‹ Back

Research Article | 🔓 **Open Access** | (cc) (i) ▼

Application of Targeted Y-Chromosomal Capture Enrichment to Increase the Resolution of Native American Haplogroup Q

Zehra Köksal ✉, Claus Børsting, Graciela Bailliet, Germán Burgos, Elizeu Carvalho, Andrea Casas-Vargas, Adriana Castillo, Marilia Brito Gomes, Beatriz Martínez, Humberto Ossa **… See all authors** ⌄

Abstract

Y-chromosomal haplogroups and the Y-SNPs defining them are relevant for the exploration of male lineages, inference of paternal ancestry, and reconstruction of migration pathways, to name a few. Currently, over 300,000 Y-SNPs have been reported, defining 20 main haplogroups. However, ascertainment bias in the investigations has led to some haplogroups being overlooked, which hinders a representative depiction of certain populations and their migration events. For migration pattern analyses of the first settlers of the Americas, the Native American main founding lineage Q-M3 needs to be further investigated to allow clear genetic differentiation of individuals of different ethnogeographic origins. To increase the resolution within this haplogroup, a total of 7.45 Mb of the Y chromosome of 59 admixed South Americans of haplogroup Q was targeted for sequencing using hybridization capture enrichment. Data were combined with 218 publicly available sequences of Central and South Americans of haplogroup Q. After rigorous data processing, variants not meeting the quality criteria were excluded and 4128 reliable Y-SNPs were reported. A total of 2224 Y-SNPs had previously unknown positions in the phylogenetic tree, and 1291 of these are novel. The phylogenetic relationships between the Y-SNPs were established using the software SNPtotree in order to report a redesigned phylogenetic tree containing 300 branches, defined by 3400 Y-SNPs. The new tree introduces 117 previously undescribed branches and is the most comprehensive phylogenetic tree of the Native American haplogroup Q lineages to date. The 214 sequences were assigned to 135 different low- to high-resolution branches, while in the previous phylogenetic tree, only 195 sequences could be sorted into 14 low-resolution branches with the same quality criteria.

‹ Back

▼

1. Introduction

The Y chromosome plays a big role in evolutionary, population, and forensic genetics. The importance of the Y chromosome in these fields lies in its unique combination of paternal inheritance, haploidy and lack of recombination (in the majority of chromosome), and its large size compared to mitochondrial DNA, which is the maternally inherited counterpart among the haploid markers.

The ability to look back into human history is based on the identification of Y-chromosomal SNPs (Y-SNPs) that are used to define Y-chromosomal haplogroups. Y-SNPs can be applied for paternal ancestry inference, reconstruction of migration patterns, and (in association with Y-STRs) estimation of the ages of pedigrees and time to most recent common ancestors (TMRCAs). This is possible, because related males share nearly identical Y chromosomes [1, 2]. Only when a mutation in the Y chromosome is introduced into the germline of a man will this mutation be inherited to the respective son and all male descendants of the son. Thus, haplogroups and characteristic Y-SNPs of founding ancestors are shared by modern men belonging to the same paternal lineage. More recent mutation events further diversified the Y chromosomes and formed additional branches that are characteristic for certain geographic regions, populations, and time ranges [2].

The human Y chromosome has been systematically neglected in genome-wide association studies, because it has half of the copies of the autosomes and a highly complex structure [3]. It varies considerably in length (from 45.2 to 84.9 Mb) [4] and consists of large portions of repetitive sequences, which are mostly pseudoautosomal, heterochromatic, ampliconic, and X-transposed regions, where it is difficult to identify true variants [5]. There are nine unique regions spanning 8.97 Mb [6] that can be investigated with massively parallel sequencing (MPS). MPS enables the efficient identification of large numbers of SNPs and continuously replaces methods targeting limited numbers of Y-SNPs, like the single base extension (SBE) method, which combines multiplex PCRs, SBE reaction(s), and capillary electrophoresis (CE) [7–10].

High-throughput targeted MPS can be applied to sequence Y-SNPs [11–13] and autosomal SNPs [14, 15] using PCR enrichment. Another option is hybridization capture enrichment, which uses oligonucleotides (baits) to capture the regions of interest. This method is widely used in ancient DNA analysis and enables sequencing of most regions of interest within the Y chromosome [16–21].

SNP database ISOGG Y-DNA Haplogroup Tree 2019–2020 and the YFull Ytree v10.07.00 (last accessed February 2023). The most thoroughly researched haplogroup R is represented by around 22% of all Y-SNPs reported in databases. Other haplogroups, like haplogroup Q that is widely distributed in Central and North Asia and the Americas [22], lack further resolution in the subhaplogroups. Currently, only 3%–5% of the total Y-SNPs reported in databases represent sublineages within haplogroup Q.

The low resolution is particularly evident for the lineage Q1b1a1a-M3 that is the most frequent among the founding lineages of the first settlers of the Americas [23–26], followed by Q1b1a1a1-M848 [27, 28], and the much less prevalent Native American lineage C3-M217∗ [25, 29, 30]. Despite the high genetic variability within many descendants of Native Americans belonging to lineage Q1b1a1a-M3, the lack of high-resolution Y-SNPs inside this lineage causes most Native American Y chromosomes (61%–87%) to be characterized as Q1b1a1a-M3 [11, 25, 31–33], preventing the identification and geographic allocation of high-resolution haplogroups, which are the basis for interpreting migration patterns.

In this study, we aimed to further resolve the most diffused Native American founding lineage Q1b1a1a-M3 by reporting (novel) SNPs within this lineage and their phylogenetic relationships. Variant information was obtained by sequencing targeted regions representing 7.45 Mb of the Y chromosome in males from different regions in South America and complementing this dataset with published sequencing data of other Central and South American Y chromosomes belonging to haplogroup Q.

