

RESEARCH ARTICLE

# Admixture and Genetic Diversity Distribution Patterns of Non-Recombining Lineages of Native American Ancestry in Colombian Populations

Catarina Xavier<sup>1,2,3</sup>, Juan José Builes<sup>4,5</sup>, Verónica Gomes<sup>1,2</sup>, Jose Miguel Ospino<sup>5</sup>, Juliana Aquino<sup>6</sup>, Walther Parson<sup>3,7</sup>, António Amorim<sup>1,2,8</sup>, Leonor Gusmão<sup>1,2,6</sup>, Ana Goios<sup>1,2\*</sup>

**1** Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, **2** Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal, **3** Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria, **4** Instituto de Biología, Universidad de Antioquia, Medellín, Colombia, **5** Laboratorio Genes Ltda, Medellín, Colombia, **6** DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil, **7** Eberly College of Science, Penn State University, University Park, PA, United States of America, **8** Faculdade de Ciências da Universidade do Porto, Porto, Portugal

\* [aalmeida@ipatimup.pt](mailto:aalmeida@ipatimup.pt)



OPEN ACCESS

**Citation:** Xavier C, Builes JJ, Gomes V, Ospino JM, Aquino J, Parson W, et al. (2015) Admixture and Genetic Diversity Distribution Patterns of Non-Recombining Lineages of Native American Ancestry in Colombian Populations. PLoS ONE 10(3): e0120155. doi:10.1371/journal.pone.0120155

**Academic Editor:** Francesc Calafell, Universitat Pompeu Fabra, SPAIN

**Received:** August 28, 2014

**Accepted:** January 20, 2015

**Published:** March 16, 2015

**Copyright:** © 2015 Xavier et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** Ipatimup is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education, and is partially supported by FCT, the Portuguese Foundation for Science and Technology. CX, VG and AG are also supported by FCT fellowships co-financed by the European Social Fund (Human Potential Thematic Operational Programme; grants SFRH/BD/90873/2012, SFRH/BPD/76207/2011 and SFRH/BPD/43646/2008, respectively). The funders had no role in study

## Abstract

Genetic diversity of present American populations results from very complex demographic events involving different types and degrees of admixture. Through the analysis of lineage markers such as mtDNA and Y chromosome it is possible to recover the original Native American haplotypes, which remained identical since the admixture events due to the absence of recombination. However, the decrease in the effective population sizes and the consequent genetic drift effects suffered by these populations during the European colonization resulted in the loss or under-representation of a substantial fraction of the Native American lineages. In this study, we aim to clarify how the diversity and distribution of uniparental lineages vary with the different demographic characteristics (size, degree of isolation) and the different levels of admixture of extant Native groups in Colombia. We present new data resulting from the analyses of mtDNA whole control region, Y chromosome SNP haplogroups and STR haplotypes, and autosomal ancestry informative insertion-deletion polymorphisms in Colombian individuals from different ethnic and linguistic groups. The results demonstrate that populations presenting a high proportion of non-Native American ancestry have preserved nevertheless a substantial diversity of Native American lineages, for both mtDNA and Y chromosome. We suggest that, by maintaining the effective population sizes high, admixture allowed for a decrease in the effects of genetic drift due to Native population size reduction and thus resulting in an effective preservation of the Native American non-recombining lineages.

design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors declare that no competing interests exist. Although two authors are affiliated to Laboratorio Genes Ltda., this does not alter the authors' adherence to PLOS ONE policies on sharing data and materials.

## Introduction

Colombia was the major entrance point into South America during the peopling of the continent by the Paleoindians [1]. From here, two major migratory routes took place resulting in the colonization of South America: (a) along the Pacific coastline and the Andean regions and (b) towards the Amazonian plains (Fig. 1) [2–4].

The demography of Native ethnic groups in Colombia endured several changes since pre-colonial times, through the Spanish domain and into present day. Several linguistic groups co-exist in Colombia since before the Spanish colonization, some of the most relevant being: the Chibchan, Carib and Arawakan, in the Atlantic coast, Chocoan in the Pacific coast and Paezan, Barbacoan and Quechua in the Southern Andean region (Fig. 1) [5–7]. The colonization by the Spaniards reshaped Native slavery in 1503, and later introduced the African slave trade, leading to severe alterations in the demography of Native groups [5–7], so that the genetic background of Colombia results from a mixture of Native American, African and European contributions. Furthermore, in the initial stages of the colonization, only a minority of European (around 10%) passengers and African slaves to America were women [10–12], leading to a differential male/female admixture ratio visible in several Colombian regions [12,13]. Nowadays Colombia presents 87 recognized ethnic groups and even though Castilian is the official language in Colombia, there are still 64 indigenous languages spoken throughout the country [7]. The Native groups inhabit mainly rural zones and small villages of the country or indigenous reserves; however there is a small minority that lives in the cities normally due to the lack of lands in the reserves or to difficulties in re-adapting to the social/cultural indigenous lifestyle [7].

Colombia has been the target of several forensic and anthropological genetic studies [2–4,12–30]. The genetic composition of Colombian populations has shown to be consistent with that from other South American populations with clear admixed profiles [12]. The proportions of Native American (NAM), African (AFR) and European (EUR) lineages vary throughout the country [10,13,25,27,31], and several small, isolated populations maintain a non-admixed NAM or AFR composition [28,32]. Different sizes and degrees of isolation of the settlements could have led to differential drift effects after the admixture events and to fluctuations of the frequencies of parental lineages in the admixed populations.

In the present work, we have focused on understanding how the different characteristics (source populations, size and degree of isolation, assessed by the differential input of maternal and paternal non-NAM lineages) of the remnant Native groups in Colombia have shaped the diversity and distribution of their uniparentally transmitted lineages, and whether the autosomal ancestry informative insertion-deletion (AIM-InDel) markers confirm these patterns.

