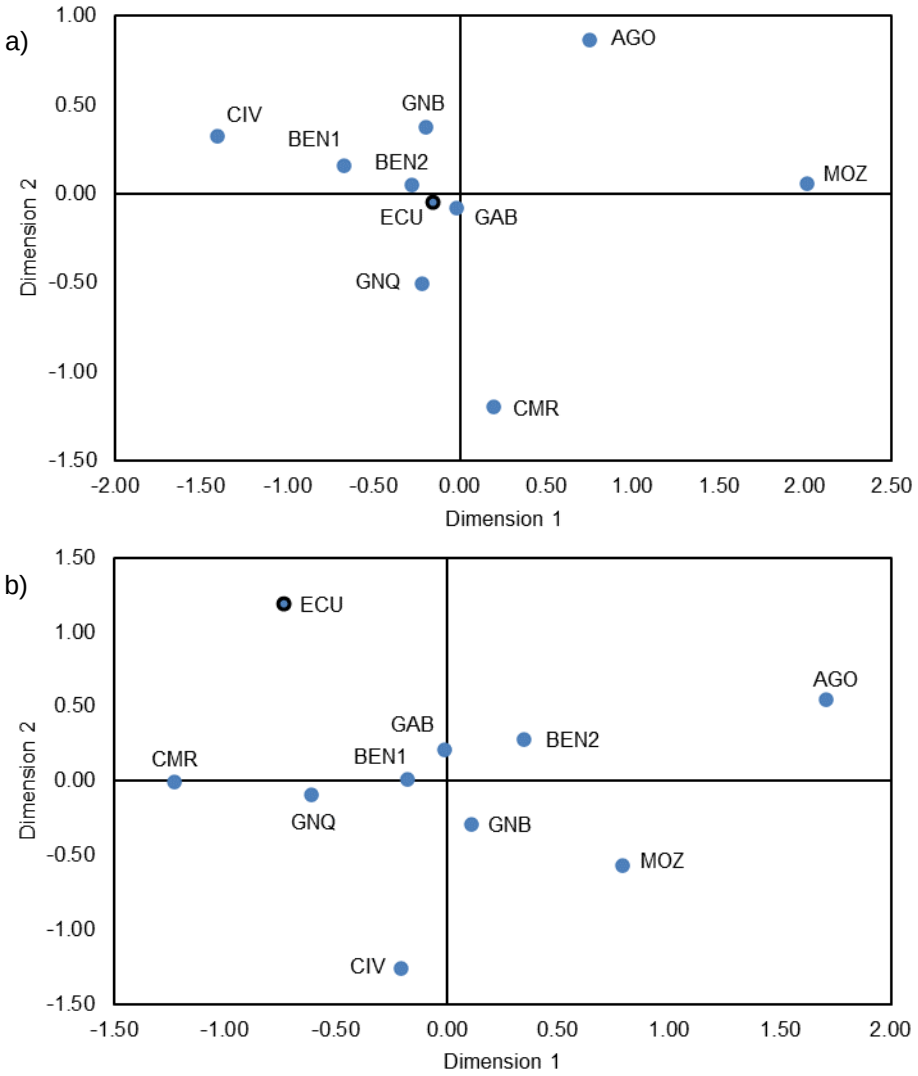


Figure 17: MDS plot of the R_{ST} (a) and F_{ST} (b) genetic distances between the Ecuadorian population and African populations, based on 10 Y-STR loci within African lineages [Stress=0.0662166 (a) and 0.0903862 (b)]. The present study population is highlighted with a black circle. (GAB-Guinea Bissau, CIV-Ivory Coast, BEN-Benin, CMR-Cameroon, GNQ-Equatorial Guinea, GAB-Gabon, AGO-Angola and MOZ-Mozambique; more details about the population codes can be seen in Appendix Table S6).

F_{ST} pairwise genetic distances analysis based on haplogroups was also performed (Appendix Table S15). In this analysis data from 7 African haplogroups [A, B-M60, E-M96*(xM33, M2, M35, M75), E-M33, E-M2, E-M35*(xM78, M81, M123), E-M75; the haplogroup R-V88 was not included due to the lack of resolution in most studies analysed] were compared between Ecuadorian population and populations from Guinea-Bissau (152), Ivory Coast (66), Benin (66, 107), Cameroon (23), Equatorial

Guinea (83), Gabon (23) and Mozambique (155). Non-significant distances were obtained between Ecuadorians and populations from Ivory Coast (F_{ST}=0.01284; P=0.10088), Benin (F_{ST}=0.00789; P=0.09326 and F_{ST}=0.00786; P=0.11672) and Equatorial Guinea (F_{ST}=-0.00025; P=0.36600). In the graphic representation (Figure 18) of these results, it can be seen the proximity of these populations. Furthermore, the populations from Guinea-Bissau and Cameroon are more distant from the aforementioned populations than Gabon and Mozambique.

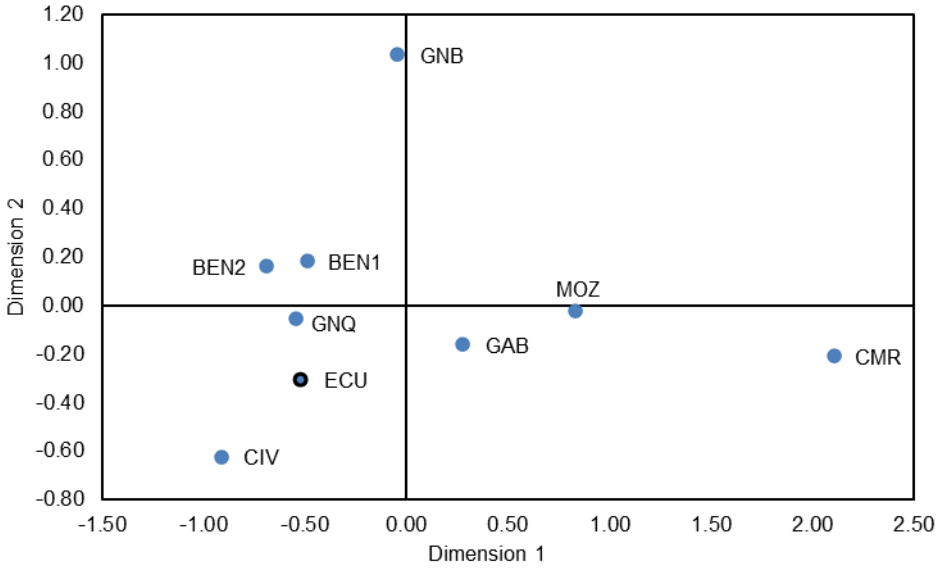


Figure 18: MDS plot of the F_{ST} genetic distances between the Ecuadorian population and African populations, based on African Y haplogroups information (Stress=0.0126493). The present study population is highlighted with a black circle. (GAB-Guinea Bissau, CIV-Ivory Coast, BEN-Benin, CMR-Cameroon, GNQ-Equatorial Guinea, GAB-Gabon and MOZ-Mozambique; more details about the population codes can be seen in Appendix Table S6).

A PCA based on haplogroup frequencies was performed in order to understand which African lineages were influencing the relationship between populations (Figure 19).The first and second components represent 45.27% and 28.52% of the total variance, respectively. The distribution on the plot of the different populations is similar to the results obtained with F_{ST} genetic distances. Ecuador is close to Ivory Coast, Benin and Equatorial Guinea due to the high frequencies of haplogroup E-M2 and the main reason that separates Gabon and Mozambique from this cluster is the frequency of haplogroup E-M75. Guinea-Bissau is very distant since this population present higher frequencies of the haplogroups A (mainly A-M31), E-M33 and E-M35*(xM78, M81, M123) than the other analysed populations. The position of Cameroon is mainly due to the frequencies of haplogroups B (mainly B-M150) and E-M96. It should be highlighted that in the different studies the resolution of the sub-lineages downstream to haplogroups A and B was different and so multiple sub-lineages were included in just

one haplogroup and for that reason the several Y-SNPs which define these haplogroups were not indicated.

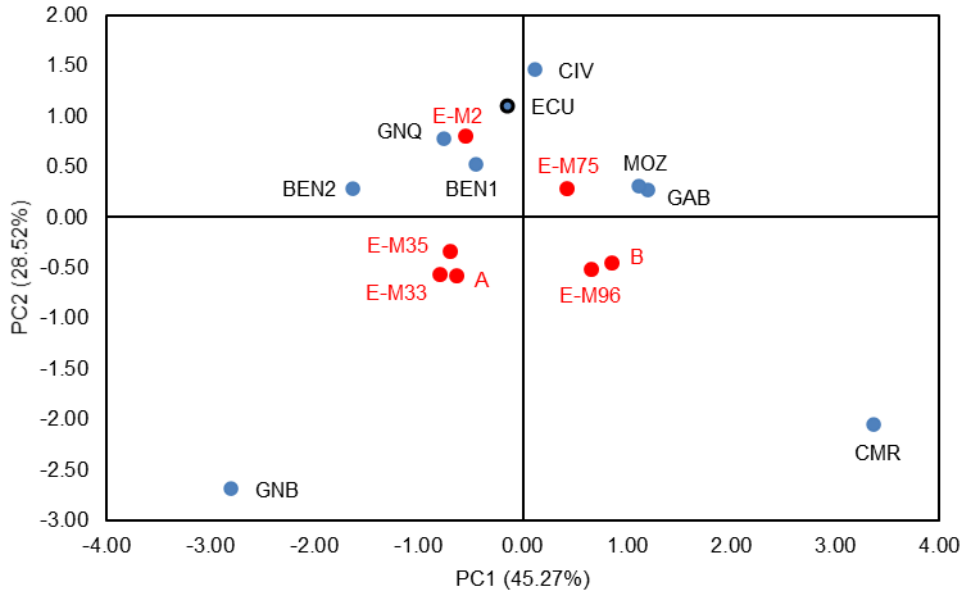


Figure 19: Principal component analysis of A, B, E-M96*(xM33, M2, M35, M75), E-M33, E-M2, E-M35*(xM78, M81, M123) and E-M75 haplogroup frequencies in the present study (highlighted with a black circle) and in populations from different African countries (GAB-Guinea Bissau, CIV-Ivory Coast, BEN-Benin, CMR-Cameroon, GNQ-Equatorial Guinea, GAB-Gabon and MOZ-Mozambique; more details about the population codes can be seen in Appendix Table S6).

The results from the previous analyses based on Y-STR and Y-SNP data seem to indicate different countries in the origin of the African lineages present in Ecuador. The more probably origins are from African countries between the Ivory Coast and Gabon, which is in agreement with historical records and genetic data from Latin American populations (50, 85, 116). According to the results obtained, Guinea-Bissau might also be a possible source in the African gene pool from Ecuador.

Haplogroup E

As previously mentioned, the haplogroup E is very common in African populations, mainly the sub-haplogroup E-M2 that is mainly distributed in Bantu people from sub-Saharan Africa. This sub-clade was also the most frequent in Ecuador. In order to identify the African origin of the E-M2 lineages three networks were built. The populations used in these networks were from Guinea-Bissau (152), Ivory Coast (66), Benin (107), Equatorial Guinea (83) and Mozambique (155). The networks were constructed based on 10 Y-STRs common between samples.

The network based on data from E-M2*(xU209, M191) lineages (Figure 20) shows a high diversity of haplotypes which indicates that is not a recent lineage and so there was time enough to obtain more diversity. Shared haplotypes were observed between populations from Ecuador, Guinea-Bissau, Ivory Coast, Benin and Equatorial Guinea. Three Ecuadorian samples share haplotypes exclusively with Guinea-Bissau and also other Afro-Ecuadorian lineages seem to be close with Guinea-Bissau, indicating that this last country can be the probable origin for that haplotypes. Moreover, it can be observed that the samples from Ecuador also show some proximity with Equatorial Guinea. This network does not show any relationship between Ecuador and Mozambique, but since only two samples from Mozambique [both belonging to the same haplogroup downstream to E-M2*(xU209, M191)] were available it cannot be discarded the presence of E-M2*(xU209, M191) lineage from this African country in Ecuador.

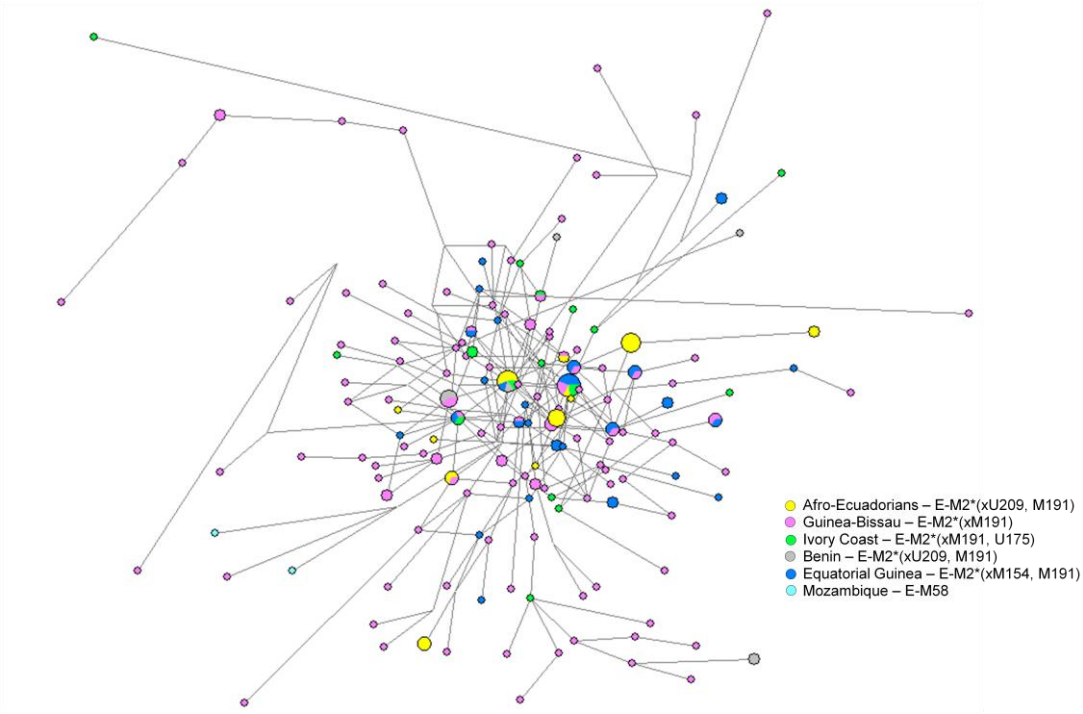


Figure 20: E-M2* median-joining network based on 10 Y-STR information. Samples from present study [Afro-Ecuadorians (n=28)], Guinea-Bissau (n=120), Ivory Coast (n=20), Benin (n=8), Equatorial Guinea (n=37) and Mozambique (n=2) were analysed.

In the second network analysis, which focus the E-M191 lineages, shared haplotypes between Ecuador, Ivory Coast, Benin and Equatorial Guinea were also observed (Figure 21). This observation could mean that even though the study of Equatorial Guinea did not genotype Y-SNPs downstream to M191, some haplotypes possibly belong to the sub-haplogroup E-U174 (sub-lineage inside of E-M191 haplogroup).

Nevertheless, this network also show shared haplotypes that are not from the same haplogroup, which means that the comparison using these 10 Y-STRs (DYS19, DYS389 I, DYS389 II, DYS390, DYS391, DYS392, DYS393, DYS438, DYS439, DYS437) might not be enough to distinguish these two lineages. It is also worth mentioning that four Ecuadorian samples share haplotypes exclusively with Benin and other 9 Ecuadorian samples share haplotypes only with Equatorial Guinea. Mozambique does not seem to be a possible origin to E-M191 lineages, since the haplotypes found in this population are distant from the Ecuadorian haplotypes.

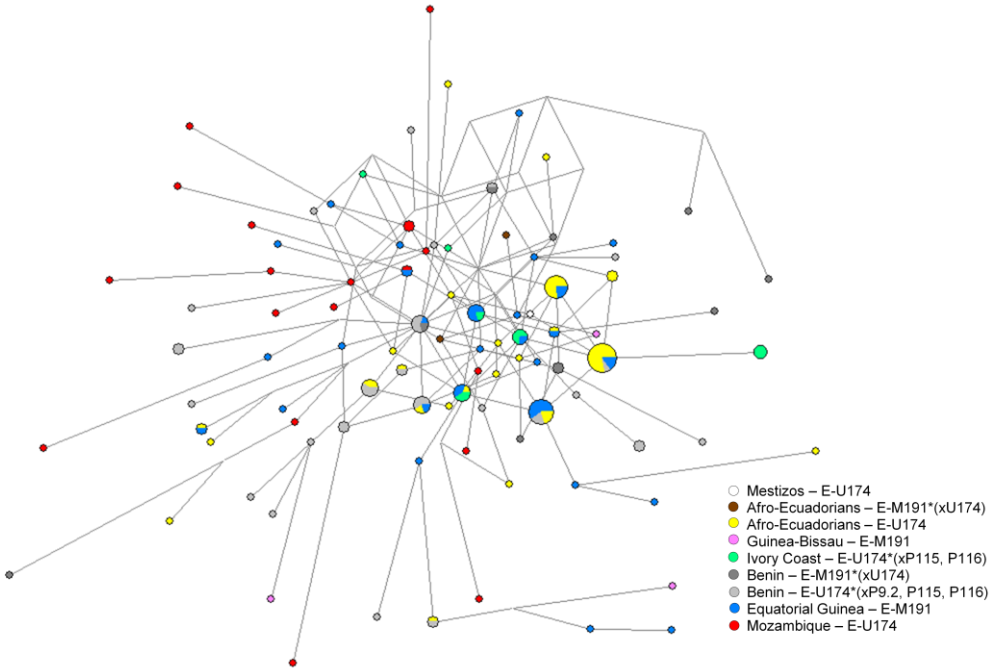


Figure 21: E-M191 phylogenetic network based on 10 Y-STRs information. Samples from present study [Mestizos (n=1) and Afro-Ecuadorians (n=44)], Guinea-Bissau (n=3), Ivory Coast (n=11). Benin (n=44), Equatorial Guinea (n=39) and Mozambique (n=19) were analysed.

The third network of E-M2 lineages was focused on the sub-haplogroup E-U209 (Figure 22). In this network, the comparison using only 10 Y-STR data seems not to be enough to distinguish the E-U209, E-U290 and E-M154 sub-lineages. In this network is clear the existence of a central haplotype with lineages radiating from it. This haplotype is shared between different lineages belonging to Ecuador, Ivory Coast and Benin. In addition, there are other shared haplotypes between Ecuadorian and African populations, two of them are shared only between Ecuador and Ivory Coast and one other is common only between Ecuador and Benin. This network is the unique one that seems to indicate Mozambique as a possible source to the Ecuadorian E-M2 lineages, since there is a haplotype shared between the samples of these two countries (and

also with Benin) and there are some Ecuadorian samples in the same branch or close to samples from Mozambique.

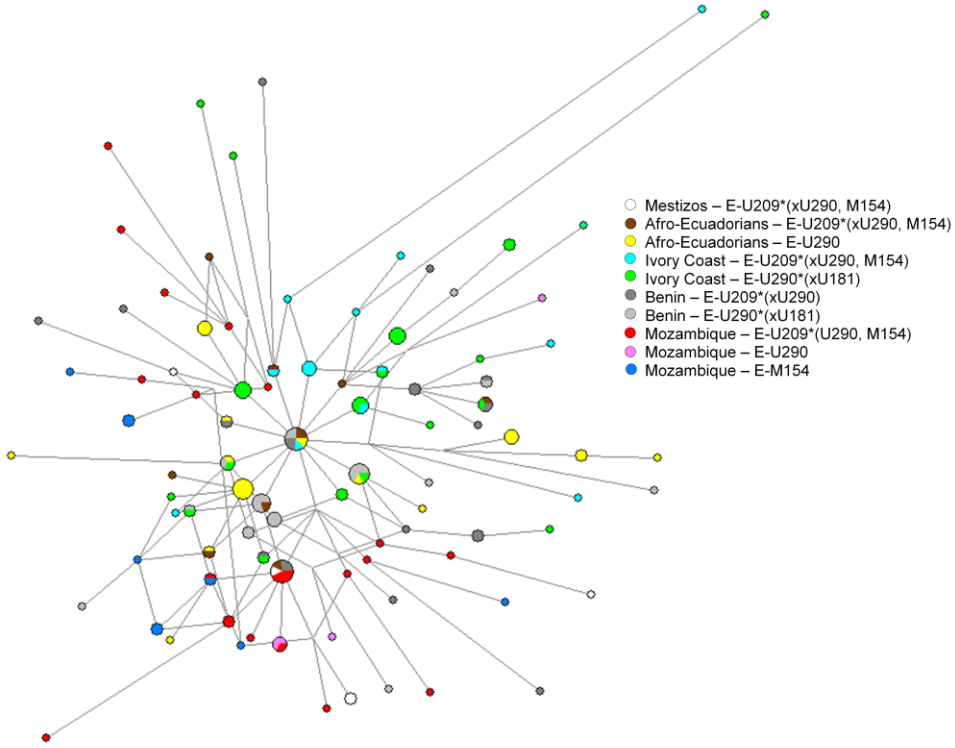


Figure 22: Relationships between Y haplotypes within E-U209 haplogroup based on information of 10 Y-STRs. Samples from present study [Mestizos (n=5) and Afro-Ecuadorians (n=33)], Ivory Coast (n=43). Benin (n=43) and Mozambique (n=35) were analysed.

One more phylogenetic network was constructed in order to evaluate the relationships between E-M75 lineages, which together with the E-M2 lineages represent more than 97% of the African E lineages in the Ecuadorian samples from this study (Figure 23). This network was constructed also based on 10 Y-STRs common between all samples [from the present study and samples from Guinea-Bissau (152), Ivory Coast (66), Benin (107), Cameroon (23), Equatorial Guinea (83), Gabon (23), and Mozambique (155)]. One shared haplotype was observed between Ecuador and Gabon (the two haplotypes does not belong to the same haplogroup. However, the Gabon samples were not typed to the Y-SNP M85). Taking this into consideration and the fact that some haplotypes from different haplogroups are very close in the network it can be concluded that some of the samples that were not typed for markers downstream to M75 Y-SNP probably belong to the sub-haplogroup E-M85. This network also shows that all Ecuadorian samples are close to samples from Gabon. However, it is important to refer that the sample size from Gabon is very big when compared with the other populations analysed and so this proximity should be carefully interpreted. It should

also be highlighted that two haplotypes from Ecuador and Ivory Coast are separated only by one step.

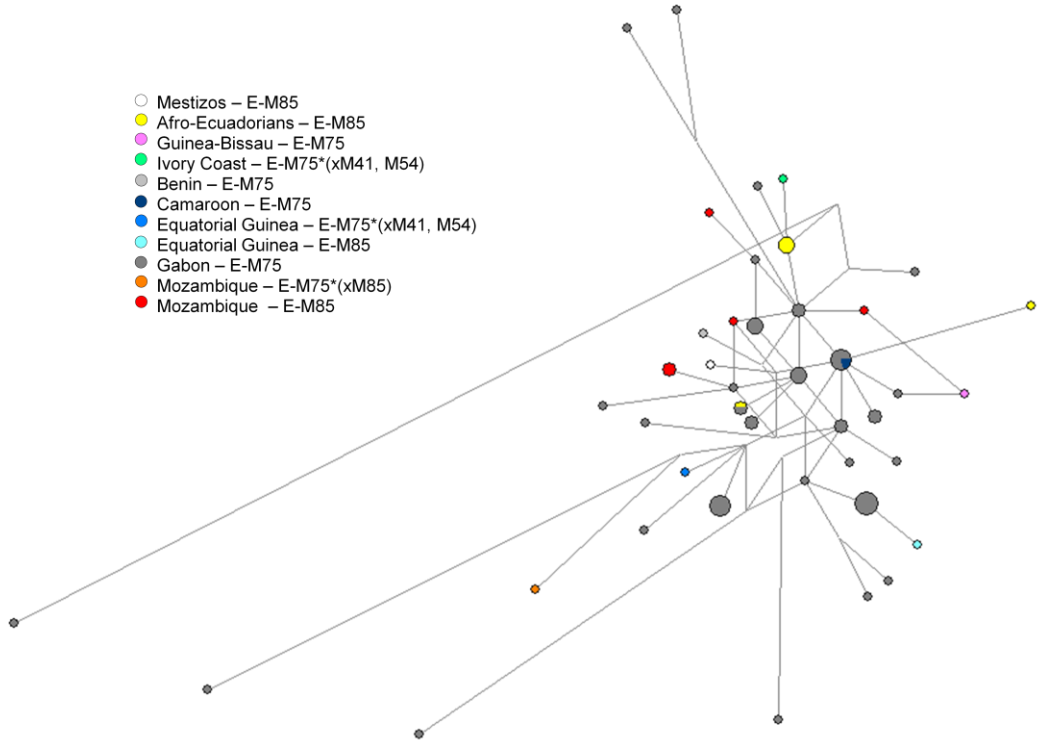


Figure 23: E-M75 Phylogenetic network (constructed based on 10 Y-STR loci). Samples from present study [Mestizos (n=1) and Afro-Ecuadorians (n=5)], Guinea-Bissaou (n=1), Ivory Coast (n=1). Benin (n=1), Camaroon (n=1), Equatorial Guinea (n=2), Gabon (n=49) and Mozambique (n=7) were analysed.

Furthermore, as the first network, the three last networks show a high haplotype diversity indicating that these lineages are ancient. Nevertheless, it seems that there was not time enough and/or, as already mentioned, the number of Y-STRs used in analyses was not enough to allow a clear genetic separation between the sub-lineages analysed in each network.

The total networks constructed inside of haplogroup E indicate that the African E lineages present in Ecuadorian populations have multiple ancestries, which is in agreement with the previous results, presented in this study, about the ancestry of all African lineages from Ecuador. Thus, populations from regions between Ivory Coast and Gabon and also from Guinea-Bissau seem to have contributed to the African gene pool present in Ecuador. A contribution from Mozambique is also possible to the E-U209*(xU290, M154) lineages found in Ecuador. It is worth mentioning that Angola was not included in any network since was not found data with the same resolution of the other studies used. Nevertheless, this country was also an important source of slaves

during the colonial period and therefore, it would be important in the future analyses if populations from Angola had also contributed to the E-lineages found in Ecuador.

European ancestry

In this section, only the European haplogroups will be considered. The samples of the two groups from Ecuador were pooled together after confirm that these two ethic groups have no statistical significant differences between them (R_{ST}=0.00812; P=0.16038).

R_{ST} and F_{ST} pairwise genetic distances were calculated based on 15 Y-STRs and 11 haplogroups [E-M78, E-M81, E-M123, G-M201, I-M70, J-12f2*(xM172), K-M9, R-M207*(xSRY1083.2, M269), R-SRY1083.2, R-M269], respectively. The comparisons were performed with data from Portugal (4, 22, 65, 78, 124), Spain (4, 6, 31, 65, 90, 145, 168), United Kingdom (113), France (54), Italy (25), Germany (142), Poland (142), Bulgaria (100) and Turkey (42) (note that not all of these populations have data available for both Y-STRs and Y-SNPs).The several regions belonging to the same country and study were pooled when no significant differences were found between them. All these populations were selected in order to represent different European regions, mainly the ones that were reported as the origin of Europeans in Ecuador, and also one population (Turkey) was included to represent the Middle East.