2. Material and Methods

2.1. Ethics Statement

All samples were collected under written informed consent and approved to be used in connection to this study by the following Ethics Committees: Comité de Ética de la Investigación en Seres Humanos (CEISH-UDLA) de la Universidad de Las Américas, Ecuador (2022-ENM-001); Comité de Ética do Hospital Universitário Pedro Ernesto/Universidade do Estado do Rio de Janeiro, Brazil (CAAE: 0067.0.228.000-09 and CAAE: 53563115.2.1001.5259); Comité de Ética da Universidade Federal do Pará, Brazil (43199815.9.0000.0018); Comité de Ética en Investigaciones Biomédicas del Instituto Multidisciplinario de Biología Celular (IMBICE), Argentina (CE00023); Comité de Ética de la Fundación Favaloro, Argentina; Comité de Ética de la Universidad Nacional de Colombia; and Comité de Ética en investigación Clínica del Instituto de Previsión Social, Paraguay (2015-03-25).

Initially, a total of 2755 blood samples were collected on FTA cards (Whatman Inc., Clifton, NJ, United States) with informed consent in connection with previous studies [29, 34–38]. The DNA extraction was conducted using the BioRobot EZ1 Workstation (Qiagen, Hilden, Germany), phenol-chloroform, or standard Chelex extraction methods following the manufacturer's recommendations. DNA concentrations were determined using the Qubit™ dsDNA High Sensitivity Assay and the Qubit® 3.0 Fluorometer (Invitrogen, Carlsbad, CA, United States) according to the manufacturer's protocols. The study was conducted according to the guidelines of the Declaration of Helsinki [39].

2.3. Sample Selection

Y-STR profiles of unrelated admixed men from different South American regions were typed in connection to other studies [29, 34–38, 40] using standard STR kits. These covered partly or completely the following 27 Y-STRs: DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS449, DYS456, DYS458, DYS460, DYS481, DYS518, DYS533, DYS549, DYS570, DYS576, DYS627, DYS635, DYS643, YGATAH4, and DYF387S1. The Y-DNA Haplogroup Predictor NevGen (Nevgen.org) was applied on the Y-STR profiles to determine the most likely haplogroup. For haplogroup predictions with low confidence, up to 59 Y-SNPs were investigated using PCR-SBE-CE as previously described [37, 41–45]. Ten of the targeted SNPs belonged to haplogroup Q: M242, P36.2, M346, M3, M19, Z19319, SA01, Z19483, M557, and SA05 [46].

Among the 2755 samples of individuals from South America, 405 individuals were identified as admixed individuals with Y chromosomes that belonged to the Native American haplogroup Q. Of the 405 individuals, 59 unrelated individuals were selected for this study due to their diverse Y-STR profiles and geographic origins. Admixed individuals with a native Y chromosome were selected in an effort to encompass most of the original genetic variation that over time has been lost in Native South Americans. The individuals originated from different regions in Argentina ($N = 24$), Bolivia ($N = 2$), Brazil ($N = 4$), Chile ($N = 2$), Colombia ($N = 17$), Ecuador ($N = 8$), and Paraguay ($N = 2$). The sample selection for MPS analysis was intended to have a broad distribution of samples from different geographic regions and a high diversity of Y-STR profiles within haplogroup Q to encompass a large variation. An overview of the collection sites is included in Figure S1. The sample distributions in scatter maps (Figure S1) were generated using the scattergeo function in the package plotly v4.14.3 in Python v3.6.8.

2.4. Publicly Available WGS Data

47–54] (shown in Figure S1 and Table S1). The 218 samples can be sorted in one of three categories of (i) modern admixed (ModAdmix) (*N* = 71), (ii) ancient indigenous (AncNAM) (*N* = 104), and (iii) modern indigenous (ModNAM) individuals (*N* = 43). The ancient samples are 11,000–400 years old, while the modern samples are taken from modern-day individuals.

For the data analysis, sequencing data in CRAM (or BAM) or FASTQ format was downloaded. CRAM files were decompressed to BAM files, chromosome labels were adjusted for consistency among all datasets, BAM files were indexed, and variants were called and annotated. In sequencing data of FASTQ format, adapter sequences were trimmed, and the reads were aligned to the reference genome GRCh37.p13. In the resulting BAM file, the reads were sorted, and duplicate reads were removed. After base quality recalibration, variants were called and annotated as before. More detailed information on the toolkits and commands is given in Figure S2.

2.5. Hybridization Capture–Based Target Enrichment

2.5.1. Probe Design

Custom RNA capture probes were designed with Agilent SureDesign's SureSelect DNA software (Agilent Technologies Inc., California, United States) to cover nine unique regions within the Y chromosome (Table S2) [6] to sequence the positions that potentially include phylogenetically relevant SNPs.

The final probe set comprised two groups of probes to cover the majority of the targeted regions: Both probe sets were designed with the 90 min hybridization option for the SureSelect XT HS2 protocol and optimized performance boosting, balancing the number of probes in GC-rich regions. The first set of probes was generated with default options in the remaining parameters, while the second set had a 1× tiling, determining the probe density or coverage of target nucleotides, and the least stringent masking of repetitive regions. The second set was designed to target regions that the first set failed to cover. Details on both probe sets can be viewed in Table S3. Both sets were combined and resulted in a total of 359,954 probes with an individual probe length of 120 bp and representing a total probe size of 7.84 Mb. Using these probes, approximately 7.45 Mb of the Y chromosome was targeted for sequencing (Figure S3).

2.5.2. Library Preparation and Sequencing

Libraries were prepared with a DNA input of 8.7–15 ng using Agilent's SureSelect XT HS2 Target Enrichment for Illumina Paired-End Sequencing Library kit (Agilent Technologies, California, United States) for the 59 samples (ModAdmix) following the manufacturer's protocol. The DNA

burst, 2 × 60 s treatment time, and bath temperature of 2°C–8°C. The resulting DNA fragments were end repaired, dA-tailed, and ligated with molecular barcodes (MBCs) for distinguishing PCR duplicates. After purification, the libraries were amplified with sample-specific index primer pairs. Purified libraries were then hybridized with the RNA probes, and complementary library molecules were captured, purified, and amplified.