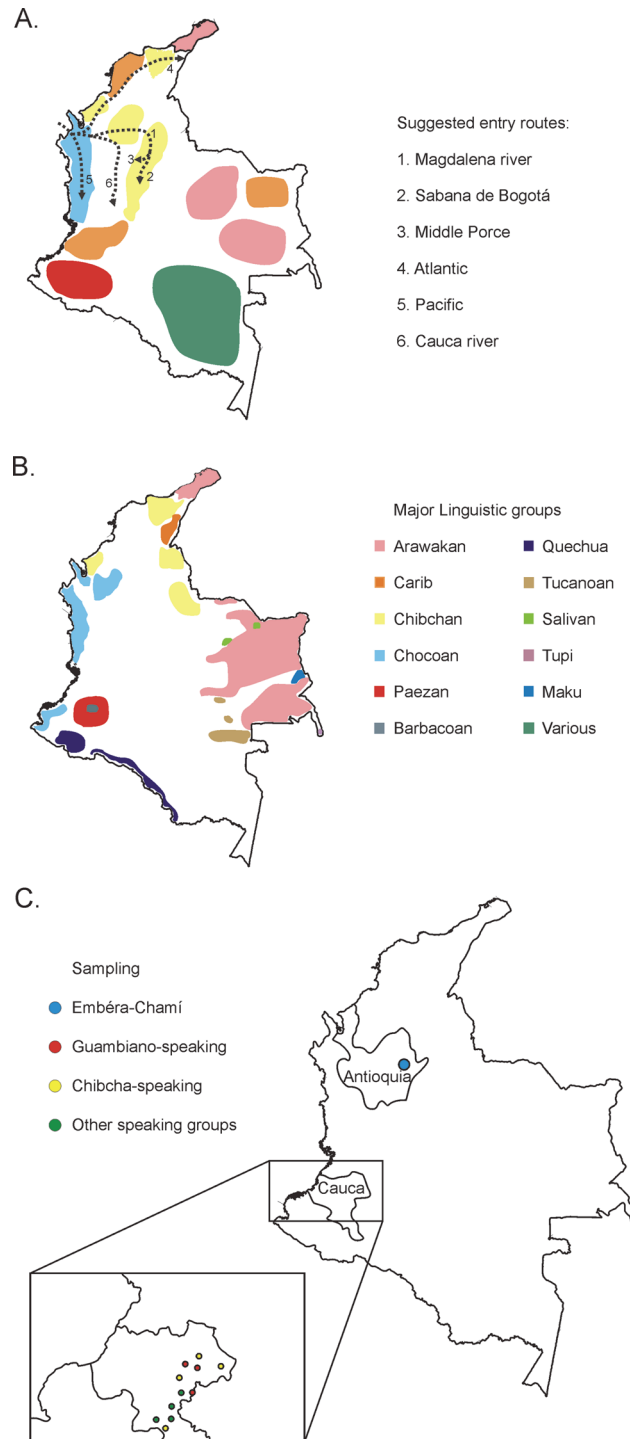
## Materials and Methods

### Ethics Statement

All samples involved in this study are long-lasting anonymized DNA extracts previously obtained from healthy individuals, from paternity cases, under informed consent for research purposes. The current study was approved by the institutional review board of IPATIMUP, and conducted in accordance with the ethical principles of the 2000 Helsinki Declaration of the World Medical Association (<http://www.uma.net/e/policy/b3.htm>).

### Sampling

Sampling was carried out in two distinct Colombian regions (Fig. 1). In the Northern Colombian province of Antioquia, we sampled 38 unrelated Emberá-Chamí individuals from a single settlement—La Po, Segovia (07°04'47"N 74°42'06"W)—whose language belongs to the Emberá



**Fig 1. Map of Colombia showing the geographic distribution of the major linguistic groups at the time of the Spanish colonization (a) and in the present day (b), the major entrance routes in Colombia during the peopling of South America (dashed arrows in a) and the location of the population groups analysed in this study (c). Data used to produce this figure are included in references [5–9].**

doi:10.1371/journal.pone.0120155.g001

group of the Chocoan linguistic family that is considered close to the Chibcha linguistic group [8,33,34]. This is a small, isolated native reserve from which nearly all possible unrelated

individuals were sampled. In Cauca (South Colombia), we sampled 58 unrelated individuals from different municipalities ([S6 Table](#)) and different ethnic and linguistic groups, collected during paternity casework in the Genes Laboratory. The major linguistic groups here represented are: (a) Guambiano ( $n = 33$ ), classified as belonging to the Barbacoan family with an undefined affiliation [[35,36](#)] and composed by individuals from Guambiano and Coconuco ethnic groups and (b) Chibcha, composed by individuals belonging to the Nasa ethnic group ( $n = 14$ ), whose common dialect, Nasa Yuwe, is becoming less and less used by younger generations. This dialect is classified as a Paezan language and inserted in the Chibchan-Paezan macro group [[5,8,33,35](#)]. The linguistic groups of the remaining samples from Cauca are Quechua ( $n = 6$ ), Emberá ( $n = 1$ ) or unidentified ( $n = 4$ ).

**MtDNA amplification and sequencing.** The analysis of the mtDNA control region was performed by PCR/sequencing reactions. The entire control region was amplified using the primers L15997 (5'-CACCATTAGCACCCAAAGCT-3') and H639 (5'-GGGTGATGT-GAGCCCGTCTA-3') [[37,38](#)]. The PCR conditions were as follow: 15 minutes at 95°C for initial denaturation, followed by 35 cycles of 30 seconds at 94°C, 90 seconds at 58°C and 90 seconds at 72°C, followed by a final extension for 10 minutes at 72°C. Amplified mtDNA was purified with ExoSAP-IT (GE Healthcare) under the following conditions: 15 minutes at the temperature of 37°C followed by 15 minutes at 80°C. Sequencing reaction was performed using the Big Dye Terminator v3.1 cycle Sequencing Kit (Applied Biosystems) and the primers L15997 and L16555 (5'-CCC ACA CGT TCC CCT TAA AT-3') for forward sequencing and H016 (5'-CCC GTG AGT GGT TAA TAG GGT-3') and H639 for reverse sequencing [[37,38](#)]. Sequencing reaction conditions were: initial denaturation of 2 minutes at 96°C, followed by 35 cycles of 15 seconds at 96°C and 2 minutes at 60°C, and a final extension for 10 minutes at 60°C. A final purification with Sephadex (Illustra Sephadex DNA Grade; GE Healthcare) was performed before the samples were run in the automatic sequencer ABI 3130XL (Genetic Analyzer 3000, Applied Biosystems).