Concerning the calculations based on the Y-STRs information (Appendix Table S16) the Ecuadorian population presented non-significant differences with Portugal (R_{ST}=-0.01133; P=0.89912) and Northwest Italy (R_{ST}=0.00341; P=0.10544). Due to the small sample size of the Portuguese population used in this analysis, the relationship between Ecuadorian and Portuguese populations should be carefully interpreted. The Spanish population, mainly Zamora (R_{ST}=0.00821; P=0.02247) and Andalusia (R_{ST}=0.01127; P=0.02287), is also close to Ecuadorians although the P-value was significant. This proximity can be seen in the MDS plot presented in Figure 24 where the Ecuadorian population is integrated in a cluster composed by the Iberian Peninsula and Northwest Italy. Germans and French are the closest populations to this cluster. Looking to Italy country is possible to observe that show high differences between the several geographic regions, being the Northwest region the most important to the European gene pool in Ecuador.

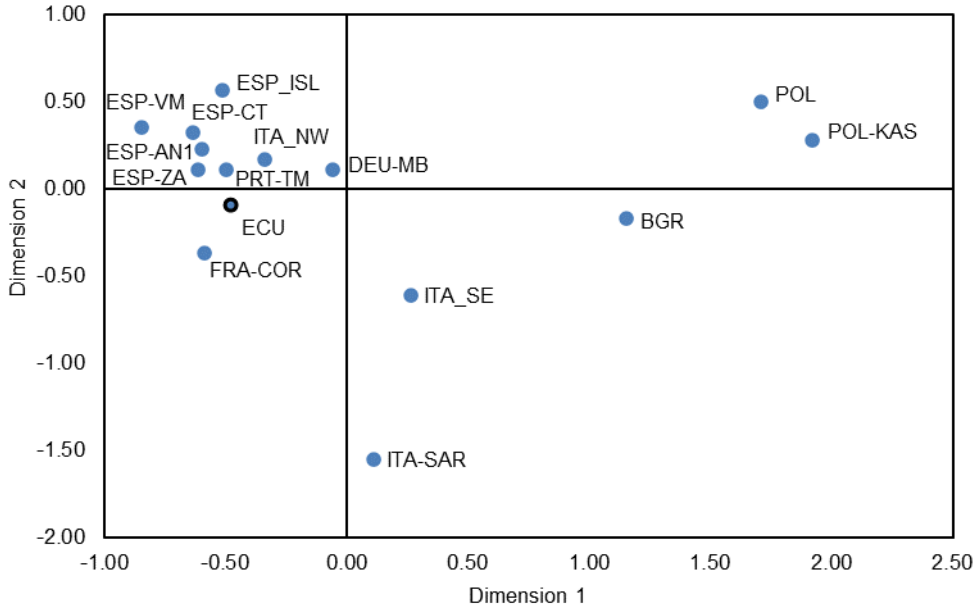


Figure 24: MDS plot of the R_{ST} genetic distances between the Ecuadorian and European populations, based on 15 Y-STR markers within European lineages (Stress=0.00369814). The present study population is highlighted with a black circle. (PRT- Portugal, ESP- Spain, FRA- France, ITA- Italy, DEU- Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix Table S7).

Observing the results obtained with the Y-SNP data (Appendix Table S17), similar conclusions were acquired. The Ecuadorian population presents non-significant values with Portugal from three different studies (values between F_{ST}=-0.00384; P=0.65162 and F_{ST}=0.00437; P=0.13811) and Northwest Italy (F_{ST}=0.00310; P=0.14246). Focusing in the colonization history of Ecuador it was expected a closer relationship with Spain than Portugal. However, observing the F_{ST} values between the Portuguese and Spanish (except Basque Country) populations used in this analysis, it can be noticed a similarity in male lineages between these two countries. This close relationship can also be noticed in the MDS presented in Figure 25. The Iberian population (except Basque Country) form a cluster with Ecuador and Northwest Italy that is surrounded by populations from Basque Country (Spain), Germany, France, Southeast Italy and United Kingdom. This last country also presents non-significant differences (F_{ST}=0.01005; P=0.20830) with Ecuador, however, due to the small sample size of the British population analysed (n=19) this result should be more explored in future.

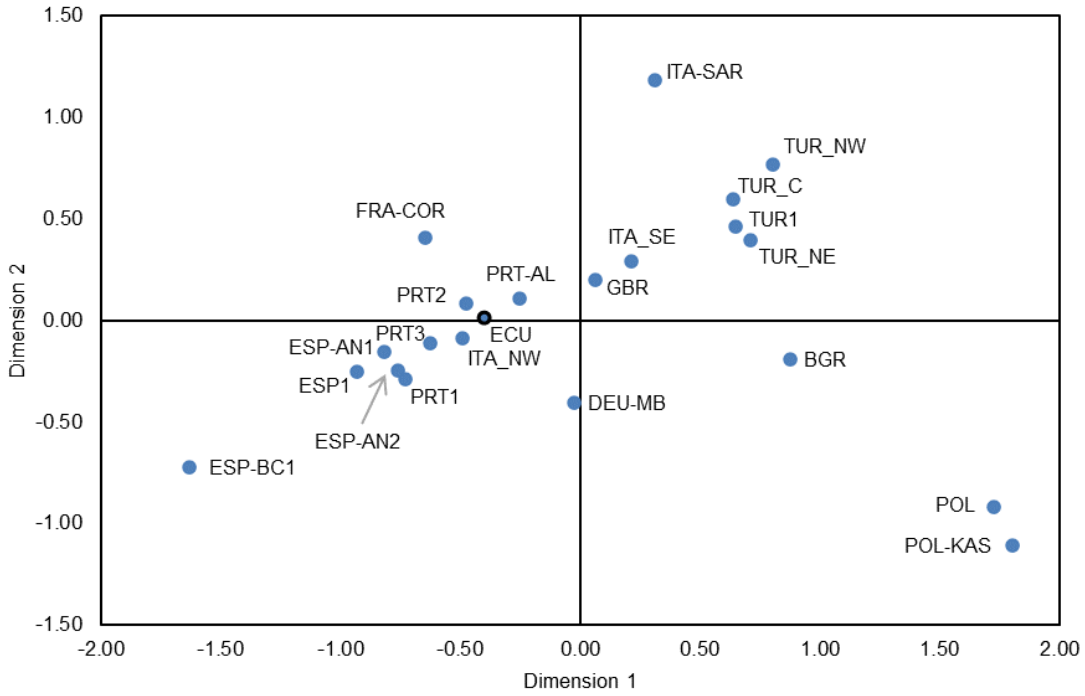


Figure 25: MDS plot of the F_{ST} genetic distances between the Ecuadorian and European populations, based on information from European Y-SNP haplogroups (Stress=0.0784008). The present study population is highlighted with a black circle. (PRT- Portugal, ESP- Spain, GBR- United Kingdom, FRA- France, ITA- Italy, DEU- Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix Table S7).

A PCA based on haplogroup frequencies was performed in order to understand which European lineages were influencing the relationship between populations (Figure 26). The first and second components represent 51.15% and 22.44% of the total variance, respectively. The result obtained supports previous results showing a closer relationship between Ecuadorian, Iberian (except Basque Country and Alentejo) and Northwest Italy populations. The position of the Iberian population is very influenced by the haplogroup R that seem also to be important for Polish populations. Moreover, there is a Portuguese population very close to the Ecuadorians, mainly due to the frequencies of the E and I lineages. Considering only the first component of the MDS, haplogroup I is influencing the proximity between Ecuadorian and Northwest Italian populations. When only the second component is considered Ecuadorian population is close to the British and Northeast Turkish populations, mainly due to the frequencies of haplogroup E.

Given all information from the previous three analyses, the importance of the Iberian (mainly Western and Southern regions) and Italian populations to the European gene pool present in Ecuador is very clear. It also suggests a possible contribution from Germany and Great Britain, but not so important as the two aforementioned populations. These results are in accordance to the historical data of the country, since Ecuador was a Spanish colony and a few centuries after the arrival of the Spaniards there were other European migrations, mainly from Italy, Germany and Great Britain (9).

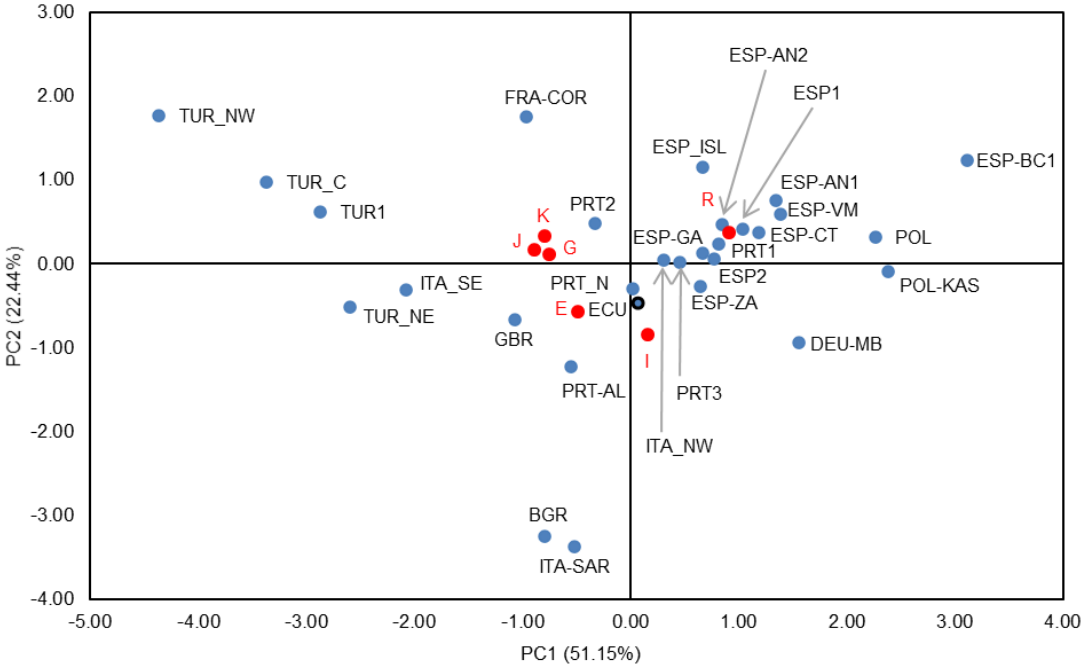


Figure 26: Principal component analysis of E (only European lineages), G, I, J, K and R (except R-V88) haplogroup frequencies in the present study Ecuadorian population (highlighted with a black circle) and in samples from different European populations (PRT- Portugal, ESP- Spain, GBR- United Kingdom, FRA- France, ITA- Italy, DEU- Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix Table S7). Samples from the same country and study were pooled when no statistically significant differences were found between them.

Haplogroup R

As already mentioned, the haplogroup R is very frequent in Europe, in particular the haplogroup R-M269 and its sub-lineages. This haplogroup presents a cline distribution in Europe with high frequencies in Western populations decreasing to the East (33, 122). The sub-haplogroup R-U106 is more frequent in the Central and Eastern portions of the continent, whereas the R-S116 is more common in West European populations (33, 122). Downstream to R-S116 the sub-lineage R-S116*(xU152, M529) is very frequent in Iberian Peninsula, being the sub-lineage R-DF27 near-specific of that region (33, 122, 169, 179, 182). The sub-lineage R-U152 is more common in

Switzerland, Italy, France and Western Poland and R-M529 is more frequent in British Isles (33, 122). According to the results of the present study, the haplogroup R is also very frequent in Ecuador, mainly the sub-haplogroup R-S116*(xU152, M529, M153, M167).

Pairwise F_{ST} genetic distances analysis was performed based on 5 different R haplogroups [R-M269*(xU106, S116), R-U106, R-S116*(xU152, M529), R-U152, R-M529] (see Appendix Table S18). The samples from this study were compared with populations from Portugal (33, 122, 179), Spain (122, 170), Ireland (122, 179), Scotland (33), England (122), France (54, 122, 179), Italy (25, 122), Netherlands (122), Germany (122, 142), Poland (142) and Turkey (122). An MDS plot based on the F_{ST} genetic distances obtained was also constructed (Figure 27). When it is considered only the R-lineages, the distance between some populations is different from the previous results [based on all European lineages (Figure 25)]. Therefore, analysing only the R-lineages is possible to increase the differentiation between geographical regions of the Europe. Ecuadorian population is in a cluster with the Iberian population, presenting non-significant differences with Portugal from two different studies (F_{ST}=-0.00277; P=0.48094 and F_{ST}=0.01016; P=0.15434) and Spain from different regions and studies (values between F_{ST}=-0.00883; P=0.85694 and F_{ST}=0.00605; P=0.19523). The cluster of South Europe is composed by Italy and Corsica (France) and, this is the closest group to the Turkish population. In the fourth quadrant are the Northern European populations [Ireland, Scotland, England, France (Brest), Netherlands, Germany and Poland].

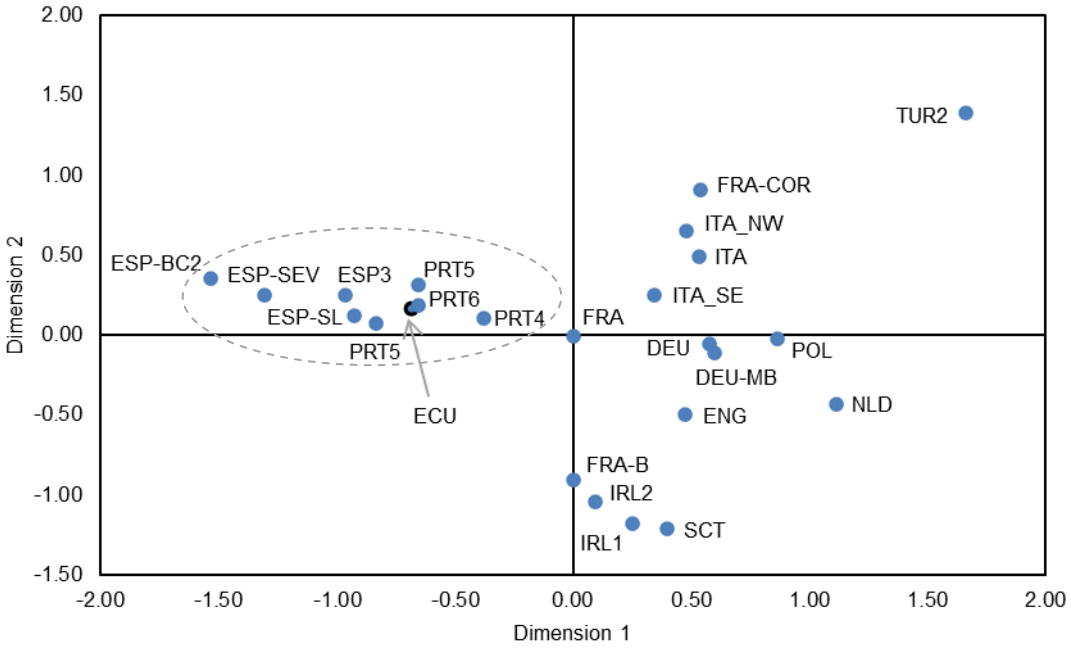


Figure 27: MDS plot of the F_{ST} genetic distances between the Ecuadorian and European populations, based on information of haplogroup R (Stress=0.1116117). The present study population is highlighted with a black circle. (PRT- Portugal, ESP- Spain, IRL- Ireland, SCT- Scotland, ENG- England, FRA- France, ITA- Italy, NLD- Netherlands, DEU- Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix Table S7).

A PC analysis was also performed and the result obtained shows a similar distribution on the plot of the different populations as previously obtained when F_{ST} pairwise genetic distances analysis was realized (Figure 28). The first and second components represent 33.81% and 28.91% of the total variance, respectively. Observing the results, it is clear that the Iberian populations present high frequencies of the haplogroup R-S116*(xU152, M529), which is also the most frequent in the Ecuadorian populations of the present study. The South European and Turkish populations present higher frequencies of haplogroups R-M269*(xU106, S116) and R-U152 than the other populations used in this analysis. Moreover, these two haplogroups are also important to the French population (FRA and Corsica). In this French population (FRA) the other haplogroups seem to be almost equally important since this population is positioned almost in the centre of the plot. The Northern European populations are divided into two clusters, being one of them composed by populations from Ireland, Scotland and Northwest France (Brest) that have higher frequencies of the haplogroup R-M529 than the other populations. As expected the cluster composed by British, Dutch, Polish and German populations has its position on the plot mainly due to the frequencies of the haplogroup R-U106. In England is also to observe a high contribution of R-M529 to its position in the plot.

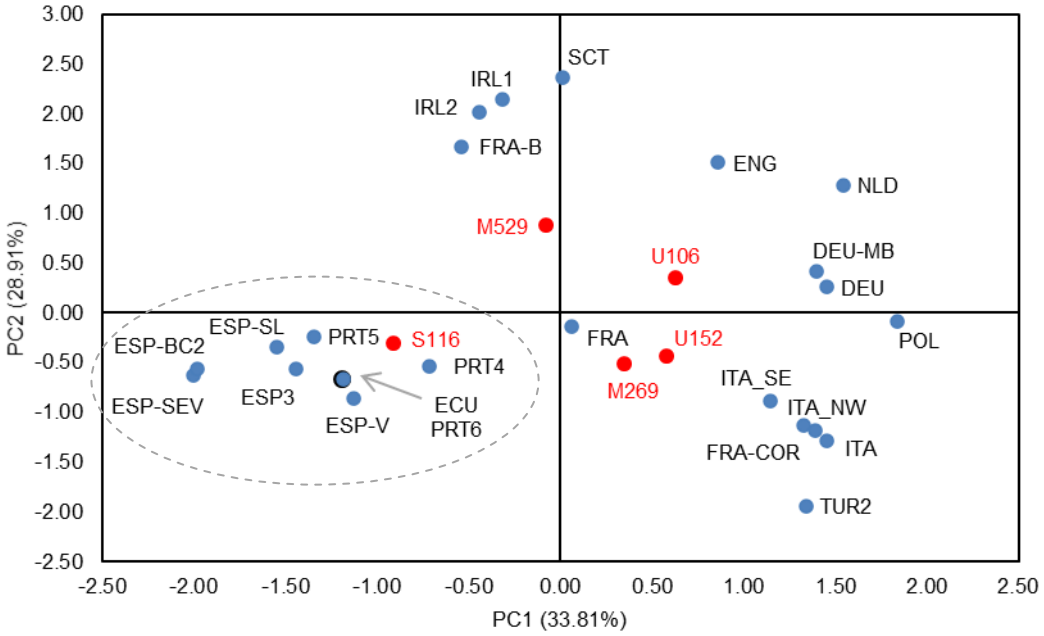


Figure 28: Principal component analysis of R-M269*(xU106, S116), R-U106, R-S116*(xU152, M529), R-U152 and R-M529 haplogroup frequencies in Ecuadorian population from the present study (highlighted with a black circle) and in different European populations (PRT- Portugal, ESP- Spain, IRL- Ireland, SCT- Scotland, ENG- England, FRA- France, ITA- Italy, NLD- Netherlands, DEU- Germany, POL- Poland, BGR- Bulgaria and TUR- Turkey; more details about the population codes can be seen in Appendix TableS7).

These analyses allowed to know that the most important contribution of R-lineages to Ecuador came from the Iberian Peninsula. Since the R-lineages were the main European lineages found in the Ecuadorian populations, this means that the most important contribution was from Iberian population, more likely from Spain than from Portugal (due to the history of the country). The other R-lineages [R-M269*(xU106, S116), R-U106, R-U152 and R-M529], have probably other European origins, possibly from Germany, Great Britain and Italy. However, more studies are needed in order to increase the knowledge about the genetic contributions of these countries into Ecuadorian male gene pool.

Conclusion

Ecuador is considered the smallest megadiverse country in the world, having also a huge human ethnic groups diversity. The three main ethnic groups that inhabit the Ecuador are the Native Americans, Mestizos and Afro-Ecuadorians, being the last two originated due to the Spanish colonization of Ecuador. During the colonial period, as in other South American countries, waves of migration from Europe (mainly Spain) and from Africa to Ecuador occurred, which allowed the admixture between different populations. In the present study and analysing Y chromosomal markers it was possible to increase the knowledge of the male history of Ecuador, mainly in Mestizo and Afro-descendant ethnic groups.

High haplotype and haplogroup diversities were obtained in the two present study population groups, being the highest diversity values found in the Mestizo population. Furthermore, statistically significant differences were found between them through pairwise genetic distances analyses based on both Y-STR and Y-SNP markers. These results were expected due to the different histories underlying the origin of these groups. Indeed, the haplogroup composition of the two groups is very different. Most of the lineages found in the Mestizo group have a European origin [haplogroups G (5.37%), I (6.71%), J (7.38%), KL (0.67%), R (except R-V88; 38.93), T (1.34%) and E-M81, E-M123 and E-M78 (10.07%)], followed by Native American origin [haplogroup Q (24.49%)] and African origin [haplogroups E (except the previously mentioned haplogroups; 6.04%)]. On the other hand, the male gene pool of the Afro-Ecuadorian population is mainly composed by African lineages [haplogroups B (1.33%), E (except the previously mentioned haplogroups; 74.00%) and R-V88 (5.88%)], followed by European lineages [haplogroups I (2.67%), J (6.00%), KL (0.67%), R (except R-V88) and E-M81, E-M123 and E-M78 (2.67%)] and only 1.33% of Native American lineage (haplogroup Q-M3).

Considering the genetic comparisons between the two Ecuadorian ethnic groups and other Latin American populations it was possible to detect that Mestizos have a closer relationship with populations with a higher European ancestry, whereas the Afro-Ecuadorians are more similar with populations with high level of African ancestry. However, since the present analyses were base on a low number of Y-STRs and Y-SNPs common to all populations, more studies focused on these ethnic groups and with a higher resolution of both Y-STRs and Y-SNPs markers are needed in order to

better understand the relationships between populations from different Latin American countries. When analysing exclusively the Native American lineages found in the Ecuadorian samples it was possible to conclude that the Ecuadorian haplotypes are very different from those in Native American populations from Ecuador. Nevertheless, the major haplogroup presented in almost all populations analysed (the Ecuadorian and also the other Native American and admixed South American populations) was the Q-M3*(xM19).This result suggests that the Y-SNPs described until now, downstream to the M3 marker, do not provide enough resolution to reveal genetic relationships between South American populations.

The most prevalent African haplogroup found in the Ecuadorian samples was the E-M2, which is also very frequent in sub-Saharan populations. Considering all African lineages present in the Ecuadorian samples, it was possible to conclude that populations from different sub-Saharan African countries contributed to the male gene pool from Ecuador. This result is in accordance to the historical data that mention different African regions as the source of African slaves brought to America. Considering the African populations that were compared with our Ecuadorian sample the most likely origin were Guinea-Bissau, Ivory Coast, Benin, Equatorial Guinea and Gabon. Mozambique also seem had contributed to the E-U209*(xU290, M154) lineages found in Ecuador.

The most probably European ancestry is also in accordance with the history of past migrations to Ecuador. The most frequent European haplogroup found Ecuador was the R-S116*(xU152, M529, M153, M167), which can also be found in Western European population, namely in Iberian Peninsula. Therefore, this European region seem to be the origin of the most European Y chromosomes present in the Ecuadorian population. Since Ecuador was a Spanish colony, the most probable origin is Spain rather than Portugal. Additional, European lineages that were also found in Ecuador seem to have other European origins, but more studies are required to understand the different European contributions to the Ecuadorian paternal gene pool.

In sum, this study allowed to increase the knowledge about the Ecuador colonization history, highlighted to the differences on degrees of admixture in both Mestizo and Afro-descendant populations where European and African men were very important to the foundation of each group, respectively. Moreover, the data obtained in this study will be helpful to future population and forensic studies.

References

1. Achilli A, Olivieri A, Semino O, Torroni A. 2018. Ancient human genomes—keys to understanding our past. *Science* 360:964-5
2. Achilli A, Perego UA, Bravi CM, Coble MD, Kong Q-P, et al. 2008. The phylogeny of the four pan-American MtDNA haplogroups: implications for evolutionary and disease studies. *PLoS One* 3:e1764
3. Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, et al. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *The American Journal of Human Genetics* 75:910-8
4. Adams SM, Bosch E, Balaresque PL, Ballereau SJ, Lee AC, et al. 2008. The genetic legacy of religious diversity and intolerance: paternal lineages of Christians, Jews, and Muslims in the Iberian Peninsula. *The American Journal of Human Genetics* 83:725-36
5. Alonso Morales LA, Casas-Vargas A, Rojas Castro M, Resque R, Ribeiro-dos-Santos ÂK, et al. 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:e0207130
6. Alvarez L, Ciria E, Marques SL, Santos C, Aluja MP. 2014. Y-chromosome analysis in a Northwest Iberian population: Unraveling the impact of Northern African lineages. *American Journal of Human Biology* 26:740-6
7. Andrade FG, Sánchez D. 2004. Genetic profile of the Kichwas (Quichuas) from Ecuador by analysis of STR loci. *Human Biology* 76:723-30
8. Ayala Mora E. 2008. *Resumen de historia del Ecuador*. Corporación Editora Nacional
9. Ayala Mora E. 2017. *Ecuador Patria de Todos – Identidad nacional, interculturalidad e integración*. Corporación Editora Nacional
10. Baeta M, Núñez C, Sosa C, Bolea M, Casalod Y, et al. 2012. Mitochondrial diversity in Amerindian Kichwa and Mestizo populations from Ecuador. *International Journal of Legal Medicine* 126:299-302
11. Baeta M, Núñez C, Villaescusa P, Ortueta U, Ibarbia N, et al. 2018. Assessment of a subset of Slowly Mutating Y-STRs for forensic and evolutionary studies. *Forensic Science International: Genetics* 34:e7-e12
12. Baeza C, Guzmán R, Tirado M, López-Parra AM, Rodríguez T, et al. 2007. Population data for 15 Y-chromosome STRs in a population sample from Quito (Ecuador). *Forensic Science International* 173:214-9
13. Bafalluy MB. 2012. *Estudio de la variabilidad del genoma mitocondrial y de marcadores sexuales en grupos étnicos de Ecuador*. PhD dissertation in Medicine, Universidad de Zaragoza
14. Ballantyne KN, Keerl V, Wollstein A, Choi Y, Zuniga SB, et al. 2012. A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Science International: Genetics* 6:208-1

15. Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16:37-48
16. Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. *Genetics* 141:743-53
17. Bandelt HJ, Herrnstadt C, Yao YG, Kong QP, Kivisild T, et al. 2003. Identification of Native American founder mtDNAs through the analysis of complete mtDNA sequences: some caveats. *Annals of Human Genetics* 67:512-24
18. Barbieri C, Heggarty P, Yang Yao D, Ferri G, De Fanti S, et al. 2014. Between Andes and A mazon: The genetic profile of the A rawak-speaking Yanesha. *American Journal of Physical Anthropology* 155:600-9
19. Barbieri C, Hübner A, Macholdt E, Ni S, Lippold S, et al. 2016. Refining the Y chromosome phylogeny with southern African sequences. *Human Genetics* 135:541-53
20. Batini C, Ferri G, Destro-Bisol G, Brisighelli F, Luiselli D, et al. 2011. Signatures of the preagricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages. *Molecular Biology and Evolution* 28:2603-13
21. Battaglia V, Grugni V, Perego UA, Angerhofer N, Gomez-Palmieri JE, et al. 2013. The first peopling of South America: new evidence from Y-chromosome haplogroup Q. *PLoS One* 8:e71390
22. Beleza S, Gusmão L, Lopes A, Alves C, Gomes I, et al. 2006. Micro-phylogeographic and demographic history of Portuguese male lineages. *Annals of Human Genetics* 70:181-94
23. Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, et al. 2009. Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Molecular Biology and Evolution* 26:1581-9
24. Blanco-Verea A, Jaime J, Brión M, Carracedo A. 2010. Y-chromosome lineages in native South American population. *Forensic Science International: Genetics* 4:187-93
25. Boattini A, Martinez-Cruz B, Sarno S, Harmant C, Useli A, et al. 2013. Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. *PloS One* 8:e65441
26. Bodner M, Perego UA, Huber G, Fendt L, Röck AW, et al. 2012. Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome Research* 22:811-20
27. Bonatto SL, Salzano FM. 1997. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proceedings of the National Academy of Sciences* 94:1866-71
28. Bradley B, Stanford D. 2004. The North Atlantic ice-edge corridor: a possible Palaeolithic route to the New World. *World Archaeology* 36:459-78
29. Brandini S, Bergamaschi P, Cerna MF, Gandini F, Bastaroli F, et al. 2018. The Paleo-Indian entry into South America according to mitogenomes. *Molecular Biology and Evolution* 35:299-311
30. 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. *The American Journal of Human Genetics* 62:1408-15