After pooling 16 or 24 libraries to a concentration of 1.7–2 nM, the pools were paired-end sequenced on a flow cell of type SP using the NovaSeq 6000 SP Reagent Kit v1.5 (300 cycles) and the Illumina NovaSeq 6000 System (Illumina, California, United States).

2.5.3. Bioinformatic Processing of Targeted Sequencing MPS Data

Data analysis was done with a customized bioinformatic pipeline. Detailed information on the toolkits, commands, and settings is shown in Figure S4.

In this pipeline, the adapter and barcode sequences were removed, and MBCs were annotated in the FASTQ files. Reads were mapped to the reference genome GRCh37. After sorting the reads in the BAM files by coordinates and adding read group information, the BAM files from different sequencing lanes per sample were merged into one BAM file per sample. PCR deduplication was performed for a read group that shares a pair of MBCs, if there were at least two read pairs with this MBC pair (parameter "m2" in Figure S4). Two complementary single-strand consensus sequences were used to generate duplex consensus sequences. Single-strand consensus reads without a complementary partner and PCR duplicates were filtered out. Thereafter, the base quality scores were recalibrated.

For estimation of the capture specificity, on-target to off-target sequenced base ratios were calculated based on output data of the picard CollectHsMetrics tool. To smooth out fluctuations in the ratios, moving averages and variations were calculated using windows of five samples upon sorting them by increasing DNA concentration for the library preparation. The average DNA concentration for the window with the highest average variation was used to determine the concentration threshold for stable capture specificity.

For coverage analyses of the processed BAM files, the repositories samtools-1.17 [55] and bedtools-2.30.0 [56] were applied using the tools samtools depth (applying Mapping Quality 25) and bedtools genomecov.

2.6. Variant Calling and Reporting in All 277 Profiles

⟨ Back

and variants within the whole Y chromosome were called in VCF files using the tool bcftools call. The minimum base quality considered was 25, read depth of one, and ploidy of one. Already known variants of the gnomAD v.3.1 database, dbSNP database (accessed February 2022), the ISOGG Y-DNA Haplogroup Tree 2019–2020 and further publications [11, 47, 57] were annotated according to the following rules: SNPs taken from ISOGG or Yfull databases were annotated with the respective SNP names. SNPs present in the dbSNP database, the gnomAD database, or other publications were labelled with the location of the variation relative to the reference genome GRCh37 and the rs-number, the comment "gnomAD," or the SNP names given in the respective publications. Lastly, novel variants were simply annotated using the locations in the Y chromosome (GRCh37) and were assigned with unique identifiers (ZK1 to ZK1291) in Table S4. For more details on the bioinformatic pipeline, please see Figure S4.

The base positions of all variants called in the 59 profiles of this study and additional positions of haplogroup Q SNPs from the database ISOGG Y-DNA Haplogroup Tree 2019–2020 were combined and targeted in all 277 profiles using the software Yleaf v.2.2 [12].

The Yleaf position file, which states the location and possible alleles of all targeted SNPs, was changed to include the respective set of variants. Here, base calling thresholds (minimum of five reads for each base, quality threshold of 25 for each read, and base majority threshold 95%) were applied on the 59 previously investigated and 218 publicly available sequencing data. A total of 4128 variable positions were reported in all 277 samples.

To ascertain the extent of resolution increased by the 4128 variants, the haplogroups of the 277 sequences were additionally inferred using Yleaf v.3.1, which targets 129,151 SNPs from the most resolved phylogenetic tree publicly available (Yfull YTree v11.01.00).

2.6.1. Determining the Phylogenic Hierarchy of Variants Using SNPtotree

Due to the rather high rates of missing data, the software SNPtotree v.1.0 [58] was applied for an initial sorting of all 4128 variants into a phylogenetic tree using default settings. The input file contained the allelic status (ancestral or derived allele) of all 4128 variable sites in all 277 samples. The software filters out variants with contradictory phylogenetic relationships based on pairwise comparisons and variants with ambiguous positions in the phylogenetic tree.

Since the software does not take into account prior knowledge on the relationships of the SNPs, the generated data was supplemented with information from the Y-DNA Haplogroup Tree 2019–2020 from ISOGG and the Yfull YTree v11.01.00 to establish the most parsimonious variant hierarchy.

Tree 2019–2020. Although the more recently proposed nomenclature [60] based on mutation is simpler, it does not allow inferring relationships between subhaplogroups within a lineage. For example, for the Q-M346 and Q-L53 lineages named according to the mutation-based nomenclature, it is not possible to infer whether there is a close relationship between them. However, when using the adopted nomenclature, it is possible to see that Q1b1-L53 is a sublineage of Q1b-M346. Lineage-based nomenclature includes all the information that mutation-based nomenclature provides.

3. Results and Discussion

A total of 277 samples were investigated in this work, comprising 59 newly sequenced individuals and publicly available sequencing data from 218 individuals. The overall 277 South American samples of haplogroup Q originated from one of three categories of (i) ModAdmix (N = 130), (ii) AncNAM (N = 104), and (iii) ModNAM individuals (N = 43). All newly sequenced samples were of ModAdmix origin and were selected based on a broad distribution of different geographic regions and high diversity of Y-STR profiles. For these samples, 7.45 Mb of informative regions within the Y chromosome was sequenced in a targeted sequencing approach.

3.1. Sequencing Data Quality of the 59 Targeted Capture Samples

For the 59 samples sequenced in this study, MBCs were used during library preparation and utilized during the bioinformatic analysis to generate duplex consensus sequences (Figure S4). Since all reads of one orientation with the same MBC (and thus copies from the same original DNA fragment molecule) were compiled to one consensus sequence, the unit "read" of the hybridization capture–based target enrichment corresponds to an unknown, but usually higher number of actually sequenced reads. In the following, we refer to the term "read" for both MBC and non-MBC-generated reads for simplicity, even though this represents an underestimation for the MBC-based reads.

For each sample, the number of typed bases that successfully aligned to the reference genome GRCh37 ranged from 14 to 217 Mb (average: 115 Mb). The fraction of aligned bases that were assigned to the Y chromosome relative to the total number of sequenced bases was 0.72 (ranging from 0.21 to 0.83) when averaged over all 59 samples. The ratio of the number of on-target to off-target bases ranged from 0.11 to 1.3 per sample, with an average of 1.00.