**Y chromosome genotyping.** Samples for Y chromosome typing include the subset of male individuals from the total group that was typed for mtDNA. From all male samples, those that belong to Native American haplogroup Q were included in a previous study by Roewer et al. [[39](#)]. The remaining samples, which had the ancestral state on Y-SNPs present in Multiplex Q, were typed for additional SNPs through a hierarchical approach using multiplex PCR and single-base extension by SNaPshot (the full set of 35 Y-SNPs investigated in the total sample is indicated in [S1 Fig](#)). First, the Multiplex 1 was typed in all non-haplogroup Q samples as described by Brion et al. [[40](#)] but excluding M22 marker. Samples showing the derived alleles at SRY10831.1 and M213 and the ancestral allele at M9, for the Y-SNPs included in Multiplex 1, were further typed using the Multiplex 2 [[40](#)]. Those samples showing the derived allele at just the SRY10831.1 locus in Multiplex 1 were further typed using the Multiplex E as previously described [[41](#)]. All the multiplexes were analysed as described in Gomes et al. [[41](#)]. All samples were also typed for 17 Y-STRs using the AmpFlSTR Yfiler kit following the manufacturers' recommendations (Applied Biosystems).

## Haplogroup assignment and statistical analysis

MtDNA sequences were compared to the revised Cambridge Reference Sequence (rCRS) [[42](#)] using the software Geneious Pro v. 5.5.6. [[43](#)], following the phylogenetic approach [[44](#)] and the nomenclature adopted by the International Society for Forensic Genetics (ISFG) [[45](#)]. Haplogroups were assigned using Haplogrep software [[46](#)] and corrected using haplogroup assignment based on full mitochondrial genome information as performed by EMMA [[47](#)], both using Phylotree mtDNA built 16, 19 February 2014 [[48](#)]. Y chromosome haplogroups were

named in accordance with Dulik et al. [49] for haplogroup Q, Trombetta et al. [50] for haplogroup E, and Karafet et al. [51] for the remaining haplogroups. The Y-STR alleles were designated according to the ISFG recommendations [52].

MtDNA sequences were submitted to EMPOP ([www.empop.org](http://www.empop.org)) [53,54] and Y-chromosome haplotypes were submitted to YHRD ([www.yhrd.org](http://www.yhrd.org)) [55] for quality control, and are available at the respective websites under accession numbers EMP00538 and YA003649 (Antioquia) and YA003818 (Cauca), respectively.

For both mtDNA and Y chromosome data, haplogroup frequencies were calculated by direct counting, diversity indices considering individuals with Native American matrilineal or patrilineal ancestry were calculated using DNAsp v. 5 [56] and Arlequin v.3.1 [57], and phylogenetic analyses were performed using Network v. 4.6.1.1 software (<http://www.fluxus-engineering.com>). Phylogenetic networks were sequentially constructed applying reduced median and median-joining methods. In the Y-chromosome network the DYS385 marker was not considered and the number of repeats at DYS389II was calculated after subtracting the number of repeats at DYS389I. STR weighting was applied in accordance with Qamar et al. [58] to obtain the most parsimonious network.

## Autosomal marker genotyping and analysis

All Antioquia population samples and 55 Cauca samples were additionally genotyped for the 46 AIM-InDel polymorphisms described in Pereira et al. [59], according to the single multiplex reaction described therein. These markers were selected to efficiently measure population admixture proportions from African, European, East Asian and Native American origins.

The apportionment of genetic ancestral contributions was estimated using the software STRUCTURE v2.3.3 [14,17]. To estimate the ancestral membership proportions, a supervised analysis was performed using prior information on the geographic origin of the reference samples from Africa, Europe and Native America. The STRUCTURE runs comprised three replicates of 100,000 burning steps followed by 100,000 Markov Chain Monte Carlo (MCMC) iterations. A tri-hybrid contribution from Native Americans, Europeans and Africans was assumed (i.e.,  $K = 3$ ). The “Use population Information to test for migrants” option was used with the Admixture model. Allele frequencies were correlated and updated using only individuals with POPFLAG = 1, (i.e., the HGDP-CEPH samples used as reference; data from Pereira et al. [59]).

## Results

MtDNA control region haplotypes observed in the studied Colombian Native American population groups, together with their classification in haplogroups, are described in detail in [S1 Table](#). MtDNA haplogroup frequencies are indicated in [Table 1](#) and [S3 Table](#).

The male haplotypes and haplogroups detected are described in [S2 Table](#), including the data already published in Roewer et al. [39] for the Native American haplogroups. Autosomal AIM-InDel genotypes per sample are described in [S5 Table](#).

## MtDNA haplogroup frequencies

All individuals except one (Cauca—haplogroup L) belong to the typical Native American mtDNA haplogroups A2, B2, C1 and D1 [60].

Emberá-Chamí individuals ( $n = 38$ ) belong to A2, B2 and D1 lineages (but not C), as previously reported for this ethnic group [12,21]. Within haplogroup A2 (0.474), two different sub-lineages were distinguished: A2ac at a predominant frequency (0.421), which was also described for other Colombian groups [21], and A2 (with polymorphisms 64T and 16126C)

**Table 1. mtDNA and Y chromosome major haplogroup frequencies and diversity indices in Colombian population groups.**

Department		Antioquia		Cauca		
		Group	Emberá Chamí	Cauca All	Guambiano-speaking	Chibcha-speaking
mtDNA	Nt		<b>38</b>	<b>58</b>	<b>33</b>	<b>14</b>
	Haplogroup frequencies	A2	0.474	0.190	0.090	0.429
		B2	0.263	0.069	0.061	0.071
		C1	0	0.638	0.758	0.286
		D1	0.263	0.086	0.091	0.143
		L2	0	0.017	0	0.071
	N		<b>38</b>	<b>57</b>	<b>33</b>	<b>13</b>
	Diversity indices	K	7	38	19	13
		S	22	71	45	47
		H	0.751 ± 0.051	0.958 ± 0.018	0.900 ± 0.043	1 ± 0.030
π		0.00785 ± 0.00035	0.00931 ± 0.00069	0.00706 ± 0.00100	0.01163 ± 0.00119	
Y chromosome	Nt		<b>24</b>	<b>48</b>	<b>29</b>	<b>12</b>
	Haplogroup frequencies	Q1a3*	0.167	0.188	0.310	0
		Q1a3a1a*	0.833	0.417	0.483	0.333
		European <sup>a</sup>	0	0.354	0.207	0.500
		African <sup>a</sup>	0	0.042	0	0.167
	N		<b>24</b>	<b>29</b>	<b>23</b>	<b>4</b>
	Diversity indices	K	13	21	16	3
		Gene Diversity	0.9239 ± 0.0323	0.9754 ± 0.0153	0.9644 ± 0.0224	0.8333 ± 0.2224

Diversity indices were calculated considering only Native American haplotypes. Nt stands for the total number of samples, N stands for the number of samples with a maternal or paternal Amerindian lineage that were considered for diversity indices calculation, K is the number of different haplotypes, S is the number of segregating sites, H is the haplotype diversity, π is the nucleotide diversity and MNPD is the mean number of pairwise differences.