31. Brion M, Sobrino B, Blanco-Verea A, Lareu M, Carracedo A. 2005. Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *International Journal of Legal Medicine* 119:10-5
32. Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC. 1991. Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *American journal of human genetics* 48:137
33. Busby GB, Brisighelli F, Sanchez-Diz P, Ramos-Luis E, Martinez-Cadenas C, et al. 2012. The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proceedings of the Royal Society B: Biological Sciences* 279:884-92
34. Butler JM. 2005. *Forensic DNA typing: biology, technology, and genetics of STR markers*. Elsevier
35. Caglià A, Tofanelli S, Coia V, Boschi I, Pescarmona M, et al. 2003. A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between F_{ST} and R_{ST} genetic distances. *Human Biology 75*:313-30
36. Campoi IC. 2019. América miscigenada: dominação e colonização espanhola através da análise do quadro Los mulatos de Esmeraldas (1599). In: ANPUH-Brasil – 30º Simpósio Nacional de História
37. Cárdenas JM, Heinz T, Pardo-Seco J, Álvarez-Iglesias V, Taboada-Echalar P, et al. 2015. The multiethnic ancestry of Bolivians as revealed by the analysis of Y-chromosome markers. *Forensic Science International: Genetics* 14:210-8
38. Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-Barrientos D, Duque C, et al. 2000. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *The American Journal of Human Genetics* 67:1287-95
39. Carvalho M, Brito P, Bento A, Gomes V, Antunes H, et al. 2011. Paternal and maternal lineages in Guinea-Bissau population. *Forensic Science International: Genetics* 5:114-6
40. Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. *Nature Genetics* 33:266-75
41. Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 355:1563-72
42. Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Human Genetics* 114:127-48
43. Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, et al. 2009. The last glacial maximum. *Science* 325:710-4
44. Coelho M, Sequeira F, Luiselli D, Beleza S, Rocha J. 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evolutionary Biology* 9:80
45. Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, et al. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* 508:488-93
46. Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, et al. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *The American Journal of Human Genetics* 74:1014-22

47. Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R. 2006. Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. *Human Mutation* 27:831-2
48. Cruciani F, Trombetta B, Sellitto D, Massaia A, Destro-Bisol G, et al. 2010. Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *European Journal of Human Genetics* 18:800-7
49. D'Atanasio E, Iacovacci G, Pistillo R, Bonito M, Dugoujon J-M, et al. 2019. Rapidly mutating Y-STRs in rapidly expanding populations: Discrimination power of the Yfiler Plus multiplex in northern Africa. *Forensic Science International: Genetics* 38:185-94
50. D'Atanasio E, Trionfetti F, Bonito M, Sellitto D, Coppa A, et al. 2020. Y Haplogroup Diversity of the Dominican Republic: Reconstructing the Effect of the European Colonization and the Trans-Atlantic Slave Trades. *Genome Biology and Evolution* 12:1579-90
51. De Acosta J. 2002. *Natural and moral history of the Indies*. Duke University Press
52. de Filippo C, Barbieri C, Whitten M, Mpoloka SW, Gunnarsdóttir ED, et al. 2011. Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Molecular Biology and Evolution* 28:1255-69
53. de Knijff P. 2000. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *The American Journal of Human Genetics* 67:1055-61
54. Di Cristofaro J, Mazières S, Tous A, Di Gaetano C, Lin AA, et al. 2018. Prehistoric migrations through the Mediterranean basin shaped Corsican Y-chromosome diversity. *PloS One* 13:e0200641
55. Diegoli TM. 2015. Forensic typing of short tandem repeat markers on the X and Y chromosomes. *Forensic Science International: Genetics* 18:140-51
56. Dillehay TD, Ramírez C, Pino M, Collins MB, Rossen J, Pino-Navarro J. 2008. Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* 320:784-6
57. Dixon EJ. 2001. Human colonization of the Americas: timing, technology and process. *Quaternary Science Reviews* 20:277-99
58. Dornelles CL, Bonatto SL, De Freitas LB, Salzano FM. 2005. Is haplogroup X present in extant South American Indians? *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists* 127:439-48
59. Dulik MC, Owings AC, Gaieski JB, Vilar MG, Andre A, et al. 2012. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan-and Eskimoan-speaking populations. *Proceedings of the National Academy of Sciences* 109:8471-6
60. ÉcVa, Cca. 2009. *Enciclopedia del saber afroecuatoriano*. Vicariato Apostólico de Esmeraldas, Quito, Ecuador
61. Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. *Nature reviews genetics* 5:435-45

62. Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*. 10: 564-7
63. Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, et al. 2008. Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *The American Journal of Human Genetics* 82:583-92
64. Finer M, Jenkins CN, Pimm SL, Keane B, Ross C. 2008. Oil and gas projects in the western Amazon: threats to wilderness, biodiversity, and indigenous peoples. *PloS One* 3:e2932
65. Flores C, Maca-Meyer N, González AM, Oefner PJ, Shen P, et al. 2004. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *European Journal of Human Genetics* 12:855-63
66. Fortes-Lima C, Brucato N, Croze M, Bellis G, Schiavinato S, et al. 2015. Genetic population study of Y-chromosome markers in Benin and Ivory Coast ethnic groups. *Forensic Science International: Genetics* 19:232-7
67. Gambéla B.2009. Billy Gambéla – Afri-Asiatic Anthropology Blog. https://billygambelaafroasiaticanthropology.wordpress.com/ Downloaded on 18 June 2020
68. Gamble C, Davies W, Pettitt P, Richards M. 2004. Climate change and evolving human diversity in Europe during the last glacial. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 359:243-5
69. García A, Pauro M, Bailliet G, Bravi CM, Demarchi DA. 2018. Genetic variation in populations from central Argentina based on mitochondrial and Y chromosome DNA evidence. *Journal of Human Genetics* 63:493-507
70. García Salazar J, Walsh C. 2020. *Cimarronaje en el Pacífico Sur*. Ediciones Abya-Yala
71. Gaviria A, Sánchez M, Morejón G, Vela M, Aguirre V, et al. 2013. Characterization and Haplotype analysis of 11 Y-STR loci in Ecuadorian population. *Forensic Science International: Genetics Supplement Series* 4:e310-e1
72. Gayà-Vidal M, Moral P, Saenz-Ruales N, Gerbault P, Tonasso L, et al. 2011. mtDNA and Y-chromosome diversity in Aymaras and Quechuas from Bolivia: Different stories and special genetic traits of the Andean Altiplano populations. *American Journal of Physical Anthropology* 145:215-30
73. Geppert M, Ayub Q, Xue Y, Santos S, Ribeiro-dos-Santos Â, et al. 2015. Identification of new SNPs in native South American populations by resequencing the Y chromosome. *Forensic Science International: Genetics* 15:111-4
74. Geppert M, Baeta M, Núñez C, Martínez-Jarreta B, Zweynert S, 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* 5:100-4
75. Goebel T, Waters MR, O'Rourke DH. 2008. The late Pleistocene dispersal of modern humans in the Americas. *Science* 319:1497-502

76. Gomes V, Sánchez-Diz P, Amorim A, Carracedo Á, Gusmão L. 2010. Digging deeper into East African human Y chromosome lineages. *Human Genetics* 127:603-13
77. Gómez-Carballa A, Pardo-Seco J, Brandini S, Achilli A, Perego UA, et al. 2018. The peopling of South America and the trans-Andean gene flow of the first settlers. *Genome Research* 28:767-79
78. Gonçalves R, Freitas A, Branco M, Rosa A, Fernandes AT, et al. 2005. Y-chromosome lineages from Portugal, Madeira and Açores record elements of Sephardim and Berber ancestry. *Annals of Human Genetics* 69:443-54
79. González-Andrade F. 2006. *Análisis molecular de variación de polimorfismos STR autosómicos y de cromosoma «Y» en grupos étnicos de Ecuador*. PhD dissertation in Medicine, Universidad de Zaragoza
80. González-Andrade F. 2011. The genetic ethnography of the Native Amerindians of Ecuador. *International Journal of Medical and Biological Frontiers* 17
81. Gonzalez-Andrade F, Roewer L, Willuweit S, Sanchez D, Martínez-Jarreta B. 2009. Y-STR variation among ethnic groups from Ecuador: Mestizos, Kichwas, Afro-Ecuadorians and Waoranis. *Forensic Science International: Genetics* 3:e83-e91
82. González-Andrade F, Sánchez D, González-Solórzano J, Gascón S, Martínez-Jarreta B. 2007. Sex-specific genetic admixture of mestizos, amerindian kichwas, and afro-ecuadorans from ecuador. *Human Biology* 79:51-77
83. González M, Gomes V, López-Parra AM, Amorim A, Carracedo A, et al. 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* 21:324-31
84. González dos Santos M. 2010. *The y chromosome haplogroup R1b1 in Central-West Africa*. Master dissertation in Forensic Genetics, Universidade do Porto
85. Gouveia MH, Borda V, Leal TP, Moreira RG, Bergen AW, et al. 2020. Origins, Admixture Dynamics, and Homogenization of the African Gene Pool in the Americas. *Molecular Biology and Evolution* 37:1647-56
86. Greenberg JH, Turner CG, Zegura SL, Campbell L, Fox JA, et al. 1986. The settlement of the Americas: A comparison of the linguistic, dental, and genetic evidence [and comments and reply]. *Current Anthropology* 27:477-97
87. Grugni V, Raveane A, Ongaro L, Battaglia V, Trombetta B, et al. 2019. Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. *BMC Biology* 17:1-14
88. Hammer MF, Zegura SL. 2002. The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. *Annual Review of Anthropology* 31:303-21
89. Hanratty DM. 1991. *Ecuador: a country study*. US Government Printing Office
90. Hernández CL, Dugoujon J-M, Sánchez-Martínez LJ, Cuesta P, Novelletto A, Calderón R. 2019. Paternal lineages in southern Iberia provide time frames for gene flow from mainland Europe and the Mediterranean world. *Annals of Human Biology* 46:63-76
91. Heyer E, Chaix R, Pavard S, Austerlitz F. 2012. Sex-specific demographic behaviours that shape human genomic variation. *Molecular Ecology* 21:597-612
92. Hoffecker JF, Elias SA, O'Rourke DH. 2014. Out of Beringia? *Science* 343:979-80

93. Hoffecker JF, Elias SA, O'Rourke DH, Scott GR, Bigelow NH. 2016. Beringia and the global dispersal of modern humans. *Evolutionary Anthropology: Issues, News, and Reviews* 25:64-78
94. Hughes JF, Rozen S. 2012. Genomics and genetics of human and primate Y chromosomes. *Annual Review of Genomics and Human Genetics* 13:83-108
95. Jobling MA, Gill P. 2004. Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics* 5:739-51
96. Jobling MA, Pandya A, Tyler-Smith C. 1997. The Y chromosome in forensic analysis and paternity testing. *International Journal of Legal Medicine* 110:118-24
97. Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics* 4:598-612
98. Jobling MA, Tyler-Smith C. 2017. Human Y-chromosome variation in the genome-sequencing era. *Nature Reviews Genetics* 18:485
99. Jota MS, Lacerda DR, Sandoval JR, Vieira PPR, Ohasi D, et al. 2016. New native South American Y chromosome lineages. *Journal of Human Genetics* 61:593-603
100. Karachanak S, Grugni V, Fornarino S, Nesheva D, Al-Zahery N, et al. 2013. Y-chromosome diversity in modern Bulgarians: new clues about their ancestry. *PLoS One* 8:e56779
101. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 18:830-8
102. Kashani BH, Perego UA, Olivieri A, Angerhofer N, Gandini F, et al. 2012. Mitochondrial haplogroup C4c: A rare lineage entering America through the ice-free corridor? *American Journal of Physical Anthropology* 147:35-9
103. Kayser M. 2017. Forensic use of Y-chromosome DNA: a general overview. *Human Genetics* 136:621-35
104. Kayser M, Kittler R, Erler A, Hedman M, Lee AC, et al. 2004. A comprehensive survey of human Y-chromosomal microsatellites. *The American Journal of Human Genetics* 74:1183-97
105. Labarga CM. Origen y evolución de las comunidades afroamericanas de Esmeraldas (Ecuador): datos históricos y genéticos. *Proc. Anales del museo de América*, *1997*:129-38: Subdirección General de Documentación y Publicaciones
106. Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. *Science* 286:964-7
107. Larmuseau MH, Vessi A, Jobling MA, Van Geystelen A, Primativo G, et al. 2015. The paternal landscape along the Bight of Benin–Testing regional representativeness of West-African population samples using Y-chromosomal markers. *PloS One* 10:e0141510
108. Leone P, Maldonado-Oyervide D, Astudillo-González O, Pérez-Villa A, Yumiceba V, et al. 2019. Mitochondrial DNA study in the Shuar ethnic group from Ecuador. *Forensic Science International: Genetics Supplement Series*
109. Lessig R, Zoledziewska M, Fahr K, Edelmann J, Kostrzewa M, et al. 2005. Y-SNP-genotyping–a new approach in forensic analysis. *Forensic Science International* 154:128-36

110. Llamas B, Fehren-Schmitz L, Valverde G, Soubrier J, Mallick S, et al. 2016. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. *Science Advances* 2:e1501385
111. López-Ramírez YL, Aguilar-Velázquez JA, López-Armenta M, Ruiz-Hernández 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* 134:199-202
112. Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, et al. 2004. The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *The American Journal of Human Genetics* 74:532-44
113. Martínez-Cruz B, Ziegle J, Sanz P, Sotelo G, Anglada R, et al. 2011. Multiplex single-nucleotide polymorphism typing of the human Y chromosome using TaqMan probes. *Investigative Genetics* 2:13
114. Mendez FL, Karafet TM, Krahn T, Ostrer H, Soodyall H, Hammer MF. 2011. Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. *Human biology* 83:39-53
115. Mezzavilla M, Geppert M, Tyler-Smith C, Roewer L, Xue Y. 2015. Insights into the origin of rare haplogroup C3* Y chromosomes in South America from high-density autosomal SNP genotyping. *Forensic Science International: Genetics* 15:115-20
116. Micheletti SJ, Bryc K, Esselmann SGA, Freyman WA, Moreno ME, et al. 2020. Genetic consequences of the transatlantic slave trade in the Americas. *The American Journal of Human Genetics* 107:265-77
117. Millstein RL, Skipper RA. 2006. Population genetics. [Preprint]
118. Moncada Madero JA. Análisis de la estructura genética de una muestra de población del departamento de Amazonas, Colombia. Master dissertation in Human Genetics, Universidad Nacional de Colombia
119. Moreno-Mayar JV, Potter BA, Vinner L, Steinrücken M, Rasmussen S, et al. 2018. Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. *Nature* 553:203-7
120. Moreno-Mayar JV, Vinner L, de Barros Damgaard P, De La Fuente C, Chan J, et al. 2018. Early human dispersals within the Americas. *Science* 362:eaav2621
121. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Proc. Cold Spring Harbor symposia on quantitative biology*, *1986*, 51:263-73: Cold Spring Harbor Laboratory Press
122. Myres NM, Rootsi S, Lin AA, Järve M, King RJ, et al. 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *European Journal of Human Genetics* 19:95-101
123. Noguera MC, Schwegler A, Gomes V, Briceño I, Alvarez L, et al. 2014. Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar. *Annals of Human Biology* 41:453-9
124. Nogueiro I, Manco L, Gomes V, Amorim A, Gusmão L. 2010. Phylogeographic analysis of paternal lineages in NE Portuguese Jewish communities. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists* 141:373-81

125. 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:e192-e5
126. Oliveira AM, Domingues PM, Gomes V, Amorim A, Jannuzzi J, et al. 2014. Male lineage strata of Brazilian population disclosed by the simultaneous analysis of STRs and SNPs. *Forensic Science International: Genetics* 13:264-8
127. Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, et al. 2012. Mitochondrial DNA signals of late glacial recolonization of Europe from near eastern refugia. *The American Journal of Human Genetics* 90:915-24
128. Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, et al. 2009. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Current Biology* 19:1-8
129. Perego UA, Angerhofer N, Pala M, Olivieri A, Lancioni H, et al. 2010. The initial peopling of the Americas: a growing number of founding mitochondrial genomes from Beringia. *Genome Research* 20:1174-9
130. Pereira V, Gusmão L. 2016. Types of Genomes, Sequences and Genetic Markers (Repeats, SNPs, Indels, Haplotypes). *Handbook of Forensic Genetics: Biodiversity and Heredity in Civil and Criminal Investigation* 2:163
131. Pinotti T, Bergström A, Geppert M, Bawn M, Ohasi D, et al. 2019. Y chromosome sequences reveal a short Beringian standstill, rapid expansion, and early population structure of Native American Founders. *Current Biology* 29:149-57. e3
132. Posth C, Nakatsuka N, Lazaridis I, Skoglund P, Mallick S, et al. 2018. Reconstructing the deep population history of Central and South America. *Cell* 175:1185-97. e22
133. Poulsen L, Børsting C, Tomas C, González-Andrade F, Lopez-Pulles R, et al. 2011. Typing of Amerindian Kichwas and Mestizos from Ecuador with the SNPforID multiplex. *Forensic Science International: Genetics* 5:e105-e7
134. Purps J, Siegert S, Willuweit S, Nagy M, Alves C, et al. 2014. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Science International: Genetics* 12:12-23
135. Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, et al. 2002. Y-chromosomal DNA variation in Pakistan. *The American Journal of Human Genetics* 70:1107-24
136. Raff JA, Bolnick DA. 2015. Does mitochondrial haplogroup X indicate ancient trans-Atlantic migration to the Americas? A critical re-evaluation. *PaleoAmerica* 1:297-304
137. Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, et al. 2015. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* 349:aab3884
138. Ralf A, van Oven M, González DM, de Knijff P, van der Beek K, et al. 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
139. Rasmussen M, Anzick SL, Waters MR, Skoglund P, DeGiorgio M, et al. 2014. The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* 506:225-9

140. Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757-62
141. Rasmussen M, Sikora M, Albrechtsen A, Korneliussen TS, Moreno-Mayar JV, et al. 2015. The ancestry and affiliations of Kennewick Man. *Nature* 523:455-8
142. Rębała K, Martínez-Cruz B, Tönjes A, Kovacs P, Stumvoll M, et al. 2013. Contemporary paternal genetic landscape of Polish and German populations: from early medieval Slavic expansion to post-World War II resettlements. *European Journal of Human Genetics* 21:415-22
143. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, et al. 2012. Reconstructing native American population history. *Nature* 488:370-4
144. Resque R, Gusmao L, Geppert M, Roewer L, Palha T, et al. 2016. Male lineages in Brazil: intercontinental admixture and stratification of the European background. *PLoS One* 11:e0152573
145. Rey-González D, Gelabert-Besada M, Cruz R, Brisighelli F, Lopez-Soto M, et al. 2017. Micro and macro geographical analysis of Y-chromosome lineages in South Iberia. *Forensic Science International: Genetics* 29:e9-e15
146. Rickards O. 1995. Analysis of the region V mitochondrial marker in two Black communities of Ecuador, and in their parental populations. *Human Evolution* 10:5-16
147. Ritter J. 2003. Articulating Blackness in Afro-Ecuadorian Marimba Performance. *Musical Cultures of Latin America: Global Effects, Past and Present*:143-53
148. Roewer L. 2019. Y-chromosome short tandem repeats in forensics—Sexing, profiling, and matching male DNA. *Wiley Interdisciplinary Reviews: Forensic Science* 1:e1336
149. Roewer L, Amemann J, Spurr N, Grzeschik K-H, Epplen J. 1992. Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Human Genetics* 89:389-94
150. Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, et al. 2005. Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Human Genetics* 116:279-91
151. Roewer L, Nothnagel M, Gusmão L, Gomes V, González M, et al. 2013. Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans. *PLoS Genetics* 9:e1003460
152. Rosa A, Ornelas C, Jobling MA, Brehm A, Villems R. 2007. Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. *BMC Evolutionary Biology* 7:1-11
153. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, et al. 2002. Genetic structure of human populations. *Science* 298:2381-5
154. Rothhammer F, Dillehay TD. 2009. The late Pleistocene colonization of South America: an interdisciplinary perspective. *Annals of Human Genetics* 73:540-9
155. Rowold D, Garcia-Bertrand R, Calderon S, Rivera L, Benedico DP, et al. 2014. At the southeast fringe of the bantu expansion: genetic diversity and phylogenetic relationships to other sub-Saharan tribes. *Meta Gene* 2:670-85
156. Rueda Nova R. 2010. *De esclavizados a comuneros en la cuenca aurífera del Río Santiago-Río Cayapas (Esmeraldas) - Etnicidad negra en construcción en Ecuador siglos XVIII-XIX*. PhD dissertation in History, Universidad Pablo de Olavide, Sevilla

157. Sandoval JR, Lacerda DR, Acosta O, Jota MS, Robles-Ruiz P, et al. 2016. The Genetic History of Peruvian Quechua-Lamistas and Chankas: Uniparental DNA Patterns among Autochthonous Amazonian and Andean Populations. *Annals of Human Genetics* 80:88-101
158. Santangelo R, González-Andrade F, Børsting C, Torroni A, Pereira V, Morling N. 2017. Analysis of ancestry informative markers in three main ethnic groups from Ecuador supports a trihybrid origin of Ecuadorians. *Forensic Science International: Genetics* 31:29-33
159. Sawe BE. 2018. Ethnic Groups Of Ecuador. World Atlas. https://www.worldatlas.com/articles/ethnic-groups-of-ecuador.html Downloaded on 28 May 2019
160. Scheib CL, Li H, Desai T, Link V, Kendall C, et al. 2018. Ancient human parallel lineages within North America contributed to a coastal expansion. *Science* 360:1024-7
161. Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, et al. 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. *The American Journal of Human Genetics* 74:1023-34
162. Sikora M, Pitulko VV, Sousa VC, Allentoft ME, Vinner L, et al. 2019. The population history of northeastern Siberia since the Pleistocene. *Nature* 570:182-8
163. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825-37
164. Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, et al. 2015. Genetic evidence for two founding populations of the Americas. *Nature* 525:104-8
165. Skoglund P, Reich D. 2016. A genomic view of the peopling of the Americas. *Current Opinion in Genetics & Development* 41:27-35
166. Slatkin M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457-62
167. Sobrino B, Brión M, Carracedo A. 2005. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International* 154:181-94
168. Solé-Morata N, Bertranpetit J, Comas D, Calafell F. 2015. Y-chromosome diversity in Catalan surname samples: insights into surname origin and frequency. *European Journal of Human Genetics* 23:1549-57
169. Solé-Morata N, Villaescusa P, García-Fernández C, Font-Porterias N, Illescas MJ, et al. 2017. Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ. *Scientific Reports* 7:1-13
170. Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, et al. 2007. Beringian standstill and spread of Native American founders. *PloS One* 2:e829
171. Tarazona-Santos E, Carvalho-Silva DR, Pettener D, Luiselli D, De Stefano GF, et al. 2001. Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *The American Journal of Human Genetics* 68:1485-96

172. Tinajero A, Barba G. 2013. Cronología de la historia resumida del Ecuador. http://www.ecuaworld.com/historia_de_ecuador.pdf Downloaded on 15 November 2019
173. Tishkoff SA, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annual Review of Genomics and Human Genetics* 4:293-340
174. Tofanelli S, Bertoncini S, Castrì L, Luiselli D, Calafell F, et al. 2009. On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. *Molecular Biology and Evolution* 26:2109-24
175. Toscanini U, Brisighelli F, Moreno F, Pantoja-Astudillo JA, Morales EA, et al. 2016. Analysis of Y-chromosome STRs in Chile confirms an extensive introgression of European male lineages in urban populations. *Forensic Science International: Genetics* 21:76-80
176. Toscanini U, Gaviria A, Pardo-Seco J, Gómez-Carballa A, Moscoso F, et al. 2018. The geographic mosaic of Ecuadorian Y-chromosome ancestry. *Forensic Science International: Genetics* 33:59-65
177. Trombetta B, Cruciani F, Sellitto D, Scozzari R. 2011. A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms. *PLoS One* 6:e16073
178. Unknown. 2013. Grupos Étnicos Ecuador. http://gruposetnicosec.blogspot.com/ Downloaded on 16 October 2020
179. Valverde L, Illescas MJ, Villaescusa P, Gotor AM, García A, et al. 2016. New clues to the evolutionary history of the main European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia. *European Journal of Human Genetics* 24:437-41
180. Van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MH. 2014. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Human Mutation* 35:187-91
181. Velez C, Palamara P, Guevara-Aguirre J, Hao L, Karafet T, et al. 2012. The impact of Converso Jews on the genomes of modern Latin Americans. *Human Genetics* 131:251-63
182. Villaescusa P, Illescas MJ, Valverde L, Baeta M, Nuñez C, et al. 2017. Characterization of the Iberian Y chromosome haplogroup R-DF27 in Northern Spain. *Forensic Science International: Genetics* 27:142-8
183. Waters MR. 2019. Late Pleistocene exploration and settlement of the Americas by modern humans. *Science* 365:eaat5447
184. Waters MR, Stafford TW. 2007. Redefining the age of Clovis: implications for the peopling of the Americas. *Science* 315:1122-6
185. Waters MR, Stafford TW, McDonald HG, Gustafson C, Rasmussen M, et al. 2011. Pre-Clovis mastodon hunting 13,800 years ago at the Manis site, Washington. *Science* 334:351-3
186. Wei L-H, Wang L-X, Wen S-Q, Yan S, Canada R, et al. 2018. Paternal origin of Paleo-Indians in Siberia: insights from Y-chromosome sequences. *European Journal of Human Genetics* 26:1687-96
187. Willuweit S, Roewer L. 2007. Y chromosome haplotype reference database (YHRD): update. *Forensic Science International: Genetics* 1:83-7

188. Willems T, Gymrek M, Poznik GD, Tyler-Smith C, Erlich Y, Group GPCY. 2016. Population-scale sequencing data enable precise estimates of Y-STR mutation rates. *The American Journal of Human Genetics* 98:919-33
189. Xavier C, Builes JJ, Gomes V, Ospino JM, Aquino J, et al. 2015. Admixture and genetic diversity distribution patterns of non-recombining lineages of Native American ancestry in Colombian populations. *PloS One* 10:e0120155
190. Y chromosome consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research* 12:339-48
191. Yanez PVG. 2013. Reconociendo el importante rol que desempeña el pueblo Montubio en el contexto sociocultural del Ecuador. *La Técnica*:6-15
192. Zambrano AK, Gaviria A, Cobos-Navarrete S, Gruezo C, Rodríguez-Pollit C, et al. 2019. The three-hybrid genetic composition of an Ecuadorian population using AIMs-InDels compared with autosomes, mitochondrial DNA and Y chromosome data. *Scientific Reports* 9:9247
193. Zambrano A, Gaviria A, Vela M, Cobos S, Leone P, et al. 2017. Ancestry characterization of Ecuador's Highland mestizo population using autosomal AIM-INDELs. *Forensic Science International: Genetics Supplement Series* 6:e477-e8
194. Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 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* 21:164-75