Figure 1 presents the capture specificity based on on-target to off-target base ratios using moving averages for sample batches of five, when sorting samples according to DNA

coefficient = 0.48, *p* value = 0.0001) when applying the Pearson correlation analysis. For lower DNA concentrations (< 0.59 ng/μL), the capture specificity was highly variable, whereas it was more uniform for samples with higher concentrations.
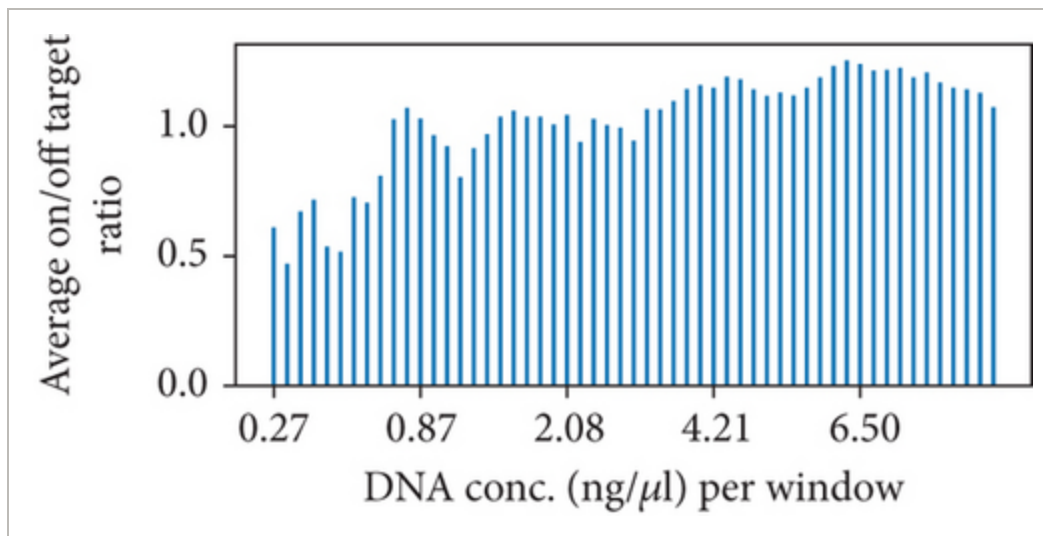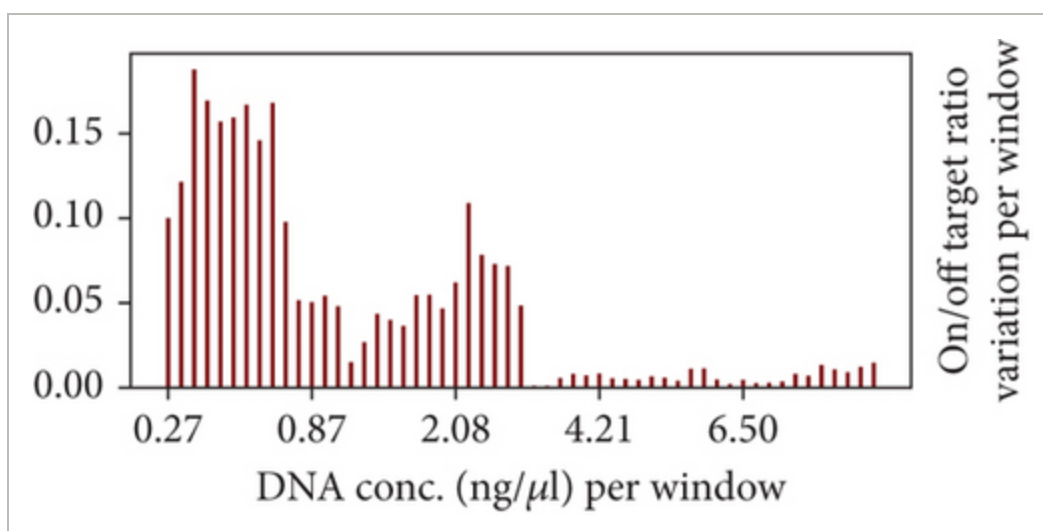

Figure 1 (a)

Open in figure viewer | ⬇**PowerPoint**

On-target to off-target ratios of sequenced bases for the 59 samples, which were sorted by increasing DNA concentrations for library preparation (abscissa, (a)), were averaged over windows of five samples for smoothing out fluctuations. Variations of on-target to off-target ratios among samples in each window (b) were highest for DNA concentrations below 0.59 ng/μL.

On-target to off-target ratios of sequenced bases for the 59 samples, which were sorted by increasing DNA concentrations for library preparation (abscissa, (a)), were averaged over windows of five samples for smoothing out fluctuations. Variations of on-target to off-target ratios among samples in each window (b) were highest for DNA concentrations below 0.59 ng/μL.

The nucleotide positions that were sequenced with read depth ≥ 1 and with base quality ≥ 25 had median read depths of 2× to 18× per sample and 10× considering all samples. The median read depth per targeted base position ranged from 0× to 389×, with the highest median sequencing depths found in the sequencing region from 28,457,993 to 28,806,758 (Figure S5).