<sup>a</sup>See S4 Table for details.

doi:10.1371/journal.pone.0120155.t001

(0.053) which was also found in Emberá and Waunana groups from Panama [61]. Considering haplogroup B2 (0.263), it is noteworthy that all samples (in all linguistic sets) belong to the B2d haplogroup. Finally two different sub-lineages were found in haplogroup D (0.263): D1+16239T+16286T (0.158) that was not previously described, and D1 (0.105).

In the Cauca population (n = 58) all four typical Native American haplogroups are present, with haplogroup C1 at a highest frequency (0.638), followed by A2 (0.190), D1 (0.086) and B2 (0.069). Because this population sample is heterogeneous in what concerns the origin of its individuals (both in location and in linguistic groups), we have analysed its major speaking groups separately.

Guambiano-speaking group (n = 33) presents only Native American mtDNA haplogroups, with haplogroups A2 (0.091) and B2 (0.061) each one represented by one sub-lineage: A2ac and B2d. Within haplogroup C (0.758), the two most frequent sub-haplogroups observed were C1b (0.485) and C1d (0.273), both recognized as founder haplogroups in the Americas [60]. The majority of samples (0.394; C1b+16169T and C1b12) found within C1b also presented the 16169T polymorphism, which is also present in the Chibcha-speaking group. Within C1d three haplotypes were observed, C1d (+194T) being the most frequent (0.212). Haplogroup D1 (0.091) was represented by two haplotypes (see S3 Table for further details on the least frequent lineages).

Chibcha speaking group ( $n = 14$ ) presents one African haplogroup, L2a1c1 that descends from L2a1 which is a typical African lineage in Bantu populations [62]. All four major Native American haplogroups are present in this group. A2 (0.429) is represented by 5 different haplotypes, B2d (0.071) by one, C1b (0.214) by three, C1d by one (0.071) and D1f (0.143) by two (S3 Table).

## Y chromosome haplogroup frequencies

The Y chromosome haplogroup frequencies observed in our samples are indicated in Table 1 and S4 Table.

Regarding Y chromosome haplogroups there is clearly a higher proportion of non-Native American lineages than for mtDNA, particularly in the populations from the Cauca department where various European lineages were found, in agreement with the asymmetric gender introgression pattern historically documented.

The Emberá-Chamí individuals from Antioquia present only Native American haplogroups, with the M3 derived lineage Q1a3a1a\* at a higher frequency (0.833) than the M3 ancestral Q1a3\* (0.167).

In the Cauca populations, Native American (0.604), European (0.354), and sub-Saharan African (0.042) haplogroups were found.

The Guambiano-speaking group shows both the Q1a3a1a\*-M3 (0.483) and the Q1a3\*-M346 (0.310) lineages at similar proportions, and the remaining 0.207 of the Y chromosomes belong to the European haplogroups R1b1-P25 and J2-M172.

In the Chibcha-speaking group the Native American lineage Q1a3\*-M346 is absent, and only four individuals present the M3 derived lineage Q1a3a1a\* (0.333), while the non-Native American lineages are present at a higher frequency (0.667) and represented by haplogroups R1b1-P25, G-M201, J2-M172, E1b1b1a1-M78 and E1b1b1c-M123 (0.500) that are typical in Europeans, and by two samples from the sub-Saharan African E1b1a1\*-M2 haplogroup (0.167). Despite its small sample size, Sub-Saharan haplogroups were detected only in the Chibcha group, similarly to the results obtained for mtDNA.

## Diversity Indices

Maternal or paternal diversities were calculated only within lineages of Native American ancestry (Table 1).

MtDNA diversity indices showed generally higher values in the Cauca samples than in the Emberá-Chamí. While the Emberá-Chamí population presents only 7 different haplotypes in 38 samples ( $H = 0.751 \pm 0.051$ ), the 57 Cauca individuals presented 38 different haplotypes ( $H = 0.958 \pm 0.018$ ). Both major speaking-groups from Cauca present higher diversity values than the Emberá-Chamí population: the Guambiano-speaking group presents 19 different haplotypes in 33 individuals ( $H = 0.900 \pm 0.043$ ), and the Chibcha-speaking group is highly diverse with 13 different haplotypes in 13 individuals ( $H = 1 \pm 0.030$ ).

Because the Chibcha sample presented few male individuals with Native American Y lineages ( $N = 4$ ), the diversity indices for this population should be regarded with caution. For the other populations, however, the Y chromosome diversity indices follow the same trend than that described for mtDNA. The Emberá-Chamí population presents lower gene diversity ( $0.924 \pm 0.032$ ) than the Guambiano ( $0.964 \pm 0.022$ ) or the Cauca population considered as a whole ( $0.975 \pm 0.015$ ), and the same is observed for the number of different haplotypes: 13 in 24 individuals for the Emberá-Chamí and 16 in 23 individuals for the Guambiano-speaking group. Despite this observation, the Emberá-Chamí population presents a slightly higher mean

number of pairwise differences (MNPD) ( $9.214 \pm 4.390$ ) than the Guambiano-speaking group ( $8.075 \pm 3.891$ ) or the total Cauca population ( $8.325 \pm 3.971$ ).

## Phylogenetic analysis

The network analysis shows clearly the presence of the four typical mtDNA Native American lineages in the Colombian populations (Fig. 2A), with the Antioquia Emberá-Chamí population presenting only 5 mtDNA founder haplotypes while the Cauca populations present a higher number of lineages. In the Y chromosome network (Fig. 2B) the same differences between populations can be visualised, although the lineages are more diverse than in the mtDNA.

It is also noticeable that, for both mtDNA and Y chromosome, no haplotypes are shared between Antioquia and Cauca groups, although in each group there are various haplotype clusters distant from each other. This observation, together with the values of MNPD described above, is probably the result of bottleneck events acting on genetically differentiated and somewhat isolated Native American groups, as well as of the small sizes of the samples analysed.