Appendix

List of Tables

Table S1: Y-STR profiles and corresponding haplogroups (predicted and the results obtained in the present study) of the Ecuadorian samples. The region of birth of each individual is also presented in this table.....XX

Table S2: Primers used for PCR of the multiplex E1 and Q.....XXXII

Table S3: Primers used for SNaPshot reaction of multiplexes E1 and Q. The primers of multiplex (Mx) E2 whose sequence were changed are also presented. Before the sequence of the primer is indicated if the primer is forward (For) or reveres (Rev).....XXXII

Table S4: The concentrations in the SNaPshot reaction of the primers from multiplexes 1, 2, B, E2 (except the Y-SNPs mentioned in table S3), R1 and R2 that were optimised from the conditions previously described by Brion et al. (31), Gomes et al. (76) and Resque et al. (144).....XXXIV

Table S5: Populations codes of the Latin American populations used in the different analyses performed.....XL

Table S6: Populations codes of the African populations used in the different analyses performed.....XLI

Table S7: Populations codes of the European and Middle East populations used in the different analyses performed.....XLII

Table S8: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other three studies about Ecuadorian population. The non-significant values are highlighted in bold ($P \geq 0.05$). Population codes can be seen in Table S5.....XLIII

Table S9: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations. The non-significant values are highlighted in bold ($P \geq 0.05$). Population codes can be seen in Tables S5-S7.....XLIV

Table S10: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations. The non-significant values are highlighted in bold ($P \geq 0.05$). Population codes can be seen in Tables S5-S7.....XLVI

Table S11: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations. These comparisons were based on Y-STR data and only the Q-lineages were considered. Non-significant values are highlighted in bold ($P \geq 0.05$). Population codes can be seen in Table S5.....XLVII

Table S12: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations. These comparisons were based on Y-STR data and only the Q-lineages were considered. Non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S5.....XLVIII

Table S13: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations within the haplogroup Q. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S5.....XLIX

Table S14: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and African populations. These comparisons were based on Y-STRs data and only the African lineages were considered. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S6.....L

Table S15: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and African populations. These comparisons were based on Y-STRs data and only the African lineages were considered. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S6.....LI

Table S16: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and African populations. These comparisons were based on Y-SNPs data and only the African lineages were considered. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S6.....LII

Table S17: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and European populations. These comparisons were performed considering only the European lineages. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S7.....LIII

Table S18: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and European populations. These comparisons were performed considering only the European lineages. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S7.....LIV

Table S19: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and European populations. These comparisons were performed considering only the European lineages belonging to haplogroup R. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S7.....LVI

List of Figures

Figure S1: Phylogenetic tree of the Y chromosome haplogroups studied with the biallelic markers displayed in each branch. The colours correspond to the different multiplexes used (grey - Mx1; red - MxB; blue - MxE1 and E2; orange - Mx2; purple - MxQ; green - MxR1 and R2). The Y-SNPs that are analysed in more than one multiplex are highlighted in bold. The Y-SNPs only belonging to MxE1 are in a darker blue than the ones from MxE2. The Y-SNPs only belonging to MxR1 are in a darker green than the ones from MxR2. The marker genotyped in a singleplex reaction is in black. XXXI

Figure S2: Discrimination of multiplex 2 fragments in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one. The ladder (L) is also shown in the figure. Note that the 12f2a is an indel and so the result can be determined in the gel, being the insertion considered to be the ancestral state..... XXXV

Figure S3: Discrimination of multiplex 1 fragments in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one. The ladder (L) is also shown in the figure..... XXXV

Figure S4: Discrimination of multiplex B fragments in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one. XXXVI

Figure S5: Discrimination of the fragments of multiplexes E1 (a) and E2 (b) in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one.....XXXVI

Figure S7: Discrimination of the fragments of multiplexes R1 (a) and R2 (b) in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one.....XXXVII

Figure S6: Discrimination of multiplex Q fragments in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one. XXXVII

Figure S10: Electropherogram of the SBE reaction of a sample belonging to haplogroup T-M70. The Y-SNP and the two possible alleles detected with multiplex 1 are indicated with arrows. The alleles present in this sample are highlighted in bold. XXXVIII

Figure S9: Electropherogram of the SBE reaction of a sample belonging to haplogroup G-M201. The Y-SNP and the two possible alleles detected with multiplex 2 are indicated with arrows. The alleles present in this sample are highlighted in bold.XXXVIII

Figure S8: Electropherogram of the SBE reaction of a sample belonging to haplogroup B-M112*(xM30). The Y-SNP and the two possible alleles detected with multiplex B are indicated with arrows. The alleles present in this sample are highlighted in bold.XXXVIII

Figure S11: Electropherogram of the SBE reaction of two samples belonging to haplogroups E-M2*(xU209, M191) (a) and E-M81 (b). The Y-SNP and the two possible alleles detected with multiplexes E1 (a) and E2 (b) are indicated with arrows. The alleles present in these samples are highlighted in bold.....XXXVIII

Figure S12: Electropherogram of the SBE reaction of a sample belonging to haplogroup Q-M3*(xM19, Z19319, Z19483, M557, SA05). The Y-SNP and the two possible alleles detected with multiplex Q are indicated with arrows. The alleles present in this sample are highlighted in bold. XXXIX

Figure S13: Electropherogram of the SBE reaction of two samples belonging to haplogroups R-S116*(xU152, M529, M153, M167) (a) and R-V88 (result known after a sequencing reaction for the V88 Y-SNP) (b). The Y-SNP and the two possible alleles detected with multiplexes R1 (a) and R2 (b) are indicated with arrows. The alleles present in these samples are highlighted in bold.....XXXIX

Table S1: Y-STR profiles and corresponding haplogroups (predicted and the results obtained in the present study) of the Ecuadorian samples. The region of birth of each individual is also presented in this table.

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
M_S1	Andes	18	13	19	29	14	10	22	12	11	13	15	17	23	23	12	13	10	13	17	11,14	16	12	R1b	100	R-S116*
M_S2	Andes	18	13	21	32	13	10	25	12	10	10	14	17	21	23	13	11	13	13	18	15,20	15	12	E1b1b	85.82	E-M123
M_S3	Andes	17	13	19	29	15	11	22	12	12	13	14	17	23	24	12	13	10	13	17	11,14	16	11	R1b	100	R-S116*
M_S4	Andes	19	13	19	29	14	11	22	12	12	12	15	16	23	23	12	13	10	13	17	11,14	15	10	R1b	100	R-S116*
M_S5	Andes	17	13	21	29	13	10	23	13	11	11	14	17	23	25	15	14	10	13	16	14,19	15	12	Q M346>> M3	92.92	Q-M3*
M_S6	Andes	18	12	19	28	14	11	22	13	12	12	15	16	23	24	14	13	10	13	16	11,14	16	12	R1b	100	R-S116*
M_S7	Andes	16	13	21	28	17	9	22	9	12	10	14	18	21	25	13	11	12	13	16	12,12	14	11	I2a1a Sardinian M26	99.86	I-M26
M_S8	Coast	18	12	21	28	16	10	24	12	9	10	16	17	22	23	11	11	11	14	16	14,15	15	12	G2a2 >> M278	99.99	G-M201
M_S9	Andes	18	13	19	29	14	11	23	12	11	12	15	17	25	24	12	13	10	13	17	11,15	16	12	R1b	100	R-S116*
M_S10	Andes	18	13	20	31	13	10	25	13	13	11	14	17	22	23	12	14	10	14	16	12,17	15	12	Q M346>> M3	92.88	Q-M3*
M_S11	Andes	16	12	20	28	14	10	25	13	11	10	16	16	24	22	11	11	12	13	14	13,15	14	11	I1	100	I-M170*
M_S12	Andes	17	14	20	30	13	9	27	11	11	10	14	21	21	24	10	11	12	13	18	13,14	15	12	E1b1b	100	E-M81
M_S13	Andes	15	13	20	32	13	10	25	12	10	10	14	20	21	24	12	11	12	14	19	16,17	16	11	E1b1b	99.91	E-M123
M_S14	Andes	19	13	18	29	14	11	22	13	12	12	14	18	24	24	12	13	10	13	17	11,13	16	11	R1b	100	R-S116*
M_S15	Andes	18	13	19	29	14	11	23	12	11	12	15	17	24	24	12	13	10	13	17	11,15	16	12	R1b	100	R-S116*
M_S16	Andes	18	13	20	29	14	10	22	12	12	12	15	18	23	25	12	13	10	13	20	12,14	16	12	R1b	100	R-U152
M_S17	Andes	17	12	20	30	13	10	23	12	11	10	15	17	23	25	11	11	13	14	16	15,17	14	12	E1b1b	67.73	E-M35*
M_S18	Andes	18	14	20	30	14	11	22	13	12	12	15	17	23	24	13	13	10	13	16	11,14	14	12	R1b	100	R-U152
M_S19	Andes	18	12	19	28	14	10	24	11	12	11	14	16	22	24	12	15	10	14	15	14,19	16	11	Q M346>> M3	55.44	Q-M3*
M_S20	Andes	17	13	18	29	14	11	22	13	12	12	15	18	24	25	12	13	10	13	17	11,14	15	12	R1b	100	R-S116*
M_S21	Andes	19	14	20	30	13	11	25	11	13	11	14	16	22	24	11	14	10	13	17	13,19	17	10	Q M346>> Y4800	5.86	Q-M3*
M_S22	Andes	16	13	19	29	17	11	21	13	12	9	15	17	24	23	11	11	11	12	15	12,16	15	11	J2a1 Z387	32.74	J-M172
M_S23	Andes	17	13	21	30	15	9	28	13	12	10	14	22	23	22	11	11	12	13	16	13,16	13	11	E1b1b	5.23	E-M35*
M_S24	Andes	19	13	20	29	14	10	22	13	12	12	15	17	23	24	13	13	11	13	18	11,14	17	12	R1b	100	R-S116*
M_S25	Andes	19	13	19	29	14	11	21	13	12	12	15	16	23	24	14	13	11	13	20	11,14	16	12	R1b	100	R-S116*
M_S26	Andes	21	14	20	31	13	11	26	12	13	11	14	16	22	24	13	15	10	17	13	12,19	16	11	Q M346>> M3	0.09	Q-M3*
M_S27	Andes	16	14	20	31	14	10	24	12	12	10	14	17	21	23	12	11	9	12	16.1	13,19	14	11	J1a2a1a2 P58	45.87	J-12f2a*

Table S1 (continued)

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
M_S28	Andes	18	14	21	33	13	10	25	11	10	10	14	17	21	23	13	11	13	13	18	14,19	15	12	E1b1b	72.54	E-M123
M_S29	Andes	18	13	20	30	14	11	25	12	11	12	14	18	22	24	11	14	9	13	17	16,18	15	11	Q M346>> M3	89.55	Q-M3*
M_S30	Amazonia	17	13	19	29	14	11	22	13	12	12	15	17	23	24	12	13	10	13	17	11,14	15	12	R1b	100	R-S116*
M_S31	Coast	16	12	20	28	14	10	25	13	11	10	16	16	24	22	11	11	12	13	15	13,15	14	11	I1	100	I-M170*
M_S32	Galápagos Islands	15	13	19	29	13	10	24	10	11	11	14	17	22	23	12	14	10	14	15	14,16	15	13	Q M346>> Z780	92.62	Q-M346*
M_S33	Andes	17	13	21	30	13	10	24	12	11	9	14	18	22	23	11	14	11	13	15	13,18	15	12	Q M346>> M3	98.31	Q-M3*
M_S34	Coast	19	14	18	30	14	11	22	12	12	9	14	17	24	24	12	13	10	13	18	11,14	16	12	R1b	100	R-S116*
M_S35	Coast	20	13	20	29	14	10	22	13	13	12	15	19	23	24	11	13	10	13	18	11,15	16	12	R1b	100	R-S116*
M_S36	Andes	16	14	21	30	16	11	24	11	12	10	14	17	21	23	10	11	14	14	15	14,15	13	12	I2a2a M223	9.44	I-M26
M_S37	Coast	18	14	19	30	14	11	22	13	12	12	15	19	23	24	13	13	10	13	18	14,14	15	13	R1b	100	R-U152
M_S38	Coast	16	12	21	28	14	10	22	12	11	9	14	17	22	23	10	11	9	12	17	13,17	15	11	J2a1 Z6065	97.08	J-M172
M_S39	Coast	18	13	21	30	14	10	22	13	12	10	14	18	20	22	13	11	12	13	16	18,19	15	12	E1b1b	99.28	E-M78
M_S40	Coast	19	13	20	30	14	9	26	11	11	10	14	22	21	24	10	11	12	13	17	13,13	15	12	E1b1b	99.96	E-M81
M_S41	Andes	17	13	18	29	14	11	22	14	12	12	14	17	23	24	11	13	10	13	17	11,14	16	12	R1b	100	R-M153
M_S42	Coast	18	12	20	30	15	10	23	12	12	11	14	20	23	25	10	11	10	14	15	11,14	15	12	R1a	100	R-SRY10831.2
M_S43	Andes	18	13	19	29	14	10	23	12	13	12	15	18	23	24	11	13	10	13	17	11,13	18	11	R1b	100	R-S116*
M_S44	Andes	15	12	22	28	16	10	21	12	10	10	16	17	21	21	11	11	12	14	18	13,15	15	11	G2a2b1 M406	100	G_M201
M_S45	Coast	15	12	19	28	14	10	24	12	12	10	14	19	24	26	11	11	11	13	18	14,18	16	11	E2 M75	92.31	E-M85
M_S46	Andes	19	13	19	29	14	11	22	13	12	12	15	17	23	24	12	13	10	13	18	11,12	15	12	R1b	100	R-S116*
M_S47	Coast	16	12	19	28	14	11	22	13	12	12	12	16	23	24	12	13	10	13	16	10,14	16	12	R1b	99.93	R-S116*
M_S48	Andes	19	13	19	30	14	11	21	14	12	12	15	17	23	24	12	13	12	13	16	11,14	16	12	R1b	100	R-U152
M_S49	Coast	19	12	20.2	29	15	10	22	12	10	10	16	20	21	22	12	11	11	15	19	12,15	16	12	G2a2a PF3147>> L91	46.55	G-M201
M_S50	Andes	17	13	19	29	14	11	21	14	12	12	15	16	23	24	12	13	10	13	17	11,15	15	12	R1b	100	R-S116*
M_S51	Amazonia	18	13	19	29	14	11	22	12	12	12	15	17	21	23	13	13	10	13	18	11,14	15	12	R1b	100	R-S116*
M_S52	Andes	17	14	18	31	14	11	22	13	12	12	14	18	23	24	13	13	10	13	16	11,11	15	11	R1b	100	R-M529
M_S53	Andes	18	14	20	30	14	11	22	13	12	12	15	17	23	24	12	13	10	13	16	11,14	14	12	R1b	100	R-S116*
M_S54	Andes	21	13	19	31	14	11	23	14	12	12	15	18	23	24	12	13	10	13	18	11,13	17	12	R1b	100	R-S116*
M_S55	Andes	17	13	19	29	14	10	22	13	11	12	15	17	23	24	13	13	10	13	17	11,14	15	12	R1b	100	R-S116*
M_S56	Andes	19	13	18	29	14	11	22	13	12	12	15	17	23	24	11	13	10	13	18	11,15	16	12	R1b	100	R-S116*
M_S57	Andes	18	13	19	29	14	11	22	13	12	12	14	16	23	25	12	14	11	14	18	11,15	16	12	R1b	100	R-S116*
M_S58	Andes	19	14	20	31	13	10	26	12	12	11	14	17	23	23	13	16	10	14	16	15,19	15	12	Q M346>> M3	57.53	Q-Z19483

Table S1 (continued)

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
M_S59	Andes	18	13	21	30	13	10	23	13	11	11	14	17	22	25	12	16	10	14	17	15,16	15	11	Q M346>> M3	99.64	Q-M3*
M_S60	Andes	16	12	19	28	14	10	25	11	11	11	14	17	22	25	12	14	9	13	16	15,19	16	11	Q M346>> Z780	71.62	Q-M3*
M_S61	Coast	19	13	19	31	14	10	24	12	10	11	14	17	22	25	12	14	11	13	16	15,17	16	12	Q M346>> M3	98.79	Q-M3*
M_S62	Coast	19	13	21	28	17	10	22	10	12	10	15	21	22	23	11	11	11	13	19	12,12	13	12	I2a1a Sardinian M26	100	I-M26
M_S63	Coast	19	13	21	30	14	10	26	13	11	10	14	18	22	23	13	11	9	12	18.2	13,15	15	11	J1a2a1a2 P58	99.95	J-12f2a*
M_S64	Coast	18	13	19	29	14	10	22	13	11	12	16	16	23	24	13	13	11	13	16	11,14	16	12	R1b	100	R-U152
M_S65	Coast	16	13	20	30	13	9	27	12	12	10	14	15	21	23	14	11	11	14	17	16,17	15	12	E1b1b	85.85	E-M123
M_S66	Andes	15	13	20	32	13	10	25	12	10	10	14	20	21	24	12	11	12	14	19	16,17	16	11	E1b1b	99.91	E-M123
M_S67	Andes	19	12	21	28	15	10	22	14	9	10	16	18	21	22	12	11	11	14	16	12,16	17	12	G2a2b2a1c CTS342	89.05	G-M201
M_S68	Andes	19	13	17	29	14	11	22	13	12	12	14	17	23	24	12	13	10	13	17	11,14	16	11	R1b	100	R-S116*
M_S69	Coast	17	13	20	28	14	12	22	13	12	12	15	16	24	24	13	15	10	12	19	11,16	16	12	R1b	97.53	R-U152
M_S70	Andes	18	13	22	28	17	11	22	10	13	10	15	19	23	23	11	11	12	13	17	12,12	14	12	I2a1a Sardinian M26	100	I-M26
M_S71	Andes	21	13	20	29	13	9	27	12	11	10	14	22	21	24	10	11	12	13	19	13,14	18	12	E1b1b	99.96	Q-M346*
M_S72	Coast	17	12	21	29	15	10	21	15	9	10	16	19	23	23	12	11	11	14	16	14,16	16	12	G2a2b2a1c CTS342	65.98	G-M201
M_S73	Andes	18	13	19	28	15	10	23	13	12	10	14	21	21	22	11	15	11	13	17	11,14	16	12	N1a2 CTS6380	0.17	KL-M9*
M_S74	Coast	16	13	20	31	13	10	24	12	12	10	14	17	20	23	12	11	14	13	16	15,16	15	10	E1b1b	85.07	E-M78
M_S75	Coast	18	13	20	29	14	10	25	12	11	11	14	18	26	24	11	14	10	13	16	13,14	15	12	Q M346>> M3	17.21	Q-M3*
M_S76	Coast	18	13	21	27	16	10	22	9	12	10	15	18	21	22	12	11	11	12	16	12,12	14	12	I2a1a Sardinian M26	100	I-M26
M_S77	Coast	20	13	20	31	13	10	25	12	12	11	14	17	22	24	13	14	10	13	17	12,18	16	12	Q M346>> M3	98.44	Q-M3*
M_S78	Coast	18	13	19	30	13	10	24	12	12	11	14	18	22	24	13	14	10	14	15	15,16	16	12	Q M346>> M3	64.48	Q-M3*
M_S79	Andes	16	12	20	28	14	11	26	12	11	10	16	19	21	22	11	11	12	13	16	13,15	14	11	I1	100	I-M170*
M_S80	Andes	18	12	20	29	11	10	23	12	13	10	14	23	22	24	10	11	12	13	15	18,18	15	11	E1b1b	61.97	E-M78
M_S81	Coast	17	13	21	31	14	10	23	12	11	10	14	18	21	23	12	11	9	12	18.2	16,17	16	11	J1a2 > ZS6591	28.45	J-12f2a*
M_S82	Coast	18	13	19	29	14	10	22	13	12	12	15	18	23	26	12	13	10	13	17	11,14	16	12	R1b	100	R-S116*

Table S1 (continued)

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
M_S83	Coast	19	13	20	30	14	9	24	11	11	11	14	18	22	24	13	13	10	15	17	14,14	14	12	Q M346>> M3	29.68	Q-M3*
M_S84	Andes	18	13	19	29	14	11	20	13	11	12	15	16	23	26	11	13	10	13	17	11,14	16	12	R1b	100	R-U106
M_S85	Andes	19	13	19	30	14	10	22	12	12	12	15	17	23	24	11	13	10	13	17	11,14	16	12	R1b	100	R-L23*
M_S86	Coast	16	13	21	31	15	11	29	11	11	11	12	21	21	21	11	11	14	13	16	17,17	15	12	E1b1a V38	99.87	E-U209*
M_S87	Coast	19	13	19	29	14	11	22	12	13	12	15	15	24	24	12	13	10	13	17	9,15	15	11	R1b	99.99	R-M153
M_S88	Coast	19	13	19	31	13	10	24	12	11	11	14	18	23	23	12	14	10	13	16	14,17	15	11	Q M346>> M3	97.59	Q-M3*
M_S89	Coast	19	13	19	29	13	11	22	14	12	12	15	20	23	24	13	13	10	13	17	11,14	15	12	R1b	100	R-L23*
M_S90	Andes	13	14	20	30	14	10	21	11	11	10	15	13	22	23	13	11	10	12	15	14,14	16	11	J2a1 Z6065	1.58	J-M172
M_S91	Coast	15	13	21	32	15	11	28	11	11	11	14	20	21	20	11	11	14	12	16	17,17	16	12	E1b1a V38	99.59	E-U209*
M_S92	Andes	17	14	19	30	14	10	22	13	12	12	15	18	23	24	11	13	10	13	17	11,14	15	12	R1b	100	R-S116*
M_S93	Andes	18	13	20	31	13	10	25	13	12	11	14	17	22	23	13	14	10	13	17	13,17	15	11	Q M346>> L330	63.34	Q-M3*
M_S94	Andes	19	14	20	31	12	11	24	12	12	11	14	17	22	25	12	14	10	14	18	12,18	15	11	Q F1096	31.38	Q-M3*
M_S95	Andes	19	13	19	29	15	10	22	14	11	12	15	17	23	23	12	13	10	13	17	11,14	16	12	R1b	100	R-U152
M_S96	Andes	18	13	18	29	14	11	23	13	12	12	15	17	23	24	12	13	10	14	17	11,14	15	12	R1b	100	R-S116*
M_S97	Andes	16	13	19	30	15	11	22	14	12	9	14	18	21	23	11	13	10	13	18	14,16	15	11	T>PF5633	99.93	T-M70
M_S98	Andes	17	14	20	31	13	10	25	12	12	11	15	17	22	22	13	14	10	13	16	15,15	15	12	Q M346>> Z780	47.34	Q-M3*
M_S99	Andes	17	13	18	29	16	11	22	13	12	12	14	18	23	23	13	13	10	13	18	11,14	15	12	R1b	100	R-S116*
M_S100	Andes	17	13	21	31	13	10	24	12	11	9	14	18	22	23	11	14	10	13	15	14,18	15	12	Q M346>> M3	94.47	Q-M3*
M_S101	Andes	18	14	20	30	13	9	28	11	11	10	14	21	21	24	10	11	12	13	18	13,14	15	12	E1b1b	100	E-M81
M_S102	Andes	18	13	20	30	13	11	26	15	13	11	14	17	22	24	11	14	10	13	16	13,19	16	12	Q M346>> M3	87.58	Q-M3*
M_S103	Andes	18	14	21	30	15 - 16	10	21	12	10	10	16	21	21	22	12	11	11	14	15	13,14	15	13	G2a2a PF3147>> L91	65.31	G-M201
M_S104	Andes	20	13	19	32	13	10	25	12	12	11	14	18	22	24	13	14	10	13	18	12,18	16	12	Q M346>> M3	82.32	Q-M3*
M_S105	Andes	18	13	19	29	13	9	27	11	11	10	14	22	21	24	10	11	13	13	18	13,14	16	12	E1b1b	100	E-M81
M_S106	Andes	15	12	20	27	14	10	24	12	11	9	14	17	22	24	11	14	10	13	17	14,16	15	11	T>PF5633	99.61	T-M70
M_S107	Coast	16	12	19	28	14	10	21	13	14	12	16	18	23	25	14	13	11	14	15	11,14	15	12	R1b	94.56	R-U152
M_S108	Coast	16	14	21	31	17	11	25	11	11	11	14	15	21	21	12	11	13	14	16	17,19	17	11	E1b1a V38	100	E-U174
M_S109	Coast	20	13	19	29	14	11	23	13	12	13	15	17	24	24	12	13	10	13	17	11,14	16	11	R1b	100	R-S116*
M_S110	Coast	18	13	18	30	13	10	23	13	11	11	14	17	23	23	11	14	10	13	16	15,15	16	11	Q M346>> Z780	96.83	Q-M346*
M_S111	Coast	17	14	18	31	14	11	22	14	12	12	14	18	24	25	12	13	10	13	17	11,14	15	11	R1b	100	R-S116*

Table S1 (continued)

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
M_S112	Coast	18	12	20	28	13	10	24	12	11	11	14	17	23	25	13	14	10	12	16	15,15	15	13	Q M346>> Z780	45.65	Q-M3*
M_S113	Coast	17	13	19	30	13	10	24	13	10	11	14	17	22	24	13	14	10	13	17	15,17	16	12	Q M346>> M3	87.86	Q-M3*
M_S114	Coast	16	12	20	29	14	10	26	12	11	10	16	20	22	23	11	11	12	13	16	13,15	15	11	I1	100	I-M170*
M_S115	Coast	19	13	19	29	14	11	22	12	13	12	15	19	23	24	13	13	10	13	17	14,14	15	13	R1b	100	R-U152
M_S116	Coast	18	13	19	29	14	11	24	13	12	12	15	18	23	24	12	13	10	13	17	10,14	16	12	R1b	100	R-S116*
M_S117	Coast	18	13	18	29	13	11	22	12	12	12	14	17	23	24	12	13	10	13	17	11,14	16	11	R1b	100	R-S116*
M_S118	Coast	18	14	20	30	13	10	25	11	11	11	14	17	22	24	11	14	10	13	17	13,20	17	11	Q M346>> Z780	72.54	Q-M3*
M_S119	Coast	20	13	19	29	14	11	23	13	12	13	15	17	24	24	12	13	10	13	17	11,14	16	11	R1b	100	R-M529
M_S120	Coast	21	12	20	29	14	10	25	13	12	11	14	20	22	24	13	14	10	14	16	12,17	15	11	Q M346>> Z780	36.16	Q-M3*
M_S121	Coast	19	13	19	29	14	12	22	12	12	12	15	17	22	23	12	13	10	13	17	12,15	15	12	R1b	100	R-S116*
M_S122	Coast	17	13	18	29	14	10	22	13	12	12	14	19	23	24	12	13	10	13	18	11,14	16	12	R1b	100	R-M153
M_S123	Coast	17	13	21	31	14	10	25	13	12	10	14	20	21	23	12	11	9	12	17 .2	13,19	15	11	J1a2a1a2 P58	99.8	J-12f2a*
M_S124	Andes	18	13	19	29	14	11	22	14	12	12	15	17	23	24	11	14	10	11	17	12,13	15	12	R1b	100	R-S116*
M_S125	Amazonia	19	12	20	29	15	10	23	13	11	10	14	17	21	24	12	11	12	13	15	17,18	15	11	E1b1b	99.98	E-M78
M_S126	Amazonia	19	13	20	28	14	11	24	12	12	11	14	16	22	24	13	14	10	13	17	14,17	14	11	Q M346>> M3	85.74	Q-M3*
M_S127	Andes	16	13	21	31	15	11	28	11	11	11	14	20	21	21	11	11	14	13	16	16,17	15	12	E1b1a V38	100	E-U209*
M_S128	Andes	16	13	21	28	17	9	22	9	13	10	14	19	21	24	13	11	12	13	16	12,12	14	11	I2a1a Sardinian M26	99.68	I-M26
M_S129	Andes	16	13	21	29	15	9	22	12	12	9	14	17	21	23	13	11	10	12	14	13,16	17	12	J2a1 Z387	100	J-M172
M_S130	Andes	17	13	20	31	15	10	25	12	11	10	14	17	20	23	11	12	9	12	17 .2	13,18	15	11	J1a2a1a2 P58	99.92	J-12f2a*
M_S131	Coast	17	14	20	31	13	10	25	13	11	11	14	18	22	24	12	14	10	13	14	14,18	15	11	Q M346>> M3	86.14	Q-M3*
M_S132	Coast	16	13	21	31	15	11	30	11	11	11	12	21	21	21	11	11	14	13	16	17,17	15	12	E1b1a V38	99.45	E-U209*
M_S133	Coast	16	12	20	30	13	9	25	13	11	10	14	18	21	24	11	11	12	14	16	17,18	14	12	E1b1b	96.32	E-M123
M_S134	Coast	16	14	18	30	14	11	22	13	12	12	14	17	23	24	12	13	10	13	18	10,13	17	11	R1b	100	R-S116*
M_S135	Coast	18	13	19	30	13	10	24	12	10	11	14	17	22	24	11	14	11	13	17	15,17	16	12	Q M346>> M3	99.93	Q-M3*
M_S136	Coast	17	12	21	29	15	10	23	12	10	10	16	17	21	22	12	10	11	14	18	14,15	14	11	G2a1 Z6552	94.4	G-M201
M_S137	Andes	17	13	20	30	13	10	23	12	12	10	14	21	21	25	13	11	12	13	15	16,16	15	12	E1b1b	100	E-M78
M_S138	Andes	19	12	20	30	13	10	22	11	11	8	14	17	22	24	11	14	10	14	17	14,16	15	11	Q M346>> M3	11.04	Q-M346*