3.2. Comparison of Variants Reported in the 59 Sequenced Samples and the Whole Dataset of 277 Sequences

When investigating the 59 samples sequenced in this study for Y-chromosomal variants, a total of 12,572 variants were observed in at least one of the samples with a minimum base quality of 25, ploidy of one, and minimum read depth of one. The majority of the variants, 11,768, were SNPs, and 804 variants were insertions and deletions (InDels) (Figure 2). Applying the same quality thresholds, 134,833 Y-chromosomal variants were reported when complementing the dataset of 59 sequences with 218 publicly available sequences. These 134,833 variants were composed of 127,762 SNPs and 7071 InDels.
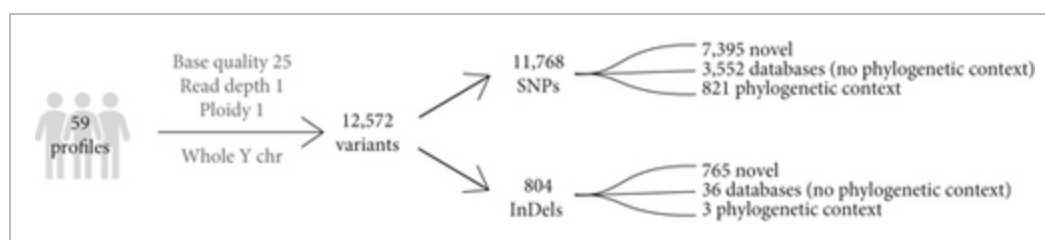

Figure 2 (a)

(a) Number of variants reported for the 59 samples sequenced in this study and (b) a combination of the 59 samples and 218 publicly available sequences using the same thresholds for variant calling.

< Back

Figure 2 (b)

Open in figure viewer ⬇PowerPoint

(a) Number of variants reported for the 59 samples sequenced in this study and (b) a combination of the 59 samples and 218 publicly available sequences using the same thresholds for variant calling.
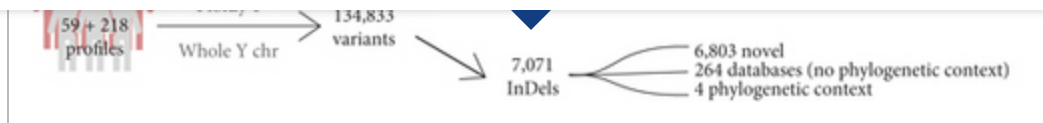
The 5-fold increase in the sample number from 59 to 277 surprisingly led to an 11-fold increase in reported variants. The variants were sorted according to the group of samples they were reported in ModAdmix, AncNAM, ModNAM, or (iv) any combination of the three groups (Figure 3). The subset of 59 sequences represented only ModAdmix samples, whereas the 218 publicly available sequences covered all three groups.
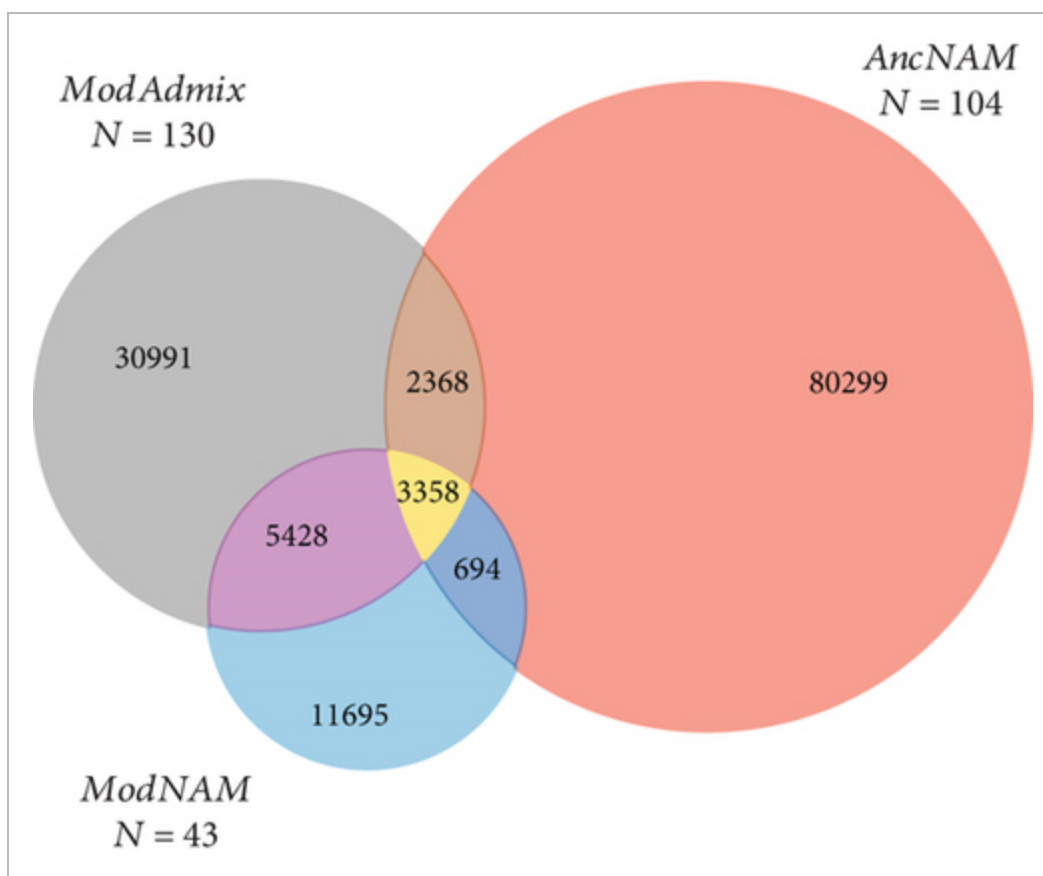

Figure 3

Open in figure viewer ⬇PowerPoint

numbers represent unique and shared variants among the three groups.

While most variants were exclusively found in one of the three sample groups, the biggest overlap of shared variants was reported between ModAdmix and ModNAM (7% of total variants). This is caused by the more similar evolutionary past, or a more recent common ancestor of current day admixed and indigenous individuals compared to the most recent common ancestor of the 11,000–400-year-old AncNAM individuals. In other words, the lineage of the AncNAM samples splits off earlier from the branch leading to the ModAdmix and ModNAM individuals. These two groups separated less than 400 years ago, and gene flow events between the groups may still be partially ongoing.

Generally, most of the variants (60%) were exclusively reported in AncNAM samples. This group was represented by only 38% of all analyzed samples and second in numbers to the ModAdmix (47%). It had a slightly wider geographic distribution from the Caribbean islands down to Tierra del Fuego in South Argentina, compared to the ModAdmix samples (Figure S1).