## Autosomal AIM-InDel results and comparison with lineage markers

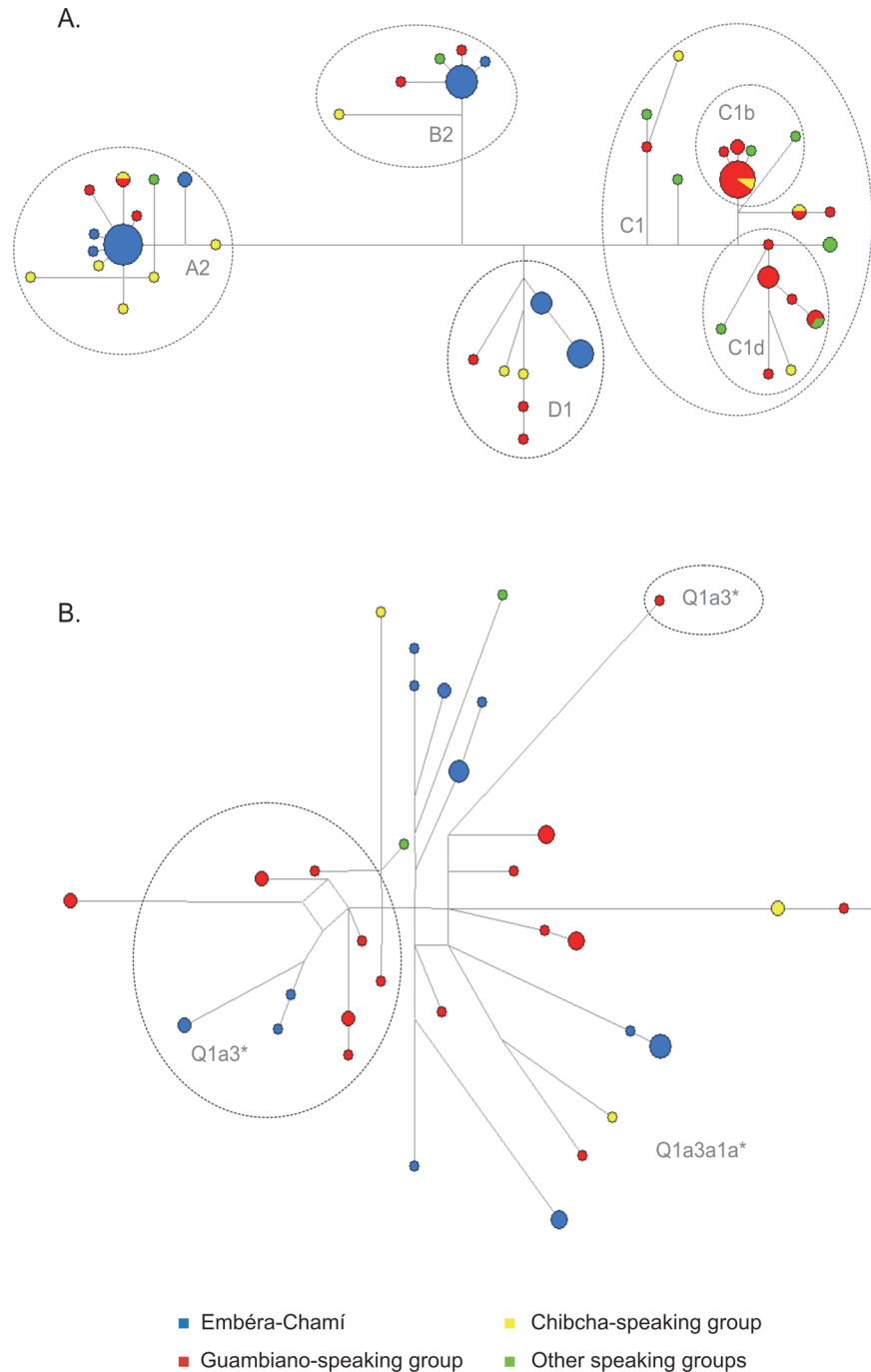
The genotyping results obtained for the 46 AIM-InDels are listed in S5 Table. These data were used to calculate the Native American (NAM), European (EUR) and African (AFR) contributions using STRUCTURE software [14,17], as detailed in M&M. Admixture proportions obtained for mtDNA, Y chromosome and AIM-InDels are compared in Table 2. The average value between lineage markers ancestries and the ancestry value obtained for the autosomal markers were compared to check for multiple events of sex-biased gene flow, as reported previously for some Colombian admixed populations [25,28]. This trend is clear in admixed populations but not in NAM populations like Cauca. Although the Cauca sample presents almost 100% of NAM mtDNA haplotypes, when averaging the proportions obtained for mtDNA and Y chromosome, the values obtained are very similar to those observed for the autosomal AIM-InDels. In the Antioquia sample, while both lineage markers presented 100% NAM ancestry, the autosomal AIM-InDels present ~3% EUR ancestry and ~2% AFR ancestry. These values may be considered residual non-NAM ancestry that could only be detected with the analysis of the autosomal InDels due to the higher number of unlinked markers studied in twice the number of chromosomes analysed in the non-recombining portions of the genome.

## Discussion

The study of mtDNA, Y chromosome and autosomal AIM-InDel data in these groups contributes to a better knowledge of the genetic composition of Colombian populations.

We show several important differences between the two samples, including different genetic composition and different degrees of admixture. Indeed, while the proportion of NAM matrilineal ancestry is well preserved in all samples, the proportion of paternal influx differed widely among samples, as reflected in the AIM-InDel data. The Emberá-Chamí sample from Antioquia presented exclusively Native American haplogroups both in mtDNA and Y chromosome, and only a residual proportion of non-Native ancestry in the autosomal AIM-InDels. The Cauca sample revealed a significant non-NAM male input (0.396), particularly within the Chibcha-speaking group, with its mtDNA haplogroups being almost exclusively NAM (only one AFR haplotype was detected), which reflects the traditional sex biased admixture between NAM and EUR men [12,13,25]. This sample harbours a proportion of EUR and AFR autosomal ancestry that is equivalent to the average between the values obtained for the lineage markers, in contrast to what was recently observed for Colombian Mestizo groups from the same departments [28], suggesting no important sex biased admixture events occurred recently. The





**Fig 2. Phylogenetic network of the Native American mtDNA (a) and Y chromosome (b) haplotypes detected in this study.** Circle size is proportional to the number of haplotypes and branch size is proportional to the number of polymorphisms that distinguish each pair of haplotypes. Dashed lines delimit different haplogroups.