Table S1 (continued)

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
M_S139	Andes	18	13	21	31	16	11	24	11	11	10	16	18	22	22	12	11	11	14	17	13,15	15	12	G2a2a PF3147>> L91	96.17	G-M201
M_S140	Andes	17	13	18	29	14	11	22	12	12	12	14	17	23	24	11	13	10	13	17	11,14	17	12	R1b	100	R-M153
M_S141	Coast	17	13	19	29	14	11	23	14	14	12	14	17	23	24	12	13	10	13	17	11,13	15	12	R1b	100	R-S116*
M_S142	Coast	18	14	18	31	14	11	23	13	12	12	14	16	23	24	12	13	10	13	18	11,13	15	11	R1b	100	R-S116*
M_S143	Coast	16	12	21	28	14	10	22	12	11	9	14	18	22	23	10	11	9	12	17	13,17	15	11	J2a1 Z6065	95.14	J-M172
M_S144	Coast	18	13	20	30	14	10	26	13	12	10	14	17	20	23	11	11	9	12	17.2	13,20	15	11	J1a2a1a2 P58	100	J-12f2a*
M_S145	Andes	17	13	19	29	14	11	22	14	12	12	15	18	23	24	12	13	10	13	17	12,15	16	12	R1b	100	R-S116*
M_S146	Andes	14	14	21	32	15	10	27	11	11	11	14	20	21	21	12	11	13	14	17	15,20	15	12	E1b1a V38	100	E-U209*
M_S147	Amazonia	19	13	20	29	13	10	24	12	11	11	13	18	22	24	12	14	10	14	18	14,18	15	12	Q M346>> M3	95.22	Q-M3*
M_S148	Amazonia	19	13	20	30	13	10	26	13	12	11	14	16	22	24	12	14	10	13	17	13,17	16	12.1	Q M346>> M3	82.44	Q-M3*
M_S149	Amazonia	18	13	19	29	14	11	22	12	12	12	15	17	23	24	12	13	10	13	19	11,15	15	12	R1b	100	R-M529
A_S150	Andes	15	13	21	30	15	11	27	11	11	11	14	21	24	22	12	11	14	13	17	15,17	15	11	E1b1a V38	100	E-U290
A_S151	Andes	15	13	21	30	15	10	23	11	10	11	14	19	22	21	12	11	15	14	18	14,16	15	11	E1b1a V38	99.69	E-M2*
A_S152	Andes	17	13	21	30	15	10	26	11	10	11	14	17	21	21	12	11	13	15	16	17,18	17	11	E1b1a V38	100	E-U174
A_S153	Costa	17	14	21	32	16	10	25	12	11	11	14	14	21	22	12	11	12	13	17	15,18	15	10	E1b1a V38	99.85	E-M2*
A_S154	Andes	19	13	20	30	13	10	24	12	11	11	14	17	22	23	11	15	10	13	17	15,17	16	11	Q M346>> M3	99.72	Q-M3*
A_S155	Costa	16	14	21	31	16	10	25	11	12	11	14	17	22	21	12	12	12	15	16	18,20	15	11	E1b1a V38	99.87	E-U174
A_S156	Costa	15	11	19	28	14	10	25	13	13	11	14	19	23	26	11	11	11	13	18	14,19	15	11	E2 M75	97.14	E-M85
A_S157	Costa	15	13	21	30	15	9	26	11	11	11	14	21	21	21	12	11	13	13	15	16,17	15	12	E1b1a V38	100	E-U209*
A_S158	Costa	19	13	23	27	15	11	24	9	12	12	14	21	18	23	11	12	8	15	16	10,12	15	11	I2a1a Sardinian M26	0	B-M112*
A_S159	Costa	16	13	20	30	15	10	25	12	12	11	14	18	21	21	12	11	13	13	18	15,16	15	12	E1b1a V38	100	E-M2*
A_S160	Costa	16	13	21	30	15	10	25	11	11	11	14	17	21	21	10	11	13	15	17	16,19	15	11	E1b1a V38	100	E-U174
A_S161	Costa	15	13	21	30	15	10	28	11	11	11	14	20	21	21	12	11	14	13	17	16,17	15	12	E1b1a V38	100	E-U290
A_S162	Costa	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,17	17	11	E1b1a V38	100	E-U174
A_S163	Costa	18	14	19	32	15	10	26	11	12	11	14	21	22	21	12	11	13	15	18	15,16	15	10	E1b1a V38	81.66	E-M2*
A_S164	Costa	17	13	20	29	15	11	27	13	12	10	15	20	20	23	11	12	13	14	15	15,16	15	11	I2a2a M223	99.99	I-M170*
A_S165	Costa	13	13	20	30	15	10	25	10	11	11	14	17	24	21	13	11	13	13	16	16,18	15	12	E1b1a V38	99.69	E-M191*
A_S166	Costa	15	13	21	30	15	10	27	11	11	11	14	19	21	21	12	11	14	13	18	15,19	15	12	E1b1a V38	100	E-U290
A_S167	Costa	17	13	19	28	17	10	22	12	12	11	14	19	23	25	12	13	12	13	15	12,12	14	11	R1b V88	8.14	R-V88
A_S168	Costa	18	13	19	29	14	11	22	13	12	12	15	18	23	23	12	13	10	13	15	11,14	16	12	R1b	100	R-S116*
A_S169	Costa	17	13	21	30	14	11	21	12	11	9	14	16	23	22	11	11	8	13	15	12,15	16	11	J2a1 L26>Z500	96.15	J-M172

Table S1 (continued)

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
A_S170	Costa	19	13	20	30	15	10	26	11	11	11	13	18	22	21	12	11	13	15	15	16,17	16	11	E1b1a V38	99.87	E-U174
A_S171	Costa	18	13	19	29	14	11	22	13	12	12	15	18	23	23	12	13	10	13	15	11,14	16	12	R1b	100	R-S116*
A_S172	Costa[3]	17	13	20	31	15	10	28	11	11	11	14	18	21	21	13	11	14	14	18	16,17	15	11	E1b1a V38	100	E-U290
A_S173	Costa[3]	18	14	20	31	16	10	25	11	11	11	14	17	21	21	12	11	12	14	17	17,19	16	11	E1b1a V38	100	E-U174
A_S174	Costa[3]	16	13	20	31	15	10	28	11	11	11	14	18	21	21	13	11	14	14	18	16,17	15	11	E1b1a V38	100	E-U290
A_S175	Costa	15	13	21	30	16	10	24	12	11	11	14	19	21	20	12	11	13	14	18	15,18	16	11	E1b1a V38	99.87	E-M2*
A_S176	Costa	15	13	20	30	15	10	26	11	12	11	14	19	23	21	12	11	12	14	19	16,16	15	10	E1b1a V38	99.26	E-M2*
A_S177	Costa	17	13	20	29	15	11	27	13	12	10	15	20	20	23	11	12	13	14	15	15,16	15	11	I2a2a M223	99.99	I-M170*
A_S178	Costa	17	13	21	31	16	10	25	11	11	11	14	17	23	21	12	10	12	14	15	17,17	17	11	E1b1a V38	99.69	E-U174
A_S179	Costa	15	13	21	30	15	10	28	11	11	11	14	21	21	22	12	11	14	13	17	15,15	15	12	E1b1a V38	100	E-U290
A_S180	Costa[3]	17	14	19	32	15	10	26	11	12	11	14	21	22	21	12	11	13	15	18	15,16	15	10	E1b1a V38	92.77	E-M2*
A_S181	Costa	16	13	21	30	17	10	28	11	12	11	14	14	21	22	12	11	12	13	16	16,17	15	10	E1b1a V38	99.89	E-M2*
A_S182	Andes	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,17	17	11	E1b1a V38	100	E-U174
A_S183	Costa	18	14	21	31	17	10	26	11	12	11	14	16	21	21	11	11	13	14	17	18,18	15	11	E1b1a V38	99.69	E-U174
A_S184	Costa	15	13	21	31	15	10	27	11	11	11	14	19	21	21	12	11	12	13	16	16,17	15	12	E1b1a V38	100	E-U209*
A_S185	Costa	18	14	18	31	14	11	22	12	13	12	14	17	23	24	12	13	10	13	17	11,14	15	11	R1b	100	R-S116*
A_S186	Costa	15	13	21	30	15	10	26	11	11	11	14	19	21	21	12	11	15	13	15	16,17	16	12	E1b1a V38	100	E-U290
A_S187	Costa	16	14	21	31	16	10	25	11	11	11	14	15	20	21	12	11	13	14	16	17,17	17	11	E1b1a V38	100	E-U174
A_S188	Costa[3]	16	12	20	29	15	10	25	11	12	11	14	17	21	21	12	11	13	14	18	15,16	16	11	E1b1a V38	99.69	E-M2*
A_S189	Costa	17	13	20	30	15	11	29	11	13	11	14	18	21	21	11	11	13	14	18	16,16	15	11	E1b1a V38	99.89	E-M2*
A_S190	Costa	14	13	22	30	15	10	29	11	11	11	14	22	20	21	13	11	14	14	17	15,19	15	12	E1b1a V38	99.45	E-U209*
A_S191	Costa	17	13	20	32	14	10	25	12	11	10	14	19	21	23	12	11	9	12	17.2	13,18	16	11	J1a2a1a2 P58	99.96	J-12f2a*
A_S192	Costa	15	13	21	32	15	10	28	11	11	11	14	18	21	21	12	11	14	13	17	14,16	15	12	E1b1a V38	100	E-U290
A_S193	Costa	16	13	21	31	15	10	26	12	11	11	14	19	21	21	12	11	15	13	18	16,16	15	12	E1b1a V38	100	E-U290
A_S194	Costa	19	14	21	31	17	10	24	11	11	12	14	17	21	21	12	11	13	15	16	17,18	16	11	E1b1a V38	100	E-U174
A_S195	Costa	17	14	21	32	16	10	26	11	11	11	14	20	22	21	13	11	13	13	17	17,17	15	11	E1b1a V38	100	E-U209*
A_S196	Costa	17	13	19	30	15	10	25	12	12	11	15	18	23	21	11	11	12	14	17	15,17	15	11	E1b1a V38	92.66	E-M2*
A_S197	Costa	14	13	21	30	15	10	29	11	12	10	14	19	21	22	12	11	14	13	18	16,16	15	12	E1b1a V38	99.87	E-U290
A_S198	Costa	15	13	21	30	15	10	28	11	11	11	14	21	21	22	12	11	14	13	17	15,15	15	12	E1b1a V38	100	E-U290
A_S199	Costa	15	12	20	29	15	10	25	11	12	11	14	17	21	21	12	11	13	14	18	15,16	16	11	E1b1a V38	99.87	E-M2*
A_S200	Costa	15	13	21	30	15	10	26	11	11	11	14	19	21	21	12	11	15	13	15	16,17	16	12	E1b1a V38	100	E-U290
A_S201	Costa	17	13	21	30	17	10	26	12	11	12	14	18	19	21	13	11	13	15	16	16,18	15	11	E1b1a V38	99.26	E-U174
A_S202	Costa	18	14	20	31	16	10	25	11	11	11	14	17	21	21	12	11	12	13	17	17,19	16	11	E1b1a V38	100	E-U174
A_S203	Costa	15	13	21	31	14	10	27	11	11	11	14	18	21	21	12	11	12	13	16	16,17	15	12	E1b1a V38	100	E-U209*
A_S204	Costa	18	14	21	31	17	10	27	12	12	11	14	16	21	21	11	11	13	15	17	18,19	15	11	E1b1a V38	99.69	E-U174
A_S205	Costa	15	13	21	30	15	10	28	11	11	11	14	20	21	21	12	11	14	13	18	16,17	15	12	E1b1a V38	100	E-U290
A_S206	Costa	19	13	21	30	17	11	27	11	11	11	14	16	20	21	12	11	13	15	16	18,19	15	10	E1b1a V38	99.26	E-U174
A_S207	Costa	15	12	21	30	15	10	28	11	11	11	14	18	21	21	15	11	14	13	17	16,18	15	11	E1b1a V38	100	E-U290

Table S1 (continued)

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
A_S208	Costa	18	14	21	31	17	10	26	11	12	11	14	16	21	21	11	11	13	14	17	18,18	15	11	E1b1a V38	99.69	E-U174
A_S209	Costa	16	12	20	29	15	10	27	11	12	11	14	19	23	21	12	11	14	14	16	16,18	17	11	E1b1a V38	99.26	E-M2*
A_S210	Costa	15	13	21	31	15	10	28	11	11	11	14	19	21	21	11	11	13	13	17	16,17	15	12	E1b1a V38	100	E-U209*
A_S211	Costa	15	13	21	30	15	10	26	11	11	11	14	19	21	21	12	11	15	13	15	16,17	16	12	E1b1a V38	100	E-U290
A_S212	Costa	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,17	17	11	E1b1a V38	100	E-U174
A_S213	Costa	16	13	21	30	15	10	26	11	12	11	14	17	21	22	12	11	13	14	16	17,17	15	11	E1b1a V38	100	E-U174
A_S214	Costa	18	13	21	30	15	10	26	11	10	11	14	17	21	21	12	11	13	15	18	15,15	15	11	E1b1a V38	100	E-M2*
A_S215	Costa	16	12	19	28	15	10	24	12	12	9	16	19	22	24	12	11	9	12	16	13,17	13	11	J2b2a M241	100	J-M172
A_S216	Costa	15	12	20	29	15	10	25	11	12	11	14	17	21	21	12	11	13	14	18	15,16	16	11	E1b1a V38	99.87	E-M2*
A_S217	Costa	17	13	20	30	15	11	28	11	13	11	14	18	21	21	11	11	13	14	18	16,16	15	11	E1b1a V38	99.87	E-M2*
A_S218	Costa	16	13	21	30	16	10	25	11	12	11	14	17	22	21	12	12	12	15	16	18,19	15	11	E1b1a V38	99.69	E-U174
A_S219	Costa	16	14	21	32	15	10	25	11	12	11	14	19	22	21	12	11	14	13	17	16,17	15	12	E1b1a V38	100	E-U290
A_S220	Costa	18	13	18	29	14	11	22	12	12	12	14	18	23	24	12	13	10	13	18	11,14	16	11	R1b	100	R-S116*
A_S221	Costa	18	14	20	30	13	9	27	11	11	10	14	22	21	24	10	11	12	13	20	13,14	17	12	E1b1b	100	E-M81
A_S222	Costa	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	18	17,17	17	11	E1b1a V38	100	E-U174
A_S223	Costa	15	13	21	31	15	10	27	11	11	11	14	19	21	21	12	11	12	13	16	16,17	15	12	E1b1a V38	100	E-U209*
A_S224	Costa	17	13	21	30	17	10	25	11	12	11	14	17	21	21	12	11	13	15	16	15,18	15	11	E1b1a V38	100	E-U174
A_S225	Costa	16	13	21	30	16	10	25	11	12	11	14	17	22	21	12	12	12	15	16	18,19	15	11	E1b1a V38	99.69	E-U174
A_S226	Costa	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,17	17	11	E1b1a V38	100	E-U174
A_S227	Costa[3]	16	13	20	31	15	10	28	11	11	11	14	18	21	21	13	11	14	14	18	16,17	15	11	E1b1a V38	100	E-U290
A_S228	Costa[3]	15	13	21	31	16	11	23	12	10	11	14	18	22	21	10	11	14	13	16	15,16	16	11	E1b1a V38	99.26	E-M2*
A_S229	Costa[3]	17	14	19	32	15	10	26	11	12	11	14	21	22	21	12	11	13	15	18	15,16	15	10	E1b1a V38	92.77	E-M2*
A_S230	Costa[3]	17	14	19	32	15	10	26	11	12	11	14	21	22	21	12	11	13	15	19	15,16	15	10	E1b1a V38	81.66	E-M2*
A_S231	Costa[3]	16	14	20	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,18	17	11	E1b1a V38	100	E-U174
A_S232	Costa[3]	15	12	19	28	14	11	26	13	11	11	15	18	23	25	11	11	12	13	17	14,14	15	11	E2 M75	100	E-M85
A_S233	Costa[3]	15	13	21	31	16	11	23	12	10	11	14	18	22	21	10	11	14	13	16	15,16	16	11	E1b1a V38	99.26	E-M2*
A_S234	Costa[3]	18	14	19	32	15	10	26	11	12	11	14	21	22	21	12	12	13	15	18	15,16	15	10	E1b1a V38	18.77	E-M2*
A_S235	Costa[3]	16	14	21	31	16	10	25	11	11	11	14	15	20	21	12	11	13	14	16	17,17	17	11	E1b1a V38	100	E-U174
A_S236	Costa	17	14	19	32	15	10	26	11	12	11	14	21	22	21	12	11	13	15	17	15,16	15	10	E1b1a V38	93.21	E-M2*
A_S237	Costa[3]	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,18	17	11	E1b1a V38	100	E-U174
A_S238	Costa	16	12	20	29	15	10	25	11	12	11	14	17	21	21	12	11	13	14	18	15,16	16	11	E1b1a V38	99.69	E-M2*
A_S239	Costa[3]	18	14	20	31	16	10	25	11	11	11	14	17	21	21	12	11	12	14	17	17,19	16	11	E1b1a V38	100	E-U174
A_S240	Costa[3]	15	12	19	28	14	11	26	13	11	11	15	18	23	25	11	11	12	13	17	14,14	15	11	E2 M75	100	E-M85
A_S241	Costa	18	13	19	29	14	11	22	12	12	13	15	16	24	27	12	13	10	13	17	11,14	16	11	R1b	100	R-S116*
A_S242	Costa	16	13	19	29	15	10	22	12	12	11	15	18	24	24	12	13	11	13	16	13,15	15	12	R1b V88	100	R-P25*
A_S243	Costa[3]	17	14	18	30	14	11	22	13	12	12	14	17	23	24	11	13	10	13	17	11,14	15	11	R1b	100	R-M153
A_S244	Costa	17	13	20	29	15	11	27	13	12	10	15	19	20	23	11	12	13	14	15	15,17	15	11	I2a2a M223	99.99	I-M170*
A_S245	Costa[3]	15	13	21	31	16	11	23	12	10	11	14	18	22	21	10	11	14	13	16	15,16	16	11	E1b1a V38	99.26	E-M2*

Table S1 (continued)

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
A_S246	Costa[3]	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,18	17	11	E1b1a V38	100	E-U174
A_S247	Costa[3]	15	12	19	28	14	11	26	13	11	11	15	17	23	25	11	11	12	13	17	14,14	15	11	E2 M75	100	E-M85
A_S248	Costa[3]	18	14	20	31	16	10	25	11	11	11	14	17	21	21	12	11	12	14	17	17,19	16	11	E1b1a V38	100	E-U174
A_S249	Costa	16	14	21	31	16	10	25	11	11	10	14	17	20	21	14	11	13	15	16	17,17	14	10	E1b1a V38	93.21	E-U174
A_S250	Costa[3]	18	14	19	32	15	10	26	11	12	11	14	21	22	21	12	12	13	15	18	15,16	15	10	E1b1a V38	18.77	E-M2*
A_S251	Costa	18	12	21	31	15	11	29	10	11	11	14	21	22	21	11	11	14	13	17	16,18	15	12	E1b1a V38	99.26	E-U290
A_S252	Costa[3]	18	14	20	31	16	10	25	11	11	11	14	17	21	21	12	11	12	14	17	17,19	16	11	E1b1a V38	100	E-U174
A_S253	Andes	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,18	17	11	E1b1a V38	100	E-U174
A_S254	Costa[3]	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,18	17	11	E1b1a V38	100	E-U174
A_S255	Costa	18	13	20	31	14	10	25	11	11	10	14	19	22	23	12	11	9	12	18.2	13,18	16	11	J1a2a1a2 P58	99.66	J-12f2a*
A_S256	Costa[3]	16	14	21	31	17	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,18	17	11	E1b1a V38	100	E-U174
A_S257	Costa	18	14	21	31	16	10	25	11	11	11	14	18	21	21	12	11	12	15	17	17,19	17	11	E1b1a V38	100	E-U174
A_S258	Costa[3]	17	14	19	32	15	10	26	11	12	11	14	21	22	21	12	11	13	15	18	15,16	15	10	E1b1a V38	92.77	E-M2*
A_S259	Andes	17	14	21	31	17	10	27	11	11	11	14	16	20	21	12	11	12	15	16	18,18	15	10	E1b1a V38	99.69	E-U174
A_S260	Andes	18	13	20	30	16	10	25	11	11	11	14	17	24	21	12	11	13	13	15	16,18	16	11	E1b1a V38	100	E-M191*
A_S261	Andes	17	13	21	30	17	10	25	13	11	11	13	17	21	21	12	11	13	13	14	17,19	15	11	E1b1a V38	100	E-U174
A_S262	Andes	15	13	21	29	14	10	25	12	11	10	16	18	22	23	11	11	12	13	16	13,14	14	11	I1	100	I-M170*
A_S263	Andes	19	13	19	29	14	10	22	13	12	12	15	17	23	25	12	11	10	11	16	11,14	15	12	R1b	100	R-L23*
A_S264	Andes	17	14	18	30	14	11	22	13	12	12	14	17	23	24	11	13	10	13	17	11,14	15	11	R1b	100	R-M153
A_S265	Andes	15	13	19	31	15	10	27	12	12	11	14	19	21	21	13	11	14	13	19	16,17	16	11	E1b1a V38	99.45	E-U290
A_S266	Andes	17	12	21	31	15	10	26	12	9	11	14	21	21	21	12	11	13	13	18	17,17	15	12	E1b1a V38	99.26	E-U290
A_S267	Costa	17	13	19	30	14	11	23	13	12	12	15	17	23	24	13	13	10	13	15	11,14	15	12	R1b	100	R-S116*
A_S268	Andes	18	13	19	30	14	10	26	13	11	11	14	15	22	24	12	11	11	14	18	14,18	15	11	Q M346>> M3	72.39	E-M96*
A_S269	Costa	18	12	20	29	13	10	23	11	12	10	14	23	22	24	10	11	12	13	15	18,18	15	11	E1b1b	99.6	E-M78
A_S270	Costa	18	14	18	31	14	11	22	13	12	12	14	18	23	24	12	13	10	13	17	11,11	15	11	R1b	100	R-S116*
A_S271	Andes	15	12	20	28	14	11	25	12	12	11	14	19	23	24	12	11	11	13	19	14,20	16	11	E2 M75	98.11	E-M85
A_S272	Andes	15	13	21	30	15	10	23	11	10	11	14	19	22	21	12	11	15	14	19	14,16	15	11	E1b1a V38	99.87	E-M2*
A_S273	Andes	15	13	21	31	15	11	27	11	11	11	14	20	22	21	11	11	14	13	16	16,17	15	12	E1b1a V38	100	E-U209*
A_S274	Andes	19	13	21	30	14	10	26	13	11	10	15	18	21	23	12	11	9	12	17.2	13,15	15	11	J1a2a1a2 P58	99.85	J-12f2a*
A_S275	Andes	19	13	19	29	14	10	22	13	12	12	15	17	23	25	12	11	10	12	16	11,14	15	12	R1b	100	R-L23*
A_S276	Andes	18	12	19	29	14	10	26	12	10	10	14	19	21	24	12	11	12	13	18	16,20	15	11	E1b1b	99.6	E-M123
A_S277	Andes	15	13	22	30	15	10	27	11	11	11	14	19	21	21	11	11	13	13	16	15,19	15	12	E1b1a V38	100	E-U209*
A_S278	Andes	14	13	19	30	15	10	28	11	12	10	14	19	21	22	12	11	14	13	18	16,16	15	12	E1b1a V38	99.25	E-U290
A_S279	Andes	15	13	19	28	14	12	22	14	12	12	15	16	23	24	12	13	10	13	18	11,14	15	13	R1b	99.99	R-U106
A_S280	Andes	14	13	21	30	15	10	26	11	12	11	14	19	21	21	11	11	14	13	18	15,15	15	12	E1b1a V38	100	E-U290
A_S281	Andes	15	13	21	31	16	10	28	11	11	11	14	19	21	21	12	11	13	13	18	16,18	15	11	E1b1a V38	100	E-U209*