While a larger number of genetically diverse individuals highly impact the total number of reported variants, so does the number of sequenced bases. The sequences from the AncNAMs covered on average 1,188,675 bp (range 1878–13,382,470 bp) of the whole Y chromosome (with a mapping quality of 25). As expected, the sequences from modern samples covered more regions due to better sample conditions: The sequencing of ModAdmix samples covered on average 9,936,870 bp (range 2,764,105–17,631,605 bp), and the sequencing of ModNAM samples covered on average 13,158,176 bp (range 8,238,675–17,109,801 bp). Even though the AncNAM samples covered around ten times fewer bases than the modern samples, more variants were reported. The exclusively ancient SNPs were mostly (99%) without phylogenetic information, and the residual SNPs belonged to haplogroup Q or parallel haplogroups. These observations hinted towards sequencing errors fabricating a large portion of the called variants in the ancient samples. For this reason, we decided to use variants identified among the 59 sequenced samples of this study combined with well-characterized haplogroup Q SNPs from databases to refine the phylogenetic structure of haplogroup Q lineages.

3.3. Selection of 4128 Variants From 59 Targeted Sequencing Samples and Databases

The set of 11,768 SNPs identified in the 59 samples (Figure 2) was supplemented with more than 5000 SNPs from haplogroup Q (Y-DNA Haplogroup Tree 2019–2020 from ISOGG), which were already reported and validated in other studies (Figure 4). These positions were investigated in the 59 admixed samples and in the 218 publicly available sequences using

These settings helped to adjust for possible sequencing errors, particularly in the 218 publicly available sequences without MBCs. A total of 4128 polymorphic variants were ascertained in all 277 sequences after application of the more stringent criteria (Table S4). A total of 2267 out of 4128 variants (55%) were found in only one individual (singletons).

Figure 4

Open in figure viewer   │   ⬇PowerPoint

Selection of variants for phylogenetic investigations. The final set of 4128 positions was reported in all 277 sequences.

After cross-referencing the 4128 variants, 31% of the reported variants were, to our knowledge, not previously observed, whereas 23% were reported in databases without a phylogenetic context (only dbSNP and/or gnomAD databases), 45% were published with a phylogenetic context within haplogroup Q (Y-DNA Haplogroup Tree 2019–2020 from ISOGG, the Yfull YTree v11.01.00, and further publications [11, 28, 46, 47, 57]), and 1% were outside haplogroup Q (Figure 4). The latter were found in haplogroups parallel to lineage Q and thus characterized as recurrent. This is in accordance with a previous study that applied a similar sequence capture protocol targeting a smaller region of 3.7 Mb, where 0.92% (123 in 13,261 SNPs) of the SNPs were reported to be recurrent [19]. Studies with whole genome sequencing data resulted in a higher SNP recurrence rate (172/5865, 2.9%) [6], which may be explained by the sequencing of repetitive regions within the Y chromosome where the risk of recurrent mutations is higher [19].

3.4. Resolution of Haplogroup Q Phylogeny

Currently, one of the best resolved and updated phylogenetic trees of Y-SNPs can be found in the Yfull YTree v11.01.00, which includes more than 300,000 SNPs and 50,000 haplogroups (last accessed February 2023). An initial haplogroup inference of the 277 samples was based on 129,151 SNPs taken from the Yfull tree. Despite the high resolution, not all samples could be distinguished based on the defined haplogroups. For 82 sequences, haplogroup predictions failed due to data qualities below the selected thresholds: read depth ≥ 5, base quality ≥ 25, and base majority ≥ 95. For the remaining sequences, the inference of 14 different haplogroups

study were used, 214 samples were allocated to 135 different and more downstream lineages as represented in Table S5.

After combination with known phylogenetic relationships of Y-SNPs (Y-DNA Haplogroup Tree 2019–2020 from ISOGG and Yfull YTree v11.01.00), a comprehensive phylogenetic tree of the lineages represented by the 277 analyzed individuals was deducted using the software SNPtotree (available via **https://github.com/ZehraKoksal/SNPtotree**). The haplogroup Q tree is presented in Table S5, with 2236 out of 4128 variants incorporated into the tree. The residual variants were filtered out, due to contradictory relationships to other markers and ambiguous tree locations. Of the 2236 variants in the tree, 883 variants were novel.

The phylogenetic tree in Table S5 is to be read from the top left to the bottom right, where subbranches are presented in the cells on the lower right cell of the respective branch. The phylogenetic tree was built on the foundation of the already established haplogroup Q SNPs and additional SNPs reported in the 59 admixed samples typed in this study. Depending on where the branch-defining SNPs have been retrieved from, the branches were highlighted in different colors (see Table S5).

Out of all 104 AncNAM samples, 38 were assigned to nine mostly upstream branches in the phylogenetic tree, which is based on SNPs taken from Y-DNA Haplogroup Tree 2019–2020 from ISOGG and ModAdmix. Although ModAdmix and AncNAM share 5726 SNPs (Figure 3), only six branches are represented by samples from both groups. A portion of the 5726 SNPs may have been lost while building the phylogenetic tree due to the increased sequencing quality thresholds and contradictory information regarding relationships between SNPs. Further, the SNPs in the phylogenetic tree, which were taken from databases, seem to not include many ancient SNPs.

Three of the nine branches are characterized by SNPs exclusively found in eight of the ancient samples. The exclusively ancient SNPs may have been lost in modern individuals over time, but this demands further research. Some SNPs found in ancient samples are located at branch tips downstream of variants found in ModAdmix samples. Thus, the upstream variants may characterize very old lineages of common ancestors of the ancient and modern samples.