doi:10.1371/journal.pone.0120155.g002

Emberá-Chamí sample demonstrates a strikingly different profile due to the complete absence of non-NAM lineages in both lineage markers analysed, an observation that is rare in extant NAM populations, particularly in what concerns the paternal lineages. Consistently, it also presents lower diversity values (particularly a small number of different haplotypes) than the

**Table 2. Native American (NAM), European (EUR) and African (AFR) admixture proportions in the Colombian samples from the departments of Cauca and Antioquia.**

	Cauca			Antioquia				
	n	NAM	EUR	AFR	n	NAM	EUR	AFR
mtDNA	58	98.30%	0.00%	1.70%	38	100.00%	0.00%	0.00%
Y chromosome	48	60.42%	35.42%	4.17%	24	100.00%	0.00%	0.00%
Average between lineage markers		79.36%	17.71%	2.94%		100.00%	0.00%	0.00%
AIM-InDels	110	77.72%	17.48%	4.81%	76	94.96%	3.32%	1.71%

n stands for the number of chromosomes analysed

doi:10.1371/journal.pone.0120155.t002

Cauca populations, which may be justified by a relatively recent bottleneck event caused by a decrease in the effective population size during the Spanish colonization; despite this, it still retains representatives of 3 of the different major NAM mtDNA lineages and of 2 different NAM Y chromosome haplogroups.

It is also noteworthy that the groups with a higher degree of Y chromosome non-NAM input show also a higher NAM mtDNA diversity. Although we can not exclude that larger Native populations may have attracted more non-Native admixture, we also suggest that populations exposed to a higher European admixture were less subjected to a decrease in effective population sizes and thus their Native non-recombining lineages were less prone to be erased by genetic drift. This protective effect of the admixture events on the genetic diversity of uniparental lineages is clear in the populations from Cauca, a less isolated and more cosmopolitan region than the Emberá-Chamí population from Antioquia.

The study of the Native American genetic diversity has been complicated by the fact that, although the more isolated populations retain the Native American genetic identity, they paid the cost of losing most of its diversity by the decrease in effective population sizes after the European colonization. However this study demonstrates that, for lineage markers, it is possible to recover more of the Native American genetic diversity from the study of a comprehensive sampling of Native American ancestry individuals in admixed populations than from the analysis of isolated communities.

## Supporting Information

### S1 Table. mtDNA haplotypes.

(XLSX)

### S2 Table. Y chromosome data.

(XLSX)

### S3 Table. Frequencies of mtDNA sublineages.

(XLSX)

### S4 Table. Y chromosome haplogroup frequencies.

(XLSX)

### S5 Table. Autosomal AIM-InDels data.

(XLSX)

### S6 Table. Population density of the sampled Cauca municipalities.

(XLSX)

**S1 Fig. Phylogenetic tree of Y-haplogroups analyzed in the present study.** The haplogroups are named in accordance with Karafet et al.[51].upgraded by Trombetta et al.[50] and Dulik et al.[49] for haplogroups E and Q, respectively. (DOCX)

## Author Contributions

Conceived and designed the experiments: AA LG AG. Performed the experiments: CX JJB VG JMO JA. Analyzed the data: CX VG WP LG AG. Contributed reagents/materials/analysis tools: JJB JMO AA LG. Wrote the paper: CX VG AA LG AG.

## References

1. Rothhammer F, Dillehay TD. The Late Pleistocene Colonization of South America: An Interdisciplinary Perspective. *Ann Hum Genet.* 2009; 73: 540–549. doi: [10.1111/j.1469-1809.2009.00537.x](https://doi.org/10.1111/j.1469-1809.2009.00537.x) PMID: [19691551](https://pubmed.ncbi.nlm.nih.gov/19691551/)
2. Keyeux G, Rodas C, Gelvez N, Carter D. Possible migration routes into South America deduced from mitochondrial DNA studies in Colombian Amerindian populations. *Hum Biol.* 2002; 74: 211–233. PMID: [12030650](https://pubmed.ncbi.nlm.nih.gov/12030650/)
3. Melton PE, Briceño I, Gómez A, Devor EJ, Bernal JE, Crawford MH. Biological relationship between central and South American Chibchan speaking populations: Evidence from mtDNA. *Am J Phys Anthropol.* 2007; 133: 753–770. PMID: [17340631](https://pubmed.ncbi.nlm.nih.gov/17340631/)
4. Casas-Vargas A, Gómez A, Briceño I, Díaz-Matallana M, Bernal JE, Rodríguez JV. High genetic diversity on a sample of pre-Columbian bone remains from Guane territories in northwestern Colombia. *Am J Phys Anthropol.* 2011; 146: 637–649. doi: [10.1002/ajpa.21626](https://doi.org/10.1002/ajpa.21626) PMID: [21990065](https://pubmed.ncbi.nlm.nih.gov/21990065/)
5. Arango R, Sánchez E. Los pueblos indígenas de Colombia 1997, desarrollo y territorio. Departamento Nacional de Planeación; Editors T, Unidad Administrativa Especial de Desarrollo Territorial; 1998.
6. Arango R, Sánchez SG. Los pueblos indígenas de Colombia en el umbral del nuevo milenio: Población, cultura y territorio: bases para el fortalecimiento social y económico de los pueblos indígenas. Departamento Nacional de Planeación; 2004.
7. DANE. Colombia una nación multicultural. Su diversidad étnica. Departamento Administrativo Nacional de Estadística; 2007.
8. Campbell L. *American Indian Languages: The historical linguistics of Native America.* Oxford University Press; 2000.
9. Aceituno FJ, Loaiza N, Delgado-Burbano ME, Barrientos G. The initial human settlement of Northwest South America during the Pleistocene/Holocene transition: Synthesis and perspectives. *Quat Int.* 2013; 301: 23–33.
10. Curtin PD. *The Atlantic Slave Trade: A Census.* University of Wisconsin Press; 1972.
11. Sánchez-Albornoz N. *La población de América Latina: desde los tiempos precolombinos al año 2000.* Alianza; 1973.
12. Mesa NR, Mondragón MC, Soto ID, Parra MV, Duque C, Ortíz-Barrientos D, et al. Autosomal, mtDNA, and Y-chromosome diversity in Amerinds: Pre- and post-Columbian patterns of gene flow in South America. *Am J Hum Genet.* 2000; 67: 1277–1286. PMID: [11032789](https://pubmed.ncbi.nlm.nih.gov/11032789/)
13. Carvajal-Carmona LG, Soto ID, Pineda N, Ortíz-Barrientos D, Duque C, Ospina-Duque J, et al. Strong Amerind/White sex bias and a possible Sephardic contribution among the founders of a Population in Northwest Colombia. *Am J Hum Genet.* 2000; 67: 1287–1295. PMID: [11032790](https://pubmed.ncbi.nlm.nih.gov/11032790/)
14. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. *Genetics.* 2000; 155: 945–959. PMID: [10835412](https://pubmed.ncbi.nlm.nih.gov/10835412/)
15. Yunis JJ, Yunis EJ, Yunis E. Genetic relationship of the Guambino, Paez, and Ingano Amerindians of Southwest Colombia using major histocompatibility complex class II haplotypes and blood groups. *Hum Immunol.* 2001; 62: 970–978. PMID: [11543899](https://pubmed.ncbi.nlm.nih.gov/11543899/)
16. Yunis JJ, Garcia O, Yunis EJ. Population frequencies for CSF1PO, TPOX, TH01, F13A01, FES/FPS and VWA in seven Amerindian populations from Colombia. *J Forensic Sci.* 2005; 50: 682–684. PMID: [15932108](https://pubmed.ncbi.nlm.nih.gov/15932108/)
17. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics.* 2003; 164: 1567–1587. PMID: [12930761](https://pubmed.ncbi.nlm.nih.gov/12930761/)