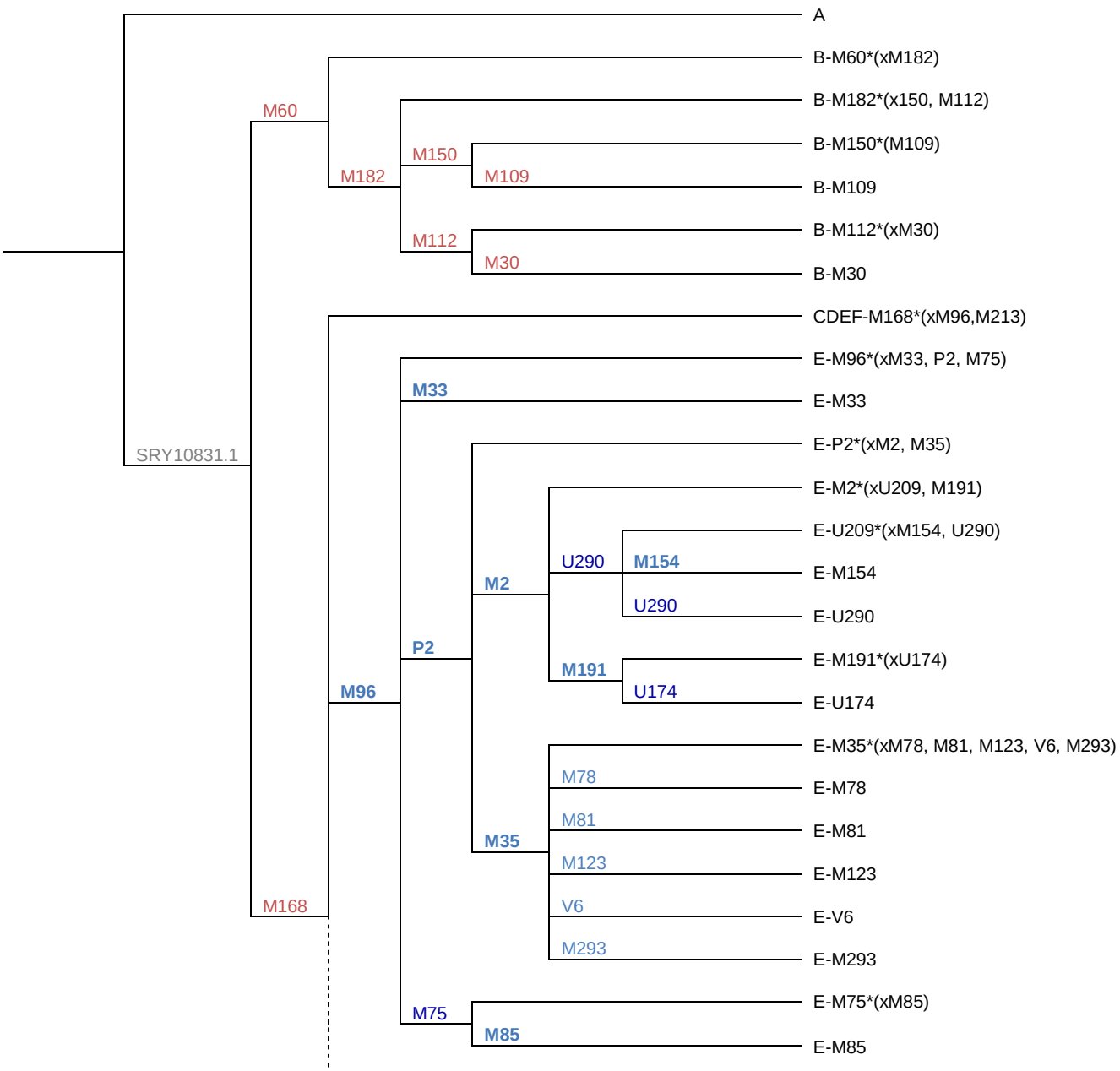
Table S1 (continued)

Sample[1]	Region of birth	Haplotype[2]																						Haplogroup Predictor (NEVGEN)		Haplogoup result
																								Haplogroup Assignment	Probability (%)	
A_S282	Andes	17	14	21	31	17	10	27	11	11	11	14	16	20	21	12	11	12	15	16	18,19	15	10	E1b1a V38	99.69	E-U174
A_S283	Andes	14	13	21	30	15	10	28	11	12	10	14	19	21	22	12	11	14	13	18	15,16	15	12	E1b1a V38	99.69	E-U290
A_S284	Andes	19	13	19	29	14	10	22	12	12	12	15	17	23	25	12	11	10	12	16	11,14	15	12	R1b	100	R-L23*
A_S285	Andes	19	14	23	32	16	10	25	11	11	10	14	18	17	23	12	11	14	13	18	11,12	13	11	B2a1 M218	100	B-M150
A_S286	Andes	18	13	21	29	15	10	24	12	11	10	14	18	21	23	13	11	9	12	17 .2	13,15	15	11	J1a2a1a2 P58	99.94	J-12f2a*
A_S287	Costa	19	13	20	29	14	12	22	13	12	12	15	16	23	25	13	13	10	12	18	11,14	17	12	R1b	100	R-S116*
A_S288	Costa	20	13	19	29	14	11	23	12	12	12	15	17	24	24	11	13	10	13	17	11,14	16	8	R1b	99.99	R-M529
A_S289	Costa	16	13	21	29	14	10	21	12	11	9	15	19	21	23	12	11	10	12	16	14,16	15	12	J2a1 PF7431	97.76	J-M172
A_S290	Costa	15	13	20	30	15	10	24	11	10	11	14	19	22	21	12	11	15	14	17	16,16	15	11	E1b1a V38	100	E-M2*
A_S291	Andes	19	14	18	32	13	10	20	13	12	11	14	20	23	25	13	14	11	13	14	14,16	15	13	Q M346>> M3	0.39	Q-M3*
A_S292	Andes	16	12	20	30	14	10	22	13	11	9	16	17	23	23	11	11	10	12	17	13,18	15	11	J2a1 M67>> S25258	21.36	J-M172
A_S293	Costa	18	13	21	31	15	10	22	11	12	11	14	16	21	21	12	11	13	15	16	17,18	17	10	E1b1a V38	99.59	E-U174
A_S294	Costa	18	14	20	30	14	10	24	12	12	10	16	15	24	23	11	15	12	12	15	17,18	15	10	L1b M317> M349	100	KL-M9*
A_S295	Costa	16	13	21	30	17	11	25	11	11	11	14	17	22	21	12	11	15	14	18	18,18	15	11	E1b1a V38	99.69	E-U174
A_S296	Costa	16	14	21	31	16	10	25	11	11	11	14	15	21	21	12	11	13	14	16	17,17	17	11	E1b1a V38	100	E-U174
A_S297	Costa	18	13	20	29	14	10	26	13	11	10	14	18	20	23	12	11	9	12	17 .2	14,16	15	11	J1a2a1a2 P58	90.39	J-12f2a*
A_S298	Costa	17	13	21	31	15	10	25	11	11	12	14	18	21	21	11	11	15	15	17	17,19	15	11	E1b1a V38	100	E-U174
A_S299	Costa	17	13	20	30	13	10	23	12	12	10	14	19	21	24	13	11	13	12	15	16,18	17	12	E1b1b	100	E-M78

[1] M- Mestizo; A- Afro-Ecuadorian

[2] The Y-STRs are in the following order: DYS576,DYS389 I,DYS448,DYS389 II,DYS19,DYS391,DYS481,DYS549,DYS533,DYS438,DYS437,DYS570,DYS635,DYS390, DYS439,DYS392,DYS643,DYS393,DYS458,DYS385,DYS456,Y GATA H4

[3] Individual that were bron in Playa de Oro



(Continues on the next page)

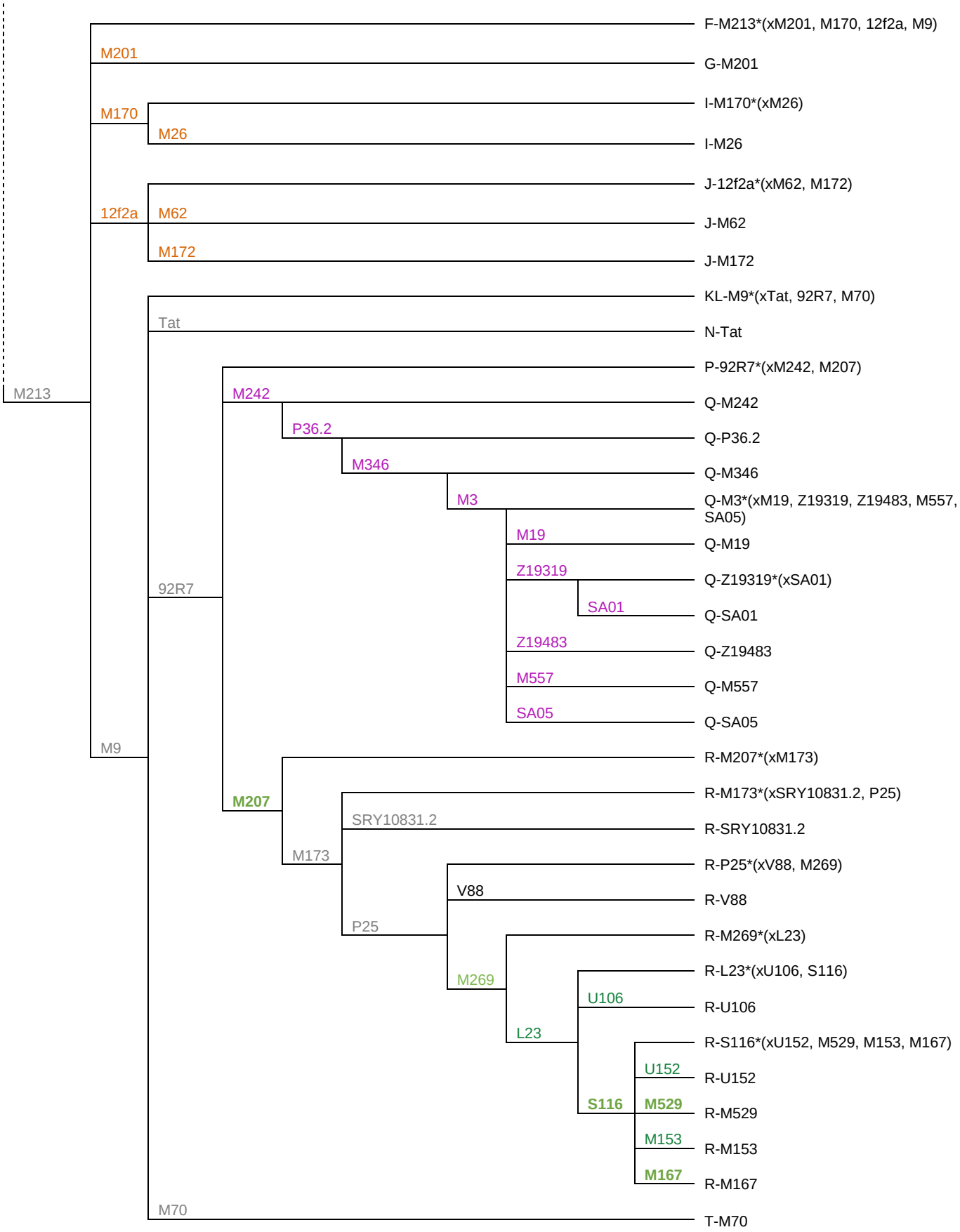


Figure S1: Phylogenetic tree of the Y chromosome haplogroups studied with the biallelic markers displayed in each branch. The colours correspond to the different multiplexes used (grey - Mx1; red - MxB; blue - MxE1 and E2; orange - Mx2; purple - MxQ; green - MxR1 and R2). The Y-SNPs that are analysed in more than one multiplex are highlighted in bold. The Y-SNPs only belonging to MxE1 are in a darker blue than the ones from MxE2. The Y-SNPs only belonging to MxR1 are in a darker green than the ones from MxR2. The marker genotyped in a singleplex reaction is in black.

Table S2: Primers used for PCR of the multiplex E1 and Q.

Y-SNP	Primer forward (5' to 3')	Primer reveres (5' to 3')	Size (bp)	Mutation	
M96	(31)	(31)	88	G/C	Multiplex E1
P2	GCTCCAGCCATCTTTTCCTTA	CTTCTCTCATGAGGGTTTTGGA	180	G/A	
M33	CACAACTTCATTGGCTACGG	GTTGAAGCCCCCAAGAGAGAC	190	A/C	
M75	TCCACACATCAAGAAAACTTGC	TTGAACAGAGGCATTTGTGA	224	G/A	
M85	TGGCATCCAATACTAGCTGATAAAC	AATGCTCACGCTTGTGTTCT	283	C/A	
M2	AAGTCCAGACCCAGGAAGGT	ACAGCTCCCCCTTTATCCTC	162	A/G	
M35	(31)	(31)	198	G/C	
U209	CCACAGGAATGCAAAAGATGTAAT	TGTGATGAGTGTCTGCCCAT	248	C/T	
U290	CCTGGAAAGCCACTAGCAAC	GTGCAGACAAAAGCGTACCA	135	T/A	
M154	TACTCACACAAACCAAGAAGAAACA	AACCATTGTGTTACATGGCCTA	130	T/C	
M191	AAAAATGGAGTGTTTATCAGAGCTT	CCCAGACACACCAAAATATCTC	122	T/G	
U174	TCCCTGCAGTGAAATAGTTTTG	AAATGGGAGTGTGGACTTGC	150	G/A	
M242	TTGTGCAAAAAGGTGACCAA	TTTCGCTTTAAGGGCTTTCA	155	C/T	Multiplex Q
P36.2	GAGGAGGGGGAGAGAGAAAA	TTCAAACAGCCCACCAGATA	299	G/T	
M346	GGCCTGAAAATGTGGAAAGA	AGCCTGGGAGACAGACACAG	247	G/C	
M3	CATTAAAGCCGGTCACAGGT	CTGCCAGGGCTTTCAAATAG	304	G/A	
M19	TCACCAGAGTTTGTGGTTGC	ACAGACACAAAGGGCCAACT	277	A/T	
Z19319	TTTGCTGAAGTTGCCTGTCA	AGTTCCAGTCAGGGCAATCA	163	C/T	
SA01	AAGATCCCACCACTGCACTC	CTCTGGCCCCTAACAAACCT	370	C/T	
Z19483	CCATGTAGGAGGAGGCAAAA	CATCACAAAAGCCAAAAGCA	253	A/G	
M557	AAGATCCCACCACTGCACTC	CTCTGGCCCCTAACAAACCT	370	C/A	
SA05	GAACCAAAGCACAGCACTCA	ATGCTCATGGCCTACACCTC	293	A/G	

Table S3: Primers used for SNaPshot reaction of multiplexes E1 and Q. The primers of multiplex (Mx) E2 whose sequence were changed are also presented. Before the sequence of the primer is indicated if the primer is forward (For) or reveres (Rev).

Y-SNP		Primer (5' to 3')	Size (bp)[1]	[µM][2]	
M96	For	(31)	40	0.22	Multiplex E1
P2	For	(76)	17	0.76	
M33	Rev	(76)	55	0.88	
M75	For	gtcgtgaaagtctgacaaACAATTATCAAACCACATCC	39	2.00	
M85	Rev	(76)	31	0.24	
M2	For	(76)	46	0.60	
M35[3]	Rev	aatgactaaactaggtgccacgtcgtgaaagtctgacaaTTTCGGAGT CTCTGCCTGTGTC	62	0.60	
U209	Rev	cgtcgtgaaagtctgacaaAAGACTGCAAGTTAAAATCA	43	1.68	
U290	For	actaggtgccacgtcgtgaaagtctgacaaTGTGGGAATTGATGG CGT	49	0.36	
M154	Rev	(76)	27	0.40	
M191	For	(76)	52	1.40	
U174	For	TgactaaactaggtgccacgtcgtgaaagtctgacaaTGCATACCAG ATTAACCCAT	58	1.14	
M81	For	gacaaTAAATTTTGTCCTTTTTTGAA	27	1.00	MxE2
M78	For	gaaagtctgacaaACACTTAACAAAGATACTTCTTTC	38	1.00	
M242	For	aaAAAAAGGTGACCAAGGTGCT	23	0.15	Multiplex Q
P36.2	Rev	gtctgacaaCATCTATCTATCCATTATTCTCTCT	35	0.45	
M346	For	ctgacaaCAGCCAAGAGGACAGTAAGA	28	0.15	
M3	Rev	TCACCTCTGGGACTGA	17	1.90	
M19	Rev	tgacaaGTAGAGACATCTGAAACCCAC	28	0.05	
Z19319	Rev	tcgtgaaagtctgacaa CCATCATCTCAACCTAAAATCC	40	0.50	
SA01	For	gtcgtgaaagtctgacaaTTTGTCAGTGTAGCAGTGG	38	4.00	
Z19483	Rev	acgtcgtgaaagtctgacaaATAAGCTGTCTGGCTATTTCT	42	2.50	
M557	Rev	tgccacgtcgtgaaagtctgacaaGAACAGGGTTGCAAACGGTA	45	0.64	
SA05	Rev	aggtgccacgtcgtgaaagtctgacaaTGTTTCTAGGGTGAGCCT GT	48	1.12	

[1] Final size with the base incorporated.
[2] The concentration in the SNaPshot reaction.
[3] In the multiplex E2 was used the same primer for this Y-SNP but with a concentration of 0.28µM.

Table S4: The concentrations in the SNaPshot reaction of the primers from multiplexes 1, 2, B, E2 (except the Y-SNPs mentioned in table S3), R1 and R2 that were optimised from the conditions previously described by Brion et al. (31), Gomes et al. (76) and Resque et al. (144).

Y-SNP	[µM]	Multiplex
SRY1532	0.075	Multiplex 1 (31)
M70	0.300	
M213	0.560	
M9	0.250	
M170	0.400	Multiplex 2 (31)
M201	0.100	
M112	0.100	Multiplex B (76)
M109	0.300	
M60	0.480	
M182	0.480	
M168	0.380	
M30	0.600	
M150	0.400	
P2	0.360	Multiplex E2 (76)
M154	0.120	
M293	0.160	
M85	0.160	
M96	0.120	
V6	0.400	
M191	0.500	
M33	0.440	
M123	0.500	
M2	0.340	
M207[1,2]	0.200	Multiplexes R1[1] and R2[2] (144)
M269[2]	0.200	
L23[1]	0.410	
U106[1]	0.040	
S116[1,2]	0.600	
U152[1]	0.200	
M529[1,2]	0.100	
M153[1]	0.200	
M167[1,2]	0.680	

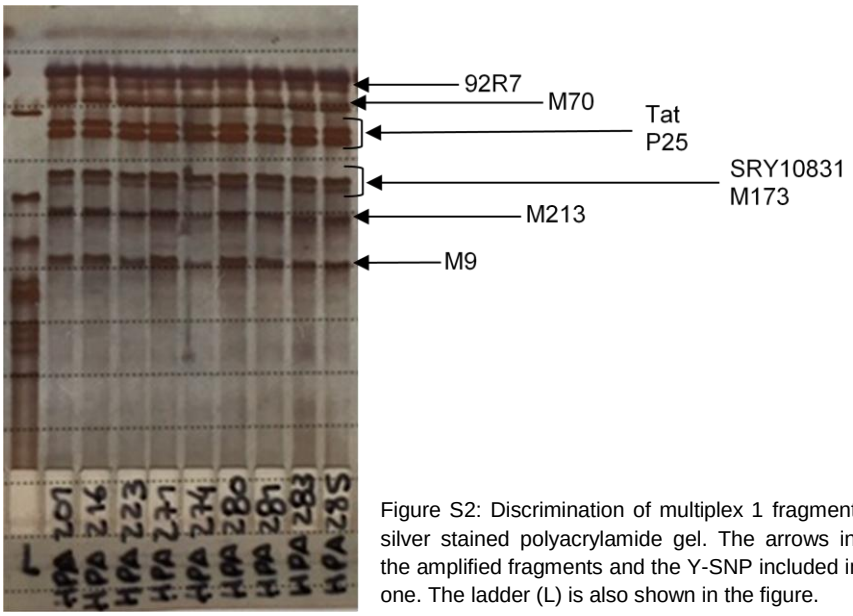


Figure S2: Discrimination of multiplex 1 fragments in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one. The ladder (L) is also shown in the figure.

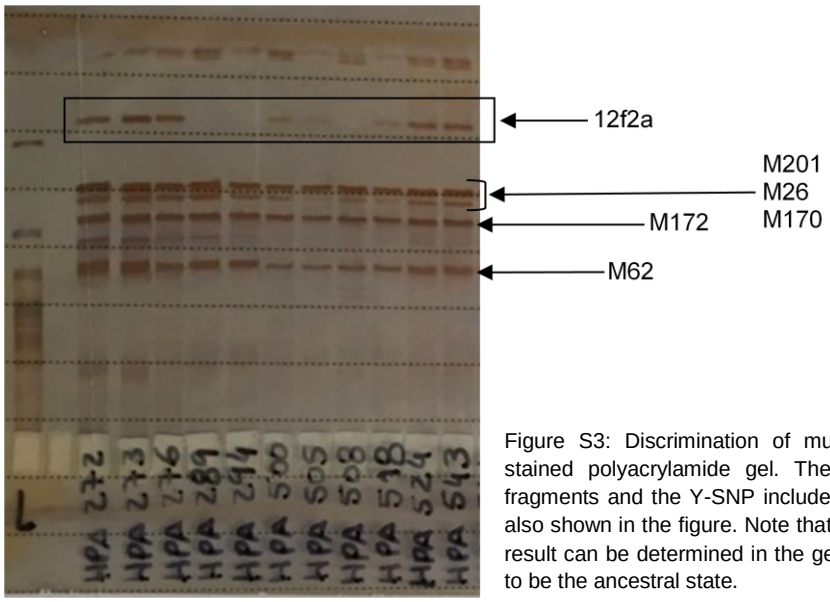


Figure S3: Discrimination of multiplex 2 fragments in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one. The ladder (L) is also shown in the figure. Note that the 12f2a is an indel and so the result can be determined in the gel, being the insertion considered to be the ancestral state.

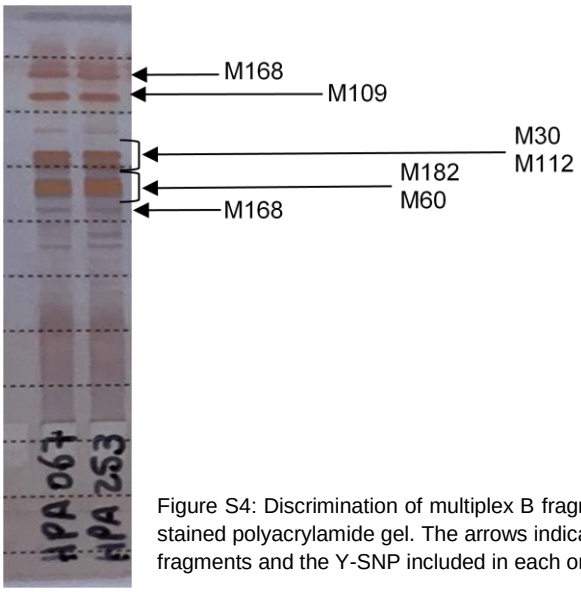


Figure S4: Discrimination of multiplex B fragments in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one.

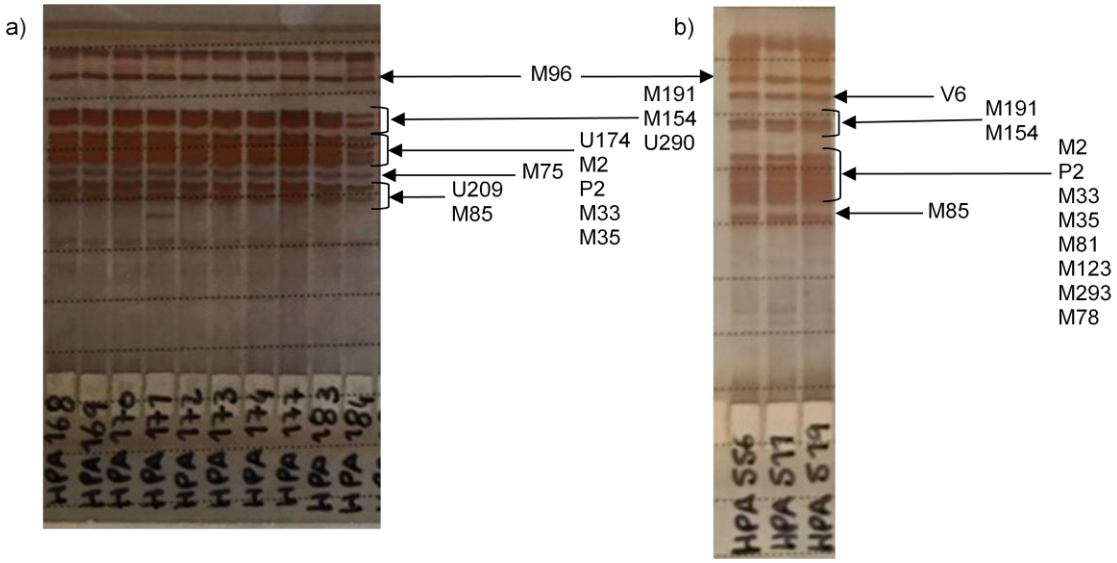


Figure S5: Discrimination of the fragments of multiplexes E1 (a) and E2 (b) in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one.

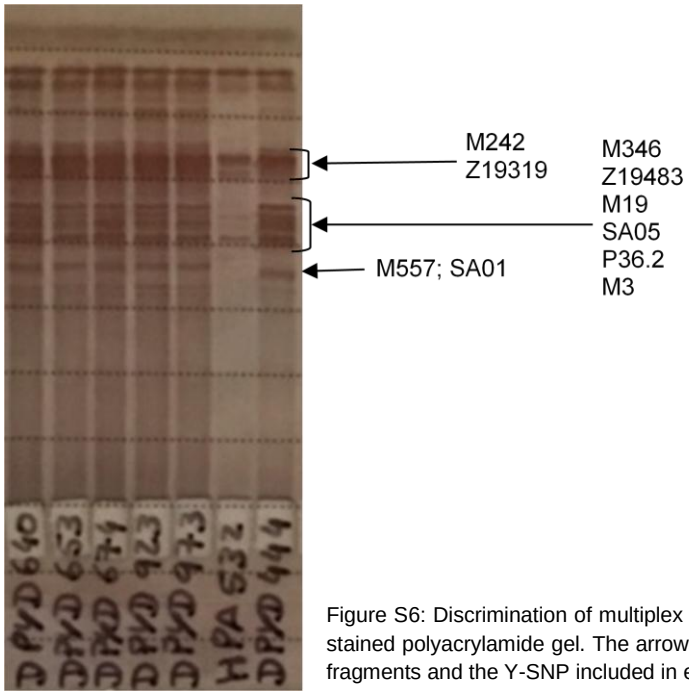


Figure S6: Discrimination of multiplex Q fragments in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one.

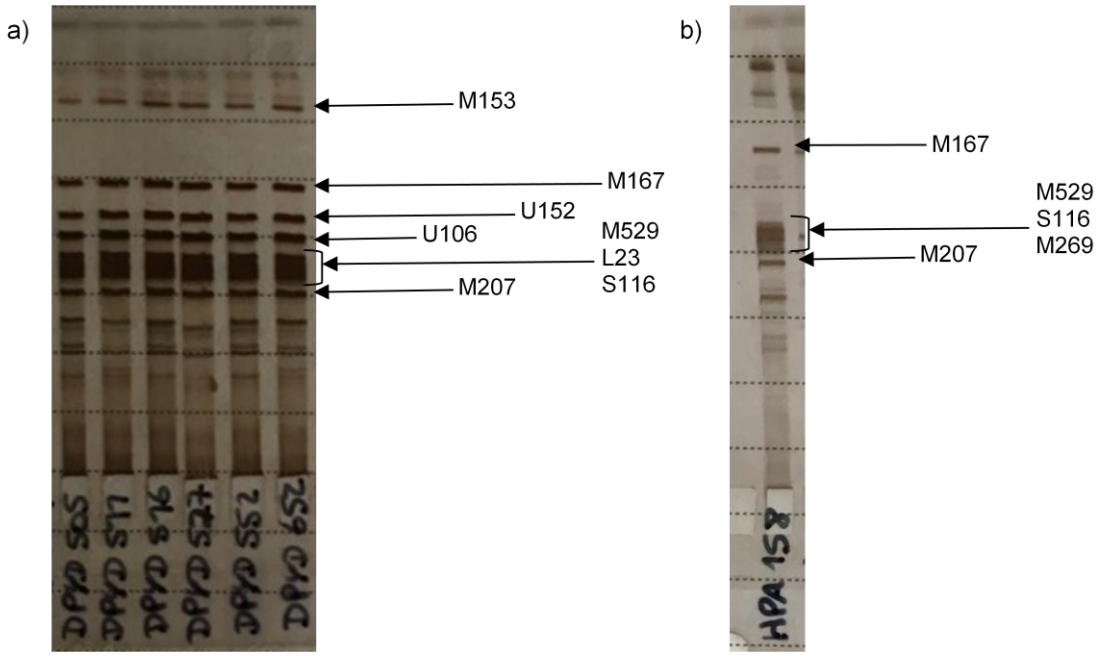


Figure S7: Discrimination of the fragments of multiplexes R1 (a) and R2 (b) in a silver stained polyacrylamide gel. The arrows indicate the amplified fragments and the Y-SNP included in each one.