Variants found in ModNAM exclusively or together with ModAdmix were presented in more intertwined tree branches. This observation is assured by a more similar evolution or more recent separation between the branches leading to current day admixed and indigenous individuals, as opposed to modern and AncNAM samples. While the ancient samples are

Here, we were able to reconstruct and refine the phylogenetic structure of around 300 branches composed of 3400 variants within haplogroup Q lineages, in particular, the Q-M3 lineage. Furthermore, it was possible to identify 117 previously undescribed subhaplogroups. However, for 17 of these branches, the local tree topology was not entirely resolved, due to an incomplete overlap of genetic information retrieved from different sources. Three main branches directly downstream of lineage Q1 and 15 branches directly downstream of lineage Q1b were reported. This counteracts the resolution of polytomies and may be the result of unexplored variants and missing data. Thus, further analysis of the given variants in a new set of samples may help to refine the tree structure and resolve polytomies.

4. Conclusion

The application of targeted capture MPS enabled the targeted sequencing of 7.45 Mb within the Y chromosome. An initial screening for variants in all 277 samples resulted in 134,833 different variants. The majority of these were reported in indigenous and ancient samples possibly due to loss of ancient SNPs and sequencing errors of the ancient samples. Furthermore, modern-day indigenous and admixed individuals had the biggest overlap of shared variants, most likely due to a more recent evolutionary past. The samples of ancient origin diverged earlier from the branch leading to both groups of modern age. Applying strict genotyping criteria, 4128 variants were reported in the 277 samples. With the help of the SNP sorting software SNPtotree and previously established phylogenetic relationships of SNPs, 3400 variants and 300 branches were compiled into a comprehensive phylogenetic tree. The established phylogenetic tree of the Native American haplogroup Q lineage provides the highest resolution reported so far. In the future, the reported variants should be investigated in more individuals, which may help resolve some branches even further and may reflect characteristic geographic patterns of the reported variants.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

1/8/24, 10:24

Application of Targeted Y-Chromosomal Capture Enrichment to Increase the Resolution of Native American Haplogroup Q - Köksal - 20...

⟨ Back

Open Research ⌄

Data Availability Statement

All relevant variant data are included in the published article and the supporting information.

Supporting Information ⌄

Filename	Description
humu3046495-sup-0001-f1.zip application/x-compressed, 2.2 MB	**Supporting Information** Additional supporting information can be found online in the Supporting Information section. Figure S1. Map of the origin of all 277 haplogroup Q samples from Central and South America included in the study. Modern admixed (ModAdmix, black), ancient indigenous (AncNAM, red), and modern indigenous (ModNAM, blue) samples are presented at the place of their origin. The point size is proportional to the number of samples. Square markers indicate 59 samples genotyped in this study; round markers indicate 218 samples from public databases. Figure S2. Linux command-based bioinformatic pipeline for retrieval, preparation, and variant calling of publicly available sequencing data in the CRAM (or BAM) or FASTQ format. Figure S3. (a) Nine unique regions within the Y chromosome [6]; (b) targeted regions in this work using RNA baits. Figure S4. Bioinformatic pipeline for targeted capture data from SureSelect XT HS2 with the commands on the left side and visual representation of the commands' effects of the respective file on the right side. Figure S5. Median and cumulative read depth of the 59 samples within the nine targeted regions (orange bins). Table S1. References of the 277 samples of ModAdmix, AncNAM, or ModNAM origin. Table S2. Nine unique regions within the Y chromosome targeted in this work. Table S3. Region sizes, number of

< Back

probes, and total probe size of the two designed probe groups. Table S4. Detailed information on the 4128 variants reported and targeted in this study. Table S5. Phylogenetic hierarchy of Y-chromosomal variants from the Native

References   ⌄

1   Jobling M. A. and Tyler-Smith C., Fathers and sons: the Y chromosome and human evolution, *Trends in Genetics*. (1995) **11**, no. 11, 449–456, https://doi.org/10.1016/s0168-9525(00)89144-1, 2-s2.0-0028892563, 8578602.

    CAS | **PubMed** | **Web of Science®** | **Google Scholar**

2   Kayser M., Forensic use of Y-chromosome DNA: a general overview, *Human Genetics*. (2017) **136**, no. 5, 621–635, https://doi.org/10.1007/s00439-017-1776-9, 2-s2.0-85015663393, 28315050.

    CAS | **PubMed** | **Web of Science®** | **Google Scholar**

3   Sun L., Wang Z., Lu T., Manolio T. A., and Paterson A. D., eXclusionarY: 10 years later, where are the sex chromosomes in GWASs?, *The American Journal of Human Genetics*. (2023) **110**, no. 6, 903–912, https://doi.org/10.1016/j.ajhg.2023.04.009, 37267899.

    CAS | **PubMed** | **Google Scholar**

4   Hallast P., Ebert P., Loftus M., Yilmaz F., Audano P. A., Logsdon G. A., Bonder M. J., Zhou W., Höps W., Kim K., Li C., Dishuck P., Porubsky D., Tsetsos F., Kwon J. Y., Zhu Q., Munson K. M., Hasenfeld P., Harvey W. T., Lewis A. P., Kordosky J., Hoekzema K., Consortium (HGSVC) THGSV, Korbel J. O., Tyler-Smith C., Eichler E. E., Shi X., Beck C. R., Marschall T., Konkel M. K., and Lee C., *Assembly of 43 diverse human Y chromosomes reveals extensive complexity and variation*, 2022, bioRxiv, https://doi.org/10.1101/2022.12.01.518658.

    Google Scholar

5   Skaletsky H., Kuroda-Kawaguchi T., Minx P. J., Cordum H. S., Hillier L., Brown L. G., Repping S., Pyntikova T., Ali J., Bieri T., Chinwalla A., Delehaunty A., Delehaunty K., Du H., Fewell G., Fulton L., Fulton R., Graves T., Hou S.-F., Latrielle P., Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W., Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K., Schultz B., Strong C., Tin-Wollam A., Yang S.-P., Waterston R. H., Wilson R. K., Rozen S., and Page D. C., The male-specific region of the human Y

⟨ Back

CAS | PubMed | Web of Science® | Google Scholar |

---

6   Wei W., Ayub Q., Chen Y., McCarthy S., Hou Y., Carbone I., Xue Y., and Tyler-Smith C., A calibrated human Y-chromosomal phylogeny based on resequencing, *Genome Research*. (2013) **23**, no. 2, 388–395, https://doi.org/10.1101/gr.143198.112, 2-s2.0-84868692234, 23038768.