18. Gaviria Ab, Ibarra AA, Jaramillo N, Palacio OD, Acosta MA, Brion A, et al. Nineteen autosomal microsatellite data from Antioquia (Colombia). *Forensic Sci Int.* 2004; 143: 69–71. PMID: [15177633](#)
19. Gaviria AA, Ibarra AA, Palacio OD, Posada YC, Triana O, Ochoa LM, et al. Y-chromosome haplotype analysis in Antioquia (Colombia). *Forensic Sci Int.* 2005; 151: 85–91. PMID: [15935946](#)
20. Palacio OD, Triana O, Gaviria A, Ibarra AA, Ochoa LM, Posada Y, et al. Autosomal microsatellite data from Northwestern Colombia. *Forensic Sci Int.* 2006; 160: 217–220. PMID: [16024199](#)
21. Torres MM, Bravi CM, Bortolini M-C, Duque C, Callegari-Jacques S, Ortiz D, et al. A revertant of the major founder Native American haplogroup C common in populations from northern South America. *Am J Hum Biol.* 2006; 18: 59–65. PMID: [16378344](#)
22. Builes JJ, Bravo ML, Gómez C, Espinal C, Aguirre D, Gómez A, et al. Y-chromosome STRs in an Antioquian (Colombia) population sample. *Forensic Sci Int.* 2006; 164: 79–86. PMID: [16289613](#)
23. Builes JJ, Martínez B, Gómez A, Caraballo L, Espinal C, Aguirre D, et al. Y chromosome STR haplotypes in the Caribbean city of Cartagena (Colombia). *Forensic Sci Int.* 2007; 167: 62–69. PMID: [16455219](#)
24. Ibarra A, Freire-Aradas A, Martínez M, Fondevila M, Burgos G, Camacho M, et al. Comparison of the genetic background of different Colombian populations using the SNPforID 52plex identification panel. *Int J Legal Med.* 2014; 128: 19–25. doi: [10.1007/s00414-013-0858-z](#) PMID: [23665814](#)
25. Bedoya G, Montoya P, García J, Soto I, Bourgeois S, Carvajal L, et al. Admixture dynamics in Hispanics: A shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci U S A.* 2006; 103: 7234–7239. PMID: [16648268](#)
26. Silva A, Briceño I, Burgos J, Torres D, Villegas V, Gómez A, et al. Análisis de ADN mitocondrial en una muestra de restos óseos arcaicos del período Herrera en la sabana de Bogotá. *Biomédica.* 2008; 28: 569–577.
27. Salas A, Acosta A, Álvarez-Iglesias V, Cerezo M, Phillips C, Lareu MV, et al. The mtDNA ancestry of admixed Colombian populations. *Am J Hum Biol.* 2008; 20: 584–591. doi: [10.1002/ajhb.20783](#) PMID: [18442080](#)
28. Rojas W, Parra MV, Campo O, Caro MA, Lopera JG, Arias W, et al. Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers. *Am J Phys Anthropol.* 2010; 143: 13–20. doi: [10.1002/ajpa.21439](#) PMID: [21086525](#)
29. Yang NN, Mazières S, Bravi C, Ray N, Wang S, Burley MW, et al. Contrasting Patterns of Nuclear and mtDNA Diversity in Native American Populations. *Ann Hum Genet.* 2010; 74: 525–538. doi: [10.1111/j.1469-1809.2010.00608.x](#) PMID: [20887376](#)
30. Usme-Romero S, Alonso M, Hernandez-Cuervo H, Yunis EJ, Yunis JJ. Genetic differences between Chibcha and Non-Chibcha speaking tribes based on mitochondrial DNA (mtDNA) haplogroups from 21 Amerindian tribes from Colombia. *Genet Mol Biol.* 2013; 36: 149–157 doi: [10.1590/S1415-47572013005000011](#) PMID: [23885195](#)
31. Gómez-Pérez L, Alfonso-Sánchez MA, Pérez-Miranda AM, García-Obregón S, Builes JJ, Bravo ML, et al. Genetic admixture estimates by Alu elements in Afro-Colombian and Mestizo populations from Antioquia, Colombia. *Ann Hum Biol.* 2010; 37: 488–500. doi: [10.3109/03014460903433810](#) PMID: [20113181](#)
32. Noguera MC, Schwegler A, Gomes V, Briceño I, Alvarez L, Uricoechea D, et al. Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolívar. *Ann Hum Biol.* 2014; 41: 453–459. doi: [10.3109/03014460.2013.852244](#) PMID: [24215508](#)
33. Greenberg JH. *Language in the Americas.* Stanford University Press; 1987.
34. Constenla UA, Margery PE. Elementos de fonología comparada Chocó. *Filología y lingüística.* 1991; 17: 137–191. PMID: [11740856](#)
35. Curnow TJ. Why Paez is not a Barbacoan language: The nonexistence of "Moguex" and the use of early sources. *International Journal of American Linguistics.* 1998; 64: 338–351.
36. Adelaar WFH, Muysken PC. *The Languages of the Andes.* Cambridge University Press; 2004.
37. Alonso A, Albarrán C, Martin P, Garcia P, Garcia O, de la Rúa C, et al. Multiplex-PCR of short amplicons for mtDNA sequencing from ancient DNA. *International Congress Series.* 2003; 1239: 585–588.
38. Parson W, Bandelt HJ. Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Sci Int Genet.* 2007; 1: 13–19. doi: [10.1016/j.fsigen.2006.11.003](#) PMID: [19083723](#)
39. Roewer L, Nothnagel M, Gusmão L, Gomes V, González M, Corach D, et al. Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in Native South Americans. *PLoS Genet.* 2013; 9: e1003460. doi: [10.1371/journal.pgen.1003460](#) PMID: [23593040](#)
40. Brion M, Sobrino B, Blanco-Verea A, Lareu MV, Carracedo A. Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *Int J Legal Med.* 2005; 119: 10–15. PMID: [15095093](#)