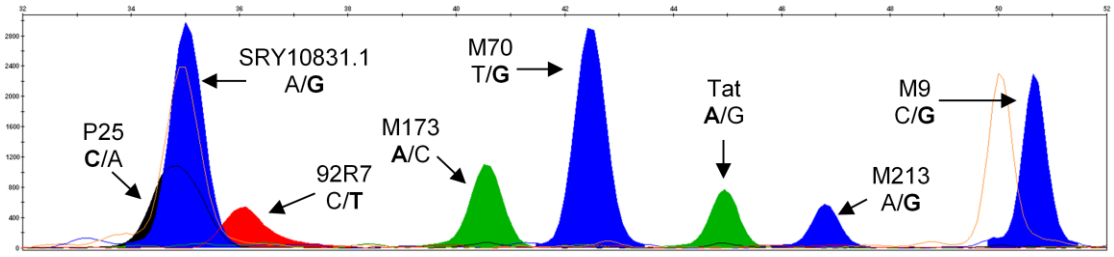


Figure S8: Electropherogram of the SBE reaction of a sample belonging to haplogroup T-M70. The Y-SNP and the two possible alleles detected with multiplex 1 are indicated with arrows. The alleles present in this sample are highlighted in bold.

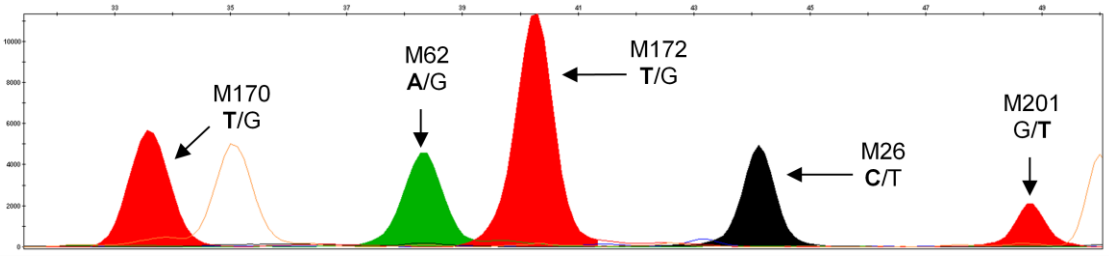


Figure S9: Electropherogram of the SBE reaction of a sample belonging to haplogroup G-M201. The Y-SNP and the two possible alleles detected with multiplex 2 are indicated with arrows. The alleles present in this sample are highlighted in bold.

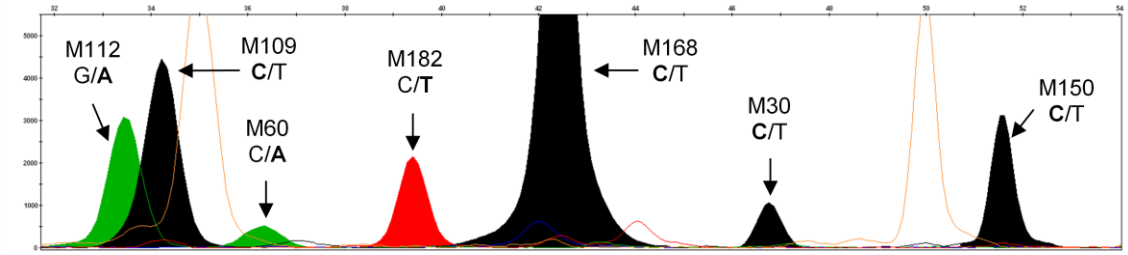


Figure S10: Electropherogram of the SBE reaction of a sample belonging to haplogroup B-M112*(xM30). The Y-SNP and the two possible alleles detected with multiplex B are indicated with arrows. The alleles present in this sample are highlighted in bold.

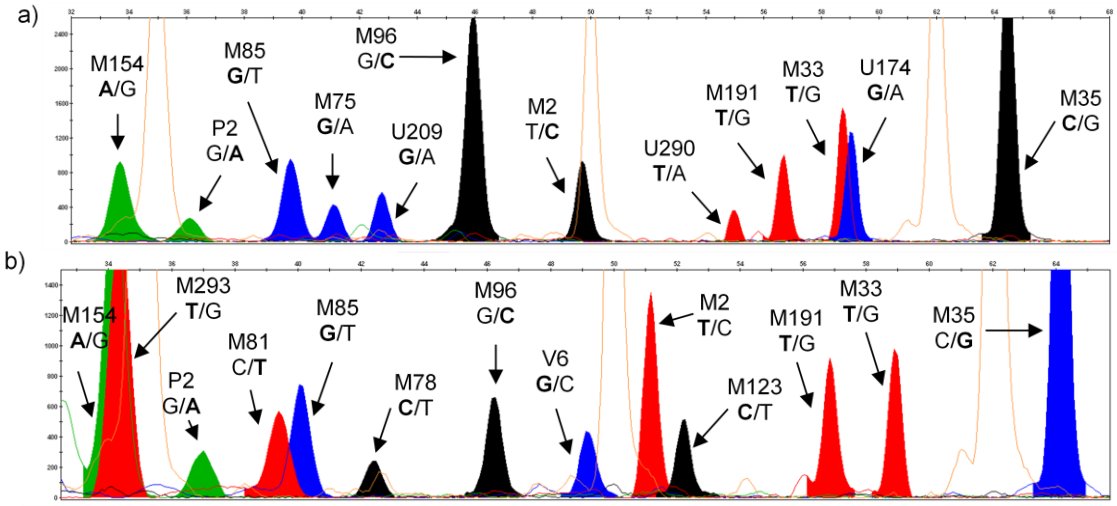


Figure S11: Electropherogram of the SBE reaction of two samples belonging to haplogroups E-M2*(xU209, M191) (a) and E-M81 (b). The Y-SNP and the two possible alleles detected with multiplexes E1 (a) and E2 (b) are indicated with arrows. The alleles present in these samples are highlighted in bold.

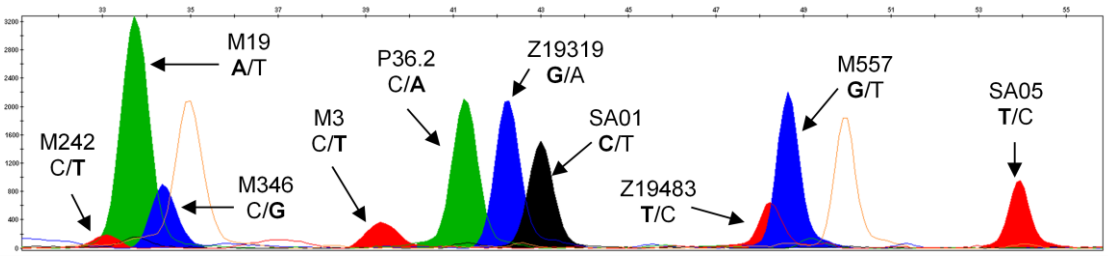


Figure S12: Electropherogram of the SBE reaction of a sample belonging to haplogroup Q-M3*(xM19, Z19319, Z19483, M557, SA05). The Y-SNP and the two possible alleles detected with multiplex Q are indicated with arrows. The alleles present in this sample are highlighted in bold.

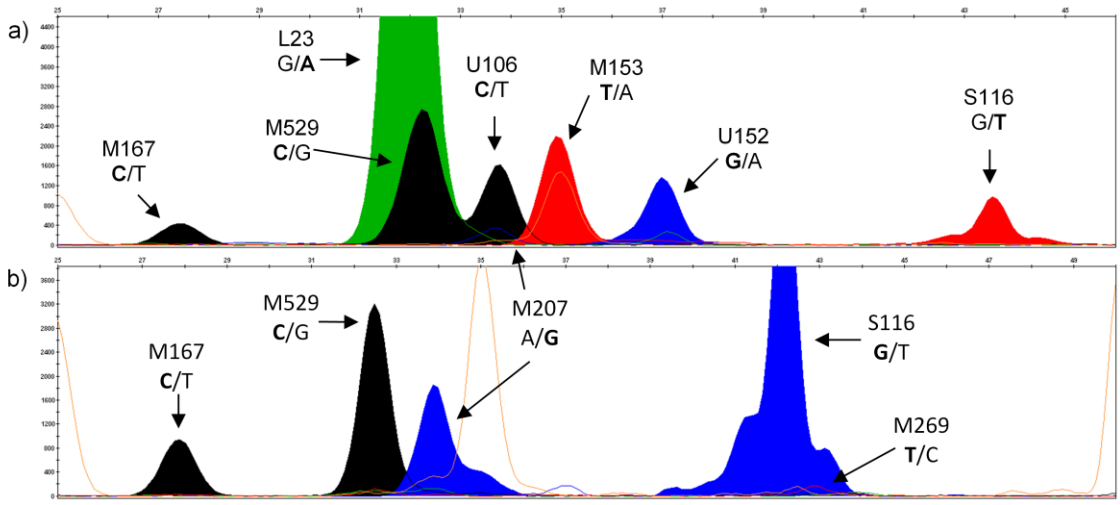


Figure S13: Electropherogram of the SBE reaction of two samples belonging to haplogroups R-S116*(xU152, M529, M153, M167) (a) and R-V88 (result known after a sequencing reaction for the V88 Y-SNP) (b). The Y-SNP and the two possible alleles detected with multiplexes R1 (a) and R2 (b) are indicated with arrows. The alleles present in these samples are highlighted in bold.

Table S5: Populations codes of the Latin American populations used in the different analyses performed.

Population code	Country	Population group/ ancestry	Data used	Reference
MEST	Ecuador	Mestizo	Y-STRs; Y-SNPs	Present study
AFRO	Ecuador	Afro-descendant	Y-STRs; Y-SNPs	Present study
ECU	Ecuador	Mestizo, Afro-descendant	Y-STRs; Y-SNPs	Present study
MEST1	Ecuador	Mestizo	Y-STRs	81
AFRO1	Ecuador	Afro-descendant	Y-STRs	81
NAM1	Ecuador	Native American	Y-STRs	81
NAM2	Ecuador	Native American	Y-STRs; Y-SNPs	74
MEST2	Ecuador	Mestizo	Y-STRs; Y-SNPs	13
NAM3	Ecuador	Native American	Y-STRs; Y-SNPs	13
NAM4	Ecuador	Native American	Y-STRs; Y-SNPs	151
COL	Colombia	Admixed American	Y-STRs	118
COL_P1	Colombia	Afro-descendant (Palenque)	Y-STRs; Y-SNPs	Data submitted for publication
COL-TH	Colombia (Tolima and Huila)	Admixed American	Y-SNPs	5
COL-BO	Colombia (Bolivar)	Admixed American	Y-SNPs	123
COL_P2	Colombia	Afro-descendant (Palenque)	Y-SNPs	123
COL_NA1	Colombia	Native American	Y-SNPs	189
COL_NA2	Colombia	Native American	Y-STRs; Y-SNPs	151
PER	Peru	Admixed American	Y-STRs	134
PER_NA1	Peru	Native American	Y-STRs; Y-SNPs	18
PER_NA2	Peru	Native American	Y-STRs; Y-SNPs	151
BRA	Brazil	Admixed American	Y-STRs	134
BRA_NA1	Brazil	Native American	Y-STRs	134
BRA-RJ	Brazil (Rio de Janeiro)	Admixed American	Y-SNPs	126
BRA_NA2	Brazil	Native American	Y-STRs; Y-SNPs	151
BOL1	Bolivia	Mestizo	Y-STRs	134
BOL2	Bolivia	Not specified	Y-SNPs	37
BOL_NA1	Bolivia	Native American	Y-STRs	134
BOL_NA2	Bolivia	Native American	Y-STRs; Y-SNPs	151
BOL_NA3	Bolivia	Native American	Y-STRs	72
CHL-I	Chile (Iquique)	Not specified (Urban populations)	Y-STRs	175
CHL-C	Chile (Concepsón)	Not specified (Urban populations)	Y-STRs	175
CHL-STP	Chile (Santiago, Temuco and Punta Arenas)	Not specified (Urban populations)	Y-STRs	175
ARG	Argentina	European	Y-STRs	134
ARG_NA1	Argentina	Native American	Y-SNPs	24
ARG_NA2	Argentina	Native American	Y-STRs; Y-SNPs	151
PAN	Panama	Admixed American	Y-STRs	134
CRI	Costa Rica	Mestizo	Y-STRs	134
NIC	Nicaragua	Mestizo	Y-STRs; Y-SNPs	125
MEX	Mexico	Mestizo	Y-STRs	111

Table S6: Populations codes of the African populations used in the different analyses performed.

Population code	Country	Data used	Reference
GNB	Guinea-Bissau	Y-STRs; Y-SNP	152
CIV	Ivory Coast	Y-STRs; Y-SNP	66
BEN1	Benin	Y-STRs; Y-SNP	66
BEN2	Benin	Y-STRs; Y-SNP	107
CMR	Cameroon	Y-STRs; Y-SNP	23
GNQ	Equatorial Guinea	Y-STRs; Y-SNP	83
GAB	Gabon	Y-STRs; Y-SNP	23
AGO	Angola	Y-STRs	44
MOZ	Mozambique	Y-STRs; Y-SNP	155

Table S7: Populations codes of the European and Middle East populations used in the different analyses performed.

Population code	Country	Data used	Reference
PRT-TM	Portugal (Trás-os-Montes)	Y-STRs	124
PRT1	Portugal	Y-SNPs	22
PRT-AL	Portugal (Alentejo)	Y-SNPs	22
PRT2	Portugal	Y-SNPs	4
PRT3	Portugal	Y-SNPs	78
PRT_N	Portugal (North)	Y-SNPs	65
PRT4	Portugal	Y-SNPs	6
PRT5	Portugal	Y-SNPs	122
PRT6	Portugal	Y-SNPs	179
ESP-ZA	Spain (Zamora)	Y-STRs	6
ESP-VM	Spain (Valencia and Murcia)	Y-STRs	168
ESP-CT	Spain (Catalunia)	Y-STRs	168
ESP_ISL	Spain (Menorca, Malloraca and Eivissa)	Y-STRs	168
ESP-AN1	Spain (Andalusia)	Y-STRs; Y-SNPs	145
ESP-AN2	Spain (Andalusia)	Y-SNPs	90
ESP1	Spain	Y-SNPs	4
ESP-BC1	Spain (Basque Country)	Y-SNPs	4
ESP-GA	Spain (Galicia)	Y-SNPs	31
ESP2	Spain	Y-SNPs	65
ESP-SL	Spain (Santander and Leon)	Y-SNPs	122
ESP-V	Spain (Valencia)	Y-SNPs	122
ESP-SEV	Spain (Sevilla)	Y-SNPs	122
ESP-BC2	Spain (Basque Country)	Y-SNPs	179
ESP3	Spain	Y-SNPs	179
GBR	United Kingdom	Y-SNPs	113
IRL1	Ireland	Y-SNPs	122
IRL2	Ireland	Y-SNPs	179
SCT	Scotland	Y-SNPs	33
ENG	England	Y-SNPs	122
FRA-COR	France (Corsica)	Y-STRs; Y-SNPs	54
FRA-B	France (Brest)	Y-SNPs	179
FRA	France	Y-SNPs	122
ITA_NW	Italy (Northwest)	Y-STRs; Y-SNPs	25
ITA_SE	Italy (Southeast)	Y-STRs; Y-SNPs	25
ITA-SAR	Italy (Sardinia)	Y-STRs; Y-SNPs	25
ITA	Italy	Y-SNPs	122
NLD	Netherlands	Y-SNPs	122
DEU	Germany	Y-SNPs	122
DEU-MB	Germany (Mecklenburg and Bavaria)	Y-STRs; Y-SNPs	142
POL-KAS	Poland (Kaszuby)	Y-STRs; Y-SNPs	142
POL	Poland (Kociewie and Kurpie)	Y-STRs; Y-SNPs	142

Table S7 (continued)

Population code	Country	Data used	Reference
BGR	Bulgaria	Y-STRs; Y-SNPs	100
TUR_NE	Turkey (Northeast)	Y-SNPs	42
TUR_NW	Turkey (Northwest)	Y-SNPs	42
TUR_C	Turkey (Centre)	Y-SNPs	42
TUR1	Turkey (West, South and East)	Y-SNPs	42
TUR2	Turkey	Y-SNPs	122

Table S8: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other three studies about Ecuadorian population. The non-significant values are highlighted in bold ($P \geq 0.05$). Population codes can be seen in Table S5.

	MEST (n=149)	AFRO (n=150)	MEST1 (n=102)	AFRO1 (n=94)	NAM1 (n=137)	NAM2 (n=49)	MEST2 (n=28)	NAM3 (n=27)
MEST (PS)	–	0.00000	**0.17068**	0.00000	0.00000	0.00000	**0.32927**	0.00178
AFRO (PS)	0.28413	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
MEST1	**0.00384**	0.35291	–	0.00000	0.00000	0.00000	**0.62469**	0.00149
AFRO1	0.10674	0.06079	0.15549	–	0.00000	0.00000	0.00030	0.00000
NAM1	0.17197	0.47719	0.13309	0.29548	–	0.00000	0.00010	0.00000
NAM2	0.15749	0.50474	0.12004	0.30311	0.08175	–	0.00010	0.00000
MEST2	**0.00215**	0.35620	**-0.00592**	0.13967	0.11181	0.11203	–	0.04099
NAM3	0.06474	0.38364	0.06759	0.17485	0.12102	0.24811	0.04911	–

Table S9: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Tables S5-S7.

	MEST (n=149)	AFRO (n=150)	COL (n=63)	COL_P1 (n=92)	PER (n=83)	PER_NA1 (n=167)	BRA (n=243)	BRA_NA1 (n=61)	BOL1 (n=44)	BOL_NA1 (n=56)	CHL-I (n=196)	CHL-C (n=198)	CHL-STP (n=584)
MEST (PS)	–	0.00000	0.03564	0.00000	0.00228	0.00000	0.01267	0.00000	0.00000	0.00000	0.01030	0.01634	0.00149
AFRO (PS)	0.22705	–	0.00000	0.00059	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
COL	0.01203	0.29564	–	0.00000	**0.05980**	0.00000	0.00000	0.00109	0.00109	0.00000	0.00089	0.00040	0.00030
COL_P1	0.15235	0.03269	0.20568	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PER	0.02110	0.29766	**0.01174**	0.20503	–	0.00000	0.00000	0.00000	0.00525	0.00000	0.00000	0.00000	0.00000
PER_NA1	0.44624	0.55317	0.44065	0.50603	0.44599	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
BRA	0.00972	0.18464	0.04954	0.12410	0.06818	0.46605	–	0.00000	0.00000	0.00000	0.00069	**0.05732**	0.00109
BRA_NA1	0.08259	0.36426	0.04210	0.27177	0.07543	0.39512	0.12630	–	0.00000	0.00000	0.00000	0.00000	0.00000
BOL1	0.08320	0.40001	0.04491	0.30645	0.03064	0.42908	0.14556	0.06664	–	**0.05673**	0.00000	0.00000	0.00000
BOL_NA1	0.15617	0.44176	0.10369	0.35258	0.08235	0.44021	0.21992	0.09154	**0.01742**	–	0.00000	0.00000	0.00000
CHL-I	0.01115	0.29142	0.03015	0.21567	0.04814	0.46429	0.01948	0.10065	0.09884	0.18734	–	0.01683	**0.26403**
CHL-C	0.00933	0.23976	0.03572	0.16794	0.06583	0.48621	**0.00471**	0.12513	0.14030	0.22519	0.00840	–	**0.20414**
CHL-STP	0.01213	0.26140	0.03974	0.19632	0.06185	0.47378	0.00986	0.11658	0.12799	0.21015	**0.00069**	**0.00108**	–
ARG	**0.00129**	0.12477	**0.01284**	0.08426	0.02411	0.31032	**0.00010**	0.05177	0.05460	0.09650	0.00413	**0.00008**	0.00313
PAN	0.02188	0.16841	0.04696	0.10912	0.06317	0.45112	0.02038	0.13227	0.11973	0.19120	0.04691	0.02558	0.04347
CRI	0.00681	0.22003	0.04824	0.14845	0.04909	0.46971	**0.00488**	0.11646	0.11971	0.19632	0.01144	0.01279	0.00882
NIC	**0.00306**	0.21487	0.02780	0.14493	0.04608	0.45689	**0.00201**	0.09619	0.11089	0.18206	0.00861	**0.00325**	0.00501
MEX	0.04564	0.33816	0.00951	0.26477	0.01874	0.40503	0.09656	0.02165	0.01460	0.04903	0.05806	0.08318	0.07920
ESP-AN1	0.01917	0.26120	0.05407	0.19098	0.07853	0.46922	0.01088	0.12678	0.13470	0.22864	**0.00199**	**0.00458**	**-0.00059**
GNQ	0.92856	0.93320	0.93374	0.92176	0.93427	0.93787	0.92699	0.93095	0.93176	0.93545	0.93418	0.93330	0.93270

Table S9 (continued)

	ARG (n=364)	PAN (n=100)	CRI (n=165)	NIC (n=165)	MEX (n=307)	ESP-AN1 (n=144)	GNQ (n=112)
MEST (PS)	**0.28175**	0.00188	0.04386	**0.14474**	0.00000	0.00198	0.00000
AFRO (PS)	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
COL	**0.14999**	0.00000	0.00020	0.00198	0.03980	0.00000	0.00000
COL_P1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PER	0.00040	0.00000	0.00000	0.00000	0.00079	0.00000	0.00000
PER_NA1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
BRA	**0.41550**	0.00129	**0.06455**	**0.17602**	0.00000	0.01168	0.00000
BRA_NA1	0.00000	0.00000	0.00000	0.00000	0.00188	0.00000	0.00000
BOL1	0.01218	0.00000	0.00000	0.00000	0.03010	0.00000	0.00000
BOL_NA1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CHL-I	0.01990	0.00000	0.00851	0.01673	0.00000	**0.19949**	0.00000
CHL-C	**0.42085**	0.00020	0.00505	**0.11781**	0.00000	**0.09266**	0.00000
CHL-STP	0.00812	0.00000	0.00594	0.03099	0.00000	**0.51084**	0.00000
ARG	–	0.03762	**0.24879**	**0.52559**	0.00000	**0.22255**	0.00000
PAN	0.01047	–	0.00069	0.00436	0.00000	0.00000	0.00000
CRI	**0.00153**	0.02933	–	**0.15345**	0.00000	0.01802	0.00000
NIC	**-0.00074**	0.01653	**0.00266**	–	0.00000	0.02277	0.00000
MEX	0.04658	0.09656	0.07861	0.06410	–	0.00000	0.00000
ESP-AN1	**0.00232**	0.04277	0.01087	0.00956	0.09124	–	0.00000
GNQ	0.86364	0.92020	0.92786	0.92870	0.92626	0.93056	–

Table S10: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Tables S5-S7.

	MEST (n=149)	AFRO (n=150)	MEST2 (n=28)	NAM3 (n=27)	NAM2 (n=49)	COL-TH (n=83)	COL-BO (n=149)	COL_P2 (n=26)	COL_P1 (n=95)	COL_NA1 (n=72)	PER_NA1 (n=163)	BRA-RJ (n=605)	BOL2 (n=226)	ARG_NA1 (n=134)	NIC (n=164)	ESP-GA (n=292)	GNQ (n=112)
MEST (PS)	–	0.00000	**0.16969**	0.00000	0.00000	0.00158	0.00099	0.00000	0.00000	0.00000	0.00000	0.00000	0.00079	0.03663	**0.06643**	0.00000	0.00000
AFRO (PS)	0.32919	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00762	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	**0.35185**
MEST2	**0.01025**	0.36690	–	0.00020	0.00000	0.00287	0.00455	0.00050	0.00000	0.00931	0.00010	0.00099	0.00277	**0.16523**	0.02505	0.00010	0.00000
NAM3	0.20265	0.50315	0.14532	–	0.00020	0.00000	0.00000	0.00000	0.00000	0.00248	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
NAM2	0.32957	0.63851	0.36559	0.24746	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00089	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
COL-TH	0.03282	0.43160	0.07522	0.33761	0.50926	–	0.00089	0.00000	0.00000	0.00000	0.00000	0.00188	0.03465	0.00000	**0.06861**	0.00545	0.00000
COL-BO	0.02316	0.26179	0.05380	0.27964	0.42788	0.03185	–	0.00683	0.00000	0.00000	0.00000	**0.12444**	0.00000	0.00000	0.03604	0.00000	0.00000
COL_P2	0.12226	0.08254	0.13869	0.35303	0.60301	0.19221	0.05892	–	**0.90813**	0.00000	0.00000	0.00040	0.00000	0.00000	0.00010	0.00000	0.01950
COL_P1	0.13291	0.08605	0.14582	0.31759	0.49891	0.19800	0.07309	**-0.01887**	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00020
COL_NA1	0.12220	0.44957	0.06311	0.09504	0.14571	0.24438	0.21171	0.28824	0.27994	–	0.01109	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PER_NA1	0.17320	0.49952	0.13842	0.14085	0.07352	0.31601	0.27534	0.37301	0.35366	0.02796	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
BRA-RJ	0.03057	0.27706	0.07316	0.30594	0.42406	0.02552	**0.00293**	0.09106	0.10461	0.23751	0.28460	–	0.00000	0.00000	0.00683	0.00000	0.00000
BOL2	0.02733	0.40916	0.07635	0.30486	0.41538	0.01428	0.04390	0.19859	0.20831	0.20959	0.24893	0.04441	–	0.00000	0.00238	0.00000	0.00000
ARG_NA1	0.01105	0.37643	**0.01169**	0.17675	0.27629	0.07144	0.06539	0.17908	0.18794	0.08582	0.11694	0.07855	0.04124	–	0.00020	0.00000	0.00000
NIC	**0.00680**	0.33491	0.03281	0.26043	0.39852	**0.00937**	0.00888	0.12314	0.13452	0.18184	0.24059	0.01013	0.01924	0.03393	–	0.00069	0.00000
ESP-GA	0.06097	0.40743	0.12267	0.37976	0.50385	0.02190	0.04545	0.20360	0.21307	0.30305	0.35215	0.02201	0.04933	0.10231	0.02369	–	0.00000
GNQ	0.31470	**-0.00022**	0.36253	0.51328	0.66064	0.41321	0.24136	0.06378	0.06956	0.45195	0.50400	0.26019	0.39172	0.36692	0.31829	0.39222	–

Table S11: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations. These comparisons were based on Y-STR data and only the Q-lineages were considered. Non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S5.