41. Gomes V, Sánchez-Diz P, Amorim A, Carracedo Á, Gusmão L. Digging deeper into East African human Y chromosome lineages. *Hum Genet.* 2010; 127: 603–613. doi: [10.1007/s00439-010-0808-5](https://doi.org/10.1007/s00439-010-0808-5) PMID: [20213473](https://pubmed.ncbi.nlm.nih.gov/20213473/)
42. Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet.* 1999; 23: 147. PMID: [10508508](https://pubmed.ncbi.nlm.nih.gov/10508508/)
43. Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, et al. Geneious V5.4, 2011.
44. Bandelt HJ, Parson W. Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *Int J Legal Med.* 2008; 122: 11–21. PMID: [17347766](https://pubmed.ncbi.nlm.nih.gov/17347766/)
45. Carracedo A, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, et al. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int.* 2000; 110: 79–85. PMID: [10808096](https://pubmed.ncbi.nlm.nih.gov/10808096/)
46. Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat.* 2011; 32: 25–32. doi: [10.1002/humu.21382](https://doi.org/10.1002/humu.21382) PMID: [20960467](https://pubmed.ncbi.nlm.nih.gov/20960467/)
47. Röck AW, Dür A, van Oven M, Parson W. Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). *Forensic Sci Int Genet.* 2013; 7: 601–609. doi: [10.1016/j.fsigen.2013.07.005](https://doi.org/10.1016/j.fsigen.2013.07.005) PMID: [23948335](https://pubmed.ncbi.nlm.nih.gov/23948335/)
48. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009; 30: E386–E394. doi: [10.1002/humu.20921](https://doi.org/10.1002/humu.20921) PMID: [18853457](https://pubmed.ncbi.nlm.nih.gov/18853457/)
49. Dulik MC, Owings AC, Gaieski JB, Vilar MG, Andre A, Lennie C, et al. Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations. *Proc Natl Acad Sci U S A.* 2012; 109: 8471–8476. doi: [10.1073/pnas.1118760109](https://doi.org/10.1073/pnas.1118760109) PMID: [22586127](https://pubmed.ncbi.nlm.nih.gov/22586127/)
50. Trombetta B, Cruciani F, Sellitto D, Scozzari R. A new topology of the human Y chromosome haplogroup E1b1 (E-P2) revealed through the use of newly characterized binary polymorphisms. *PLoS ONE.* 2011; 6: e16073. doi: [10.1371/journal.pone.0016073](https://doi.org/10.1371/journal.pone.0016073) PMID: [21253605](https://pubmed.ncbi.nlm.nih.gov/21253605/)
51. Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 2008; 8:830–838.
52. Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, et al. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Int J Legal Med.* 2006; 120: 191–200. PMID: [16998969](https://pubmed.ncbi.nlm.nih.gov/16998969/)
53. Parson W, Brandstätter A, Pircher M, Steinlechner M, Scheithauer R. EMPOP—the EDNAP mtDNA population database concept for a new generation, high-quality mtDNA database. *International Congress Series.* 2004; 1261: 106–108.
54. Parson W, Dür A. EMPOP—A forensic mtDNA database. *Forensic Sci Int Genet.* 2007; 1: 88–92. doi: [10.1016/j.fsigen.2007.01.018](https://doi.org/10.1016/j.fsigen.2007.01.018) PMID: [19083735](https://pubmed.ncbi.nlm.nih.gov/19083735/)
55. Willuweit S, Roewer L. Y chromosome haplotype reference database (YHRD): Update. *Forensic Sci Int Genet.* 2007; 1: 83–87. doi: [10.1016/j.fsigen.2007.01.017](https://doi.org/10.1016/j.fsigen.2007.01.017) PMID: [19083734](https://pubmed.ncbi.nlm.nih.gov/19083734/)
56. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinform.* 2009; 25: 1451–1452.
57. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010; 10: 564–567. doi: [10.1111/j.1755-0998.2010.02847.x](https://doi.org/10.1111/j.1755-0998.2010.02847.x) PMID: [21565059](https://pubmed.ncbi.nlm.nih.gov/21565059/)
58. Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, et al. Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet.* 2002; 70: 1107–1124. PMID: [11898125](https://pubmed.ncbi.nlm.nih.gov/11898125/)
59. Pereira R, Phillips C, Pinto N, Santos C, Santos SEBd, Amorim A, et al. Straightforward inference of ancestry and admixture proportions through Ancestry-Informative Insertion Deletion multiplexing. *PLoS ONE.* 2012; 7: e29684. doi: [10.1371/journal.pone.0029684](https://doi.org/10.1371/journal.pone.0029684) PMID: [22272242](https://pubmed.ncbi.nlm.nih.gov/22272242/)
60. Achilli A, Perego UA, Bravi CM, Coble MD, Kong QP, Woodward SR, et al. The phylogeny of the four Pan-American MtDNA haplogroups: Implications for evolutionary and disease studies. *PLoS ONE.* 2008; 3: e1764. doi: [10.1371/journal.pone.0001764](https://doi.org/10.1371/journal.pone.0001764) PMID: [18335039](https://pubmed.ncbi.nlm.nih.gov/18335039/)
61. Kolman CJ, Bermingham E. Mitochondrial and nuclear DNA diversity in the Chocó and Chibcha Amerinds of Panamá. *Genetics.* 1997; 147: 1289–1302. PMID: [9383071](https://pubmed.ncbi.nlm.nih.gov/9383071/)
62. Salas A, Richards M, Lareu M-V, Scozzari R, Coppa A, Torroni A, et al. The African diaspora: Mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet.* 2004; 74: 454–465. PMID: [14872407](https://pubmed.ncbi.nlm.nih.gov/14872407/)