	ECU (n=37)	NAM2 (n=43)	NAM4 (n=70)	COL_NA2 (n=55)	COL-TH (n=14)	PER_NA2 (n=68)	PER_NA1 (n=111)	BRA_NA2 (n=329)	BRA-RJ (n=14)	BOL_NA2 (n=73)	BOL_NA3 (n=98)	BOL2 (n=53)	ARG_NA2 (n=227)
ECU (PS)	–	0.00020	0.00238	**0.18404**	0.00000	0.00307	0.00030	0.00020	**0.20345**	**0.15573**	0.00000	0.03059	0.00000
NAM2	0.11980	–	**0.54608**	0.00000	0.00000	0.00000	0.00000	0.00000	0.04712	0.00000	0.00000	0.00010	0.00000
NAM4	0.06862	**-0.00586**	–	0.00030	0.00000	0.00000	0.00000	0.00000	**0.14444**	0.00000	0.00000	0.00010	0.00000
COL_NA2	**0.00838**	0.09600	0.06054	–	0.00030	0.00010	0.00000	0.00040	**0.29928**	0.03020	0.00000	0.04693	0.00000
COL-TH	0.19716	0.23418	0.22577	0.19025	–	0.00000	0.00000	0.00000	0.00406	0.00000	0.00000	0.00000	0.00000
PER_NA2	0.04399	0.11942	0.09568	0.05962	0.21823	–	0.00000	0.00000	0.00554	0.00000	0.00010	**0.07564**	0.00000
PER_NA1	0.05401	0.13194	0.10700	0.06423	0.28978	0.06460	–	0.00000	0.02742	0.00000	0.00000	0.00000	0.00000
BRA_NA2	0.04969	0.12605	0.09719	0.03529	0.27571	0.09182	0.07447	–	0.01030	0.00030	0.00000	0.00000	0.00000
BRA-RJ	**0.01479**	0.06492	**0.02495**	**0.00747**	0.20052	0.08383	0.04731	0.06250	–	**0.12326**	0.00000	0.02703	0.00881
BOL_NA2	**0.00917**	0.09410	0.06003	0.01788	0.21255	0.07703	0.06157	0.02818	**0.02405**	–	0.00000	0.00050	0.00000
BOL_NA3	0.13042	0.22668	0.19544	0.13255	0.33353	0.05518	0.10624	0.20725	0.15955	0.17834	–	0.00010	0.00000
BOL2	0.02269	0.09345	0.06277	0.01720	0.18776	**0.01329**	0.07003	0.07181	0.04806	0.04597	0.06830	–	0.00000
ARG_NA2	0.06442	0.17104	0.12967	0.06245	0.29815	0.10100	0.03191	0.07143	0.06349	0.08115	0.16307	0.08292	–

Table S12: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations. These comparisons were based on Y-STR data and only the Q-lineages were considered. Non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S5.

	ECU (n=37)	NAM2 (n=43)	NAM4 (n=70)	COL_NA2 (n=55)	COL-TH (n=14)	PER_NA2 (n=68)	PER_NA1 (n=111)	BRA_NA2 (n=329)	BRA-RJ (n=14)	BOL_NA2 (n=73)	BOL_NA3 (n=98)	BOL2 (n=53)	ARG_NA2 (n=227)
ECU (PS)	–	0.00000	0.00000	0.00109	0.00109	0.00911	0.00366	0.00099	**0.99990**	0.00089	0.00347	**0.51520**	0.01010
NAM2	0.09684	–	**0.88466**	0.00000	0.00059	0.00000	0.00000	0.00000	0.00158	0.00000	0.00000	0.00000	0.00000
NAM4	0.07022	**-0.00956**	–	0.00000	0.00069	0.00000	0.00000	0.00000	0.00307	0.00000	0.00000	0.00000	0.00000
COL_NA2	0.01226	0.10642	0.08078	–	0.00139	0.00000	0.00000	0.00000	**0.05009**	0.00000	0.00000	0.00000	0.00000
COL-TH	0.02079	0.12583	0.09528	0.03338	–	0.00099	0.00188	0.00277	**0.10989**	0.00188	0.00257	0.00020	0.00178
PER_NA2	0.00735	0.10031	0.07524	0.01934	0.02812	–	0.00000	0.00000	**0.07336**	0.00000	0.03594	0.02396	0.00000
PER_NA1	0.01022	0.10081	0.07660	0.02208	0.03095	0.01622	–	0.00000	**0.06138**	0.00000	0.00000	0.00089	0.00000
BRA_NA2	0.01487	0.10223	0.07895	0.02655	0.03558	0.02178	0.02442	–	0.04336	0.00000	0.00000	0.00040	0.00000
BRA-RJ	**0.00000**	0.10531	0.07502	**0.01286**	**0.02198**	**0.00769**	**0.01070**	0.01553	–	0.03386	**0.05960**	**0.66716**	**0.07445**
BOL_NA2	0.01647	0.10900	0.08396	0.02833	0.03767	0.02343	0.02610	0.03047	0.01729	–	0.00000	0.00010	0.00000
BOL_NA3	0.01101	0.10240	0.07799	0.02315	0.03210	0.00472	0.01959	0.02547	**0.01182**	0.02718	–	0.00743	0.00000
BOL2	**0.00037**	0.09481	0.06922	0.01249	0.02094	0.00460	0.00879	0.01463	**0.00038**	0.01615	0.00751	–	0.00149
ARG_NA2	0.00698	0.09607	0.07263	0.01961	0.02828	0.01485	0.01730	0.02199	**0.00816**	0.02359	0.01860	0.00811	–

Table S13: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and other Latin American populations within the haplogroup Q. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S5.

	ECU (n=37)	NAM2 (n=43)	NAM4 (n=70)	COL_NA2 (n=55)	COL-TH (n=14)	PER_NA2 (n=68)	PER_NA1 (n=111)	BRA_NA2 (n=329)	BRA-RJ (n=14)	BOL_NA2 (n=73)	BOL2 (n=53)	ARG_NA2 (n=227)
ECU (PS)	–	0.04128	**0.22790**	**0.16988**	0.00604	0.04891	0.03475	**0.26393**	**0.32225**	0.03069	**0.08554**	0.00396
NAM2	0.09307	–	**0.28928**	0.00079	0.00000	**0.99990**	**0.58598**	**0.14781**	**0.99990**	0.00010	0.03297	**0.67736**
NAM4	**0.01386**	**0.01646**	–	0.00158	0.00010	**0.61746**	**0.38561**	**0.78319**	**0.99990**	0.00020	0.01366	**0.11821**
COL_NA2	**0.03124**	0.19962	0.14067	–	**0.09930**	0.00000	0.00000	0.00010	**0.05722**	**0.42273**	0.00178	0.00000
COL-TH	0.32576	0.65718	0.56558	**0.11395**	–	0.00000	0.00000	0.00000	0.00564	**0.21473**	0.00010	0.00000
PER_NA2	0.07220	**-0.00707**	**-0.00052**	0.20379	0.67291	–	**0.99990**	**0.21869**	**0.99990**	0.00000	0.00584	**0.59252**
PER_NA1	0.08180	**-0.00358**	**-0.00023**	0.23964	0.72024	**-0.01166**	–	**0.12029**	**0.99990**	0.00000	0.00228	**0.39541**
BRA_NA2	**0.01008**	**0.01903**	**-0.00731**	0.18255	0.61696	**0.00861**	**0.00942**	–	**0.62835**	0.00000	0.00515	0.00733
BRA-RJ	**0.03010**	**0.00000**	**-0.01691**	0.12566	0.46154	**-0.03550**	**-0.03098**	**-0.00729**	–	0.01762	**0.33630**	**0.99990**
BOL_NA2	0.07931	0.24147	0.19633	**-0.00545**	**0.04425**	0.25094	0.29085	0.26867	0.17258	–	0.00030	0.00000
BOL2	**0.03639**	0.08437	0.05285	**0.11285**	0.41679	0.08296	0.09869	0.06615	**0.03207**	0.16110	–	0.00050
ARG_NA2	0.11274	**-0.00271**	**0.01159**	0.31546	0.76923	**-0.00372**	**-0.00045**	0.02055	**-0.02880**	0.37609	0.11831	–

Table S14: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and African populations. These comparisons were based on Y-STRs data and only the African lineages were considered. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S6.

	ECU (n=123)	GNB (n=157)	CIV (n=76)	BEN1 (n=170)	BEN2 (n=111)	CMR (n=54)	GNQ (n=94)	GAB (n=782)	AGO (n=230)	MOZ (n=76)
ECU (PS)	–	0.00030	0.00000	0.00248	**0.20681**	0.00050	**0.09276**	**0.70825**	0.00000	0.00000
GNB	0.02917	–	0.00000	0.00010	**0.07722**	0.00000	0.00000	0.00000	0.00000	0.00000
CIV	0.06885	0.06815	–	0.00158	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000
BEN1	0.01982	0.02347	0.02632	–	**0.17295**	0.00000	0.00010	0.00000	0.00000	0.00000
BEN2	**0.00319**	**0.00634**	0.05034	**0.00357**	–	0.00010	0.01218	0.03119	0.00000	0.00000
CMR	0.05959	0.05262	0.12571	0.08632	0.06590	–	0.00287	0.00030	0.00000	0.00000
GNQ	**0.00796**	0.04244	0.08606	0.03356	0.01784	0.04389	–	0.02049	0.00000	0.00000
GAB	**-0.00183**	0.02637	0.06081	0.02039	0.00723	0.04348	0.01003	–	0.00000	0.00000
AGO	0.04879	0.06197	0.10917	0.07776	0.05252	0.13144	0.09017	0.04141	–	0.00000
MOZ	0.10758	0.10192	0.16623	0.14055	0.10839	0.12335	0.12479	0.09377	0.08230	–

Table S15: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and African populations. These comparisons were based on Y-STRs data and only the African lineages were considered. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S6.

	ECU (n=123)	GNB (n=157)	CIV (n=76)	BEN1 (n=170)	BEN2 (n=111)	CMR (n=54)	GNQ (n=94)	GAB (n=782)	AGO (n=230)	MOZ (n=76)
ECU (PS)	–	0.00000	0.00000	0.00030	0.00000	0.00089	0.00218	0.00000	0.00000	0.00000
GNB	0.00908	–	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
CIV	0.01334	0.00463	–	0.00000	0.00000	0.00000	0.00050	0.00000	0.00000	0.00000
BEN1	0.00470	0.00242	0.00555	–	0.00832	0.00020	0.03029	0.00069	0.00000	0.00000
BEN2	0.01005	0.00347	0.00917	0.00231	–	0.00000	0.00020	0.00020	0.00000	0.00000
CMR	0.00991	0.00840	0.01267	0.00646	0.00990	–	**0.17830**	0.00218	0.00000	0.00000
GNQ	0.00553	0.00323	0.00582	0.00194	0.00497	**0.00179**	–	**0.07247**	0.00000	0.00000
GAB	0.00807	0.00362	0.00683	0.00204	0.00337	0.00496	**0.00115**	–	0.00000	0.00000
AGO	0.01751	0.01139	0.01669	0.01175	0.01129	0.01930	0.01410	0.00892	–	0.00000
MOZ	0.01278	0.00333	0.00965	0.00602	0.00680	0.01220	0.00737	0.00635	0.01008	–

Table S16: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and African populations. These comparisons were based on Y-SNPs data and only the African lineages were considered. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S6.

	ECU (n=122)	GNB (n=268)	CIV (n=76)	BEN1 (n=171)	BEN2 (n=117)	CMR (n=54)	GNQ (n=85)	GAB (n=781)	MOZ (n=75)
ECU (PS)	–	0.00000	**0.10088**	**0.09326**	**0.11672**	0.00000	**0.36600**	0.00465	0.00069
GNB	0.07000	–	0.00000	0.00020	0.00069	0.00000	0.00168	0.00000	0.00000
CIV	**0.01284**	0.09770	–	0.01921	0.04831	0.00000	**0.06386**	0.00030	0.00000
BEN1	**0.00789**	0.04042	0.02605	–	**0.60459**	0.00000	**0.86229**	0.00069	0.00040
BEN2	**0.00786**	0.04282	0.02273	**-0.00337**	–	0.00000	**0.90446**	0.00040	0.00000
CMR	0.26984	0.16658	0.34228	0.24994	0.27570	–	0.00000	0.00000	0.01931
GNQ	**-0.00025**	0.04392	**0.01818**	**-0.00648**	**-0.00788**	0.24781	–	0.00990	0.00089
GAB	0.02125	0.05482	0.05178	0.02208	0.03376	0.13995	0.02393	–	**0.13504**
MOZ	0.07545	0.05947	0.13996	0.06946	0.09108	0.05207	0.07277	**0.00660**	–

Table S17: Pairwise R_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and European populations. These comparisons were performed considering only the European lineages. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S7.

	ECU (n=139)	PRT-TM (n=30)	ESP-ZA (n=230)	ESP-AN1 (n=138)	ESP-CT (n=1413)	ESP-VM (n=330)	ESP_ISL (n=150)	FRA-COR (n=290)	ITA_NW (n=340)	ITA_SE (n=450)	ITA-SAR (n=80)	DEU-MB (n=346)	POL-KAS (n=202)	POL (n=315)	BGR (n=243)
ECU (PS)	–	**0.89912**	0.02247	0.02287	0.00228	0.00228	0.00099	0.00762	**0.10544**	0.00000	0.00000	0.00149	0.00000	0.00000	0.00000
PRT-TM	**-0.01133**	–	**0.83695**	**0.89516**	**0.93812**	**0.69161**	**0.41194**	**0.52173**	**0.90842**	0.00554	0.00010	**0.39768**	0.00000	0.00000	0.00000
ESP-ZA	0.00821	**-0.00882**	–	**0.54173**	**0.07465**	0.01653	0.00317	0.00000	0.00317	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000
ESP-AN1	0.01127	**-0.01121**	**-0.00091**	–	**0.72131**	**0.33818**	**0.21018**	0.00099	0.02445	0.00000	0.00000	0.00079	0.00000	0.00000	0.00000
ESP-CT	0.01150	**-0.01044**	**0.00217**	**-0.00143**	–	0.00564	0.01218	0.00000	0.00119	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP-VM	0.01628	**-0.00620**	0.00595	**0.00034**	0.00439	–	0.00842	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP_ISL	0.02412	**-0.00087**	0.01446	**0.00213**	0.00720	0.01019	–	0.00000	0.00238	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
FRA-COR	0.01134	**-0.00392**	0.02408	0.01991	0.02417	0.02642	0.03783	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ITA_NW	**0.00341**	**-0.01031**	0.00937	0.00793	0.00657	0.01362	0.01330	0.01797	–	0.00000	0.00000	0.00337	0.00000	0.00000	0.00000
ITA_SE	0.02933	0.03503	0.06364	0.07072	0.07622	0.08367	0.07953	0.04494	0.04149	–	0.00000	0.00000	0.00000	0.00000	0.00000
ITA-SAR	0.08562	0.09547	0.11563	0.12038	0.13852	0.14933	0.15215	0.07048	0.11464	0.05542	–	0.00000	0.00000	0.00000	0.00000
DEU-MB	0.01472	**-0.00003**	0.01599	0.01833	0.01997	0.03288	0.02496	0.02421	0.00776	0.03287	0.08512	–	0.00000	0.00000	0.00000
POL-KAS	0.16732	0.16774	0.16415	0.16898	0.17791	0.19559	0.16220	0.17559	0.15230	0.12962	0.17253	0.11686	–	**0.06415**	0.00000
POL	0.15492	0.15707	0.14988	0.15335	0.16154	0.17977	0.14738	0.16492	0.14232	0.12962	0.16067	0.11114	**0.00394**	–	0.00000
BGR	0.08513	0.08896	0.11440	0.11661	0.12310	0.13587	0.10628	0.10753	0.08452	0.05564	0.12112	0.07528	0.06309	0.05826	–

Table S18: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and European populations. These comparisons were performed considering only the European lineages. The non-significant values are highlighted in bold ($P \geq 0.05$). Population codes can be seen in Table S7.

	ECU (n=139)	PRT1 (n=582)	PRT-AL (n=64)	PRT2 (n=132)	PRT3 (n=531)	ESP-AN1 (n=138)	ESP-AN2 (n=408)	ESP1 (n=874)	ESP-BC1 (n=115)	GBR (n=19)	FRA-COR (n=291)
ECU (PS)	–	0.02257	**0.65162**	**0.13811**	**0.16068**	0.03515	0.02564	0.00099	0.00000	**0.20830**	0.00040
PRT1	0.00856	–	0.01148	0.03643	**0.18968**	**0.57004**	**0.90506**	**0.20404**	0.00000	0.02505	0.00000
PRT-AL	**-0.00384**	0.02072	–	**0.26116**	**0.10504**	0.01406	0.00950	0.00119	0.00000	**0.74250**	0.00525
PRT2	**0.00437**	0.00742	**0.00269**	–	**0.36937**	**0.05742**	0.02327	0.00455	0.00000	**0.23572**	0.03010
PRT3	**0.00236**	**0.00077**	**0.00709**	**0.00008**	–	**0.23691**	**0.08554**	0.00347	0.00000	**0.08267**	0.00000
ESP-AN1	0.01157	**-0.00121**	0.02619	**0.00903**	**0.00136**	–	**0.42580**	**0.51401**	0.00000	0.02554	0.00010
ESP-AN2	0.00869	**-0.00133**	0.02356	0.00954	**0.00205**	**-0.00045**	–	**0.21879**	0.00000	0.02327	0.00000
ESP1	0.02011	**0.00054**	0.03677	0.01420	0.00530	**-0.00095**	**0.00062**	–	0.00000	0.01119	0.00000
ESP-BC1	0.12482	0.06162	0.19374	0.12849	0.08049	0.06699	0.06106	0.04900	–	0.00000	0.00000
GBR	**0.01005**	0.05043	**-0.01465**	**0.00837**	**0.02583**	0.05775	0.05200	0.07039	0.32937	–	**0.10593**
FRA-COR	0.03404	0.03807	0.03225	0.01083	0.02783	0.03765	0.03766	0.04476	0.14858	**0.02748**	–
ITA_NW	**0.00310**	0.00713	**0.00515**	**0.00215**	**0.00232**	0.00798	0.00667	0.01248	0.09774	**0.00755**	0.02276
ITA_SE	0.04501	0.09058	0.02073	0.04159	0.06582	0.08725	0.09050	0.11266	0.22146	**-0.01904**	0.05217
ITA-SAR	0.12741	0.20503	0.09865	0.16196	0.17521	0.20789	0.20407	0.24158	0.40935	0.07216	0.19020
DEU-MB	0.02584	0.04734	0.02207	0.04393	0.04000	0.04941	0.04602	0.06375	0.15125	**0.01220**	0.07920
POL-KAS	0.31762	0.36222	0.30658	0.33078	0.33683	0.37859	0.36606	0.39400	0.53592	0.26586	0.35413
POL	0.28625	0.32880	0.27482	0.29715	0.30620	0.33950	0.33109	0.35943	0.47784	0.23040	0.32307
BGR	0.13211	0.19617	0.09901	0.15207	0.17057	0.19295	0.19382	0.22690	0.31864	0.04345	0.18036
TUR_NE	0.10424	0.18248	0.07078	0.11262	0.14667	0.17247	0.18280	0.21664	0.36132	**0.01890**	0.14275
TUR_NW	0.13216	0.21212	0.10178	0.13260	0.17594	0.21035	0.21030	0.24772	0.42134	0.04876	0.14726
TUR_C	0.11275	0.18385	0.08159	0.10668	0.14744	0.17788	0.18541	0.21559	0.39086	**0.02855**	0.12561
TUR1	0.10007	0.17282	0.07074	0.09927	0.13793	0.16302	0.17318	0.20540	0.34802	**0.02088**	0.11908

Table S18 (continued)

	ITA_NW (n=340)	ITA_SE (n=450)	ITA-SAR (n=80)	DEU-MB (n=346)	POL-KAS (n=202)	POL (n=315)	BGR (n=790)	TUR_NE (n=125)	TUR_NW (n=81)	TUR_C (n=85)	TUR1 (n=136)
ECU (PS)	**0.14246**	0.00000	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PRT1	0.00525	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PRT-AL	**0.15385**	0.00525	0.00000	0.01317	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PRT2	**0.19949**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PRT3	**0.08465**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP-AN1	0.03792	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP-AN2	0.01247	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP1	0.00050	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP-BC1	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
GBR	**0.24602**	**0.94911**	0.00713	**0.19533**	0.00000	0.00000	0.01733	**0.11128**	0.01346	**0.06871**	**0.09880**
FRA-COR	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ITA_NW	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ITA_SE	0.05122	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00089	0.00050
ITA-SAR	0.16168	0.08465	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
DEU-MB	0.03076	0.05903	0.09013	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
POL-KAS	0.32055	0.22882	0.27670	0.22763	–	**0.44708**	0.00000	0.00000	0.00000	0.00000	0.00000
POL	0.28983	0.21070	0.25937	0.20090	**-0.00065**	–	0.00000	0.00000	0.00000	0.00000	0.00000
BGR	0.14694	0.06344	0.05307	0.08647	0.13720	0.12965	–	0.00000	0.00000	0.00000	0.00000
TUR_NE	0.12865	0.02071	0.08035	0.10623	0.22200	0.20610	0.04205	–	0.00584	**0.16068**	**0.40442**
TUR_NW	0.15967	0.03939	0.11882	0.15086	0.26328	0.24165	0.09218	0.01890	–	**0.17434**	**0.05693**
TUR_C	0.13429	0.02245	0.10898	0.12771	0.25956	0.23784	0.08421	**0.00474**	**0.00546**	–	**0.56143**
TUR1	0.12114	0.01689	0.10964	0.11590	0.22289	0.20549	0.06765	**0.00019**	**0.00938**	**-0.00165**	–

Table S19: Pairwise F_{ST} genetic distances (below diagonal) and *P*-values (above diagonal) for present study (PS) Ecuadorian samples and European populations. These comparisons were performed considering only the European lineages belonging to haplogroup R. The non-significant values are highlighted in bold (*P*≥0.05). Population codes can be seen in Table S7.

	ECU (n=69)	PRT4 (n=182)	PRT5 (n=57)	PRT6 (n=69)	ESP-SL (n=111)	ESP-V (n=113)	ESP-SEV (n=71)	ESP-BC2 (n=281)	ESP3 (n=420)	IRL1 (n=81)	IRL2 (n=119)	SCT (n=108)	ENG (n=69)
ECU (PS)	–	0.04762	**0.16563**	**0.48490**	**0.19711**	**0.87150**	0.00079	0.00000	**0.32155**	0.00000	0.00000	0.00000	0.00000
PRT4	0.01688	–	0.00307	**0.12929**	0.00000	0.02346	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PRT5	**0.01016**	0.04723	–	**0.08722**	**0.44233**	0.03564	0.00564	0.00059	**0.21146**	0.00000	0.00000	0.00000	0.00000
PRT6	**-0.00277**	**0.00834**	**0.01937**	–	**0.07593**	**0.50827**	0.00030	0.00000	**0.06316**	0.00000	0.00000	0.00000	0.00000
ESP-SL	**0.00605**	0.05264	**-0.00249**	**0.01535**	–	0.02871	0.01178	0.00020	**0.51975**	0.00000	0.00000	0.00000	0.00000
ESP-V	**-0.00883**	0.01689	0.02619	**-0.00323**	0.02042	–	0.00020	0.00000	0.03653	0.00000	0.00000	0.00000	0.00000
ESP-SEV	0.07859	0.13063	0.06130	0.08568	0.03415	0.08129	–	**0.88189**	0.00455	0.00000	0.00000	0.00000	0.00000
ESP-BC2	0.10760	0.18115	0.07587	0.12541	0.04339	0.11471	**-0.00677**	–	0.00000	0.00000	0.00000	0.00000	0.00000
ESP3	**0.00078**	0.05796	**0.00407**	**0.01260**	**-0.00164**	0.01121	0.03057	0.03735	–	0.00000	0.00000	0.00000	0.00000
IRL1	0.44370	0.33209	0.46800	0.43437	0.48296	0.45596	0.62245	0.69750	0.51545	–	**0.65380**	**0.32531**	0.00000
IRL2	0.39576	0.29834	0.41384	0.38845	0.42840	0.41162	0.55330	0.63610	0.47003	**-0.00556**	–	**0.06277**	0.00000
SCT	0.48117	0.36998	0.49967	0.47257	0.51669	0.49201	0.63956	0.71072	0.54724	**0.00081**	**0.01496**	–	0.00000
ENG	0.33910	0.24513	0.34521	0.33317	0.39915	0.36098	0.52961	0.63953	0.44982	0.14674	0.13752	0.12837	–
FRA-B	0.32833	0.24224	0.34803	0.32238	0.36217	0.34701	0.48461	0.57559	0.41136	**0.01100**	-0.00085	0.04180	0.12894
FRA	0.09976	0.04619	0.13731	0.11161	0.17183	0.11240	0.31728	0.42541	0.19513	0.25585	0.22640	0.28642	0.11524
FRA-COR	0.39683	0.32388	0.45736	0.42507	0.48037	0.39656	0.58229	0.65458	0.48516	0.49524	0.47388	0.50322	0.33294
ITA	0.32046	0.23811	0.38322	0.32983	0.41704	0.32138	0.52086	0.62457	0.44257	0.40733	0.39474	0.42105	0.23998
ITA_NW	0.33386	0.26650	0.39388	0.35526	0.41878	0.33588	0.51092	0.59366	0.43963	0.41951	0.40449	0.43144	0.26905
ITA_SE	0.23627	0.15335	0.28743	0.23136	0.32248	0.23793	0.41675	0.52835	0.36119	0.31303	0.30332	0.32880	0.15736
NLD	0.45726	0.34132	0.45403	0.44376	0.51517	0.46510	0.66681	0.73439	0.53112	0.43071	0.40662	0.40876	0.11233
DEU	0.29757	0.21883	0.30425	0.29340	0.36155	0.30951	0.45759	0.56081	0.40433	0.33798	0.32449	0.33212	0.08062
DEU-MB	0.30622	0.22834	0.30564	0.30048	0.36349	0.31811	0.45692	0.55397	0.40495	0.34814	0.33322	0.34027	0.08443
POL	0.37702	0.26215	0.40154	0.36002	0.46430	0.38337	0.60402	0.70181	0.48955	0.40374	0.38976	0.39804	0.12191
TUR2	0.68529	0.49567	0.74699	0.65090	0.72473	0.64032	0.85039	0.84639	0.67326	0.68679	0.65226	0.68073	0.59542

Table S19 (continued)

	FRA-B (n=126)	FRA (n=53)	FRA-COR (n=157)	ITA (n=105)	ITA_NW (n=189)	ITA_SE (n=131)	NLD (n=46)	DEU (n=141)	DEU-MB (n=161)	POL (n=43)	TUR2 (n=91)
ECU (PS)	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PRT4	0.00000	0.00337	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PRT5	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
PRT6	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP-SL	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP-V	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP-SEV	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP-BC2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ESP3	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
IRL1	**0.11108**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
IRL2	**0.37363**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
SCT	0.00287	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ENG	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000
FRA-B	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
FRA	0.17345	–	0.00000	0.00000	0.00010	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
FRA-COR	0.43278	0.18496	–	0.00050	**0.06534**	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ITA	0.35321	0.11543	0.05220	–	**0.07158**	0.01119	0.00000	0.00000	0.00000	0.00000	0.00000
ITA_NW	0.36506	0.13031	**0.00894**	**0.01036**	–	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
ITA_SE	0.26471	0.06291	0.13645	0.02397	0.07212	–	0.00000	0.00000	0.00000	0.00000	0.00000
NLD	0.38266	0.24495	0.44998	0.34603	0.38528	0.24700	–	0.00426	0.02723	0.00238	0.00000
DEU	0.29795	0.10712	0.24382	0.15707	0.19693	0.09511	0.05164	–	**0.73854**	**0.19008**	0.00000
DEU-MB	0.30821	0.12475	0.28001	0.19434	0.23382	0.12380	0.03182	**-0.00374**	–	**0.09554**	0.00000
POL	0.35711	0.15437	0.30435	0.16125	0.22965	0.08177	0.08444	**0.00837**	**0.01601**	–	0.00000
TUR2	0.61445	0.56956	0.64618	0.48292	0.54074	0.34956	0.71543	0.50106	0.51546	0.53215	–