CAS | PubMed | Web of Science® | Google Scholar |

---

7   Fondevila M., Børsting C., Phillips C., de la Puente M., Consortium E.-N., Carracedo A., Morling N., and Lareu M. V., Forensic SNP genotyping with SNaPshot: technical considerations for the development and optimization of multiplexed SNP assays, *Forensic Science Review*. (2017) **29**, no. 1, 57–76, 28119267.

CAS | PubMed | Google Scholar |

---

8   Inagaki S., Yamamoto Y., Doi Y., Takata T., Ishikawa T., Yoshitome K., Miyaishi S., and Ishizu H., Typing of Y chromosome single nucleotide polymorphisms in a Japanese population by a multiplexed single nucleotide primer extension reaction, *Legal Medicine (Tokyo, Japan)*. (2002) **4**, no. 3, 202–206, https://doi.org/10.1016/s1344-6223(02)00032-9, 2-s2.0-0036734358, 12935671.

CAS | PubMed | Google Scholar |

---

9   Sanchez J. J., Børsting C., Hallenberg C., Buchard A., Hernandez A., and Morling N., Multiplex PCR and minisequencing of SNPs--a model with 35 Y chromosome SNPs, *Forensic Science International*. (2003) **137**, no. 1, 74–84, https://doi.org/10.1016/s0379-0738(03)00299-8, 2-s2.0-0141533239, 14550618.

CAS | PubMed | Web of Science® | Google Scholar |

---

10   Geppert M., Ayub Q., Xue Y., Santos S., Ribeiro-dos-Santos Â., Baeta M., Núñez C., Martínez-Jarreta B., Tyler-Smith C., and Roewer L., Identification of new SNPs in native South American populations by resequencing the Y chromosome, *Forensic Science International: Genetics*. (2015) **15**, 111–114, https://doi.org/10.1016/j.fsigen.2014.09.014, 2-s2.0-84921423367, 25303787.

CAS | PubMed | Web of Science® | Google Scholar |

---

11   Köksal Z., Burgos G., Carvalho E., Loiola S., Parolin M. L., Quiroz A., Ribeiro Dos Santos Â., Toscanini U., Vullo C., Børsting C., Gusmão L., and Pereira V., Testing the Ion AmpliSeq™ HID Y-SNP Research Panel v1 for performance and resolution in admixed South Americans of haplogroup Q, *Forensic Science*

‹ Back

PubMed | Google Scholar

12　Ralf A., van Oven M., Montiel González D., de Knijff P., van der Beek K., Wootton S., Lagacé R., and Kayser M., 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*. (2019) **41**, 93–106, https://doi.org/10.1016/j.fsigen.2019.04.001, 2-s2.0-85065045160, 31063905.

CAS | PubMed | Web of Science® | Google Scholar

13　Claerhout S., Verstraete P., Warnez L., Vanpaemel S., Larmuseau M., and Decorte R., CSYseq: the first Y-chromosome sequencing tool typing a large number of Y-SNPs and Y-STRs to unravel worldwide human population genetics, *PLoS Genetics*. (2021) **17**, no. 9, article e1009758, https://doi.org/10.1371/journal.pgen.1009758.

PubMed | Google Scholar

14　Pereira V., Santangelo R., Børsting C., Tvedebrink T., Almeida A. P. F., Carvalho E. F., Morling N., and Gusmão L., Evaluation of the precision of ancestry inferences in South American admixed populations, *Frontiers in Genetics*. (2020) **11**, https://doi.org/10.3389/fgene.2020.00966, 32973885.

PubMed | Google Scholar

15　Truelsen D., Tvedebrink T., Mogensen H., Sharafi Farzad M., Shan M. A., Morling N., Pereira V., and Børsting C., Assessment of the effectiveness of the EUROFORGEN NAME and precision ID ancestry panel markers for ancestry investigations, *Scientific Reports*. (2021) **11**, no. 1, https://doi.org/10.1038/s41598-021-97654-0, 34545122.

PubMed | Google Scholar

16　Hofreiter M., Sneberger J., Pospisek M., and Vanek D., Progress in forensic bone DNA analysis: lessons learned from ancient DNA, *Forensic Science International: Genetics*. (2021) **54, article 102538**, https://doi.org/10.1016/j.fsigen.2021.102538, 34265517.

PubMed | Google Scholar

17　Yan S., Wang C.-C., Zheng H.-X., Wang W., Qin Z.-D., Wei L.-H., Wang Y., Pan X.-D., Fu W.-Q., He Y.-G., Xiong L.-J., Jin W.-F., Li S.-L., An Y., Li H., and Jin L., Y chromosomes of 40% Chinese descend from three

1/8/24, 10:24   Application of Targeted Y-Chromosomal Capture Enrichment to Increase the Resolution of Native American Haplogroup Q - Köksal - 20…

‹   Back

PubMed | **Google Scholar**

---

18   Rohrlach A. B., Papac L., Childebayeva A., Rivollat M., Villalba-Mouco V., Neumann G. U., Penske S., Skourtanioti E., van de Loosdrecht M., Akar M., Boyadzhiev K., Boyadzhiev Y., Deguilloux M.-F., Dobeš M., Erdal Y. S., Ernée M., Frangipane M., Furmanek M., Friederich S., Ghesquière E., Hałuszko A., Hansen S., Küßner M., Mannino M., Özbal R., Reinhold S., Rottier S., Salazar-García D. C., Diaz J. S., Stockhammer P. W., de Togores Muñoz C. R., Yener K. A., Posth C., Krause J., Herbig A., and Haak W., Using Y-chromosome capture enrichment to resolve haplogroup H2 shows new evidence for a two-path Neolithic expansion to Western Europe, *Scientific Reports*. (2021) **11**, no. 1, **https://doi.org/10.1038/s41598-021-94491-z**, **34294811**.

PubMed | **Google Scholar**

---

19   Hallast P., Batini C., Zadik D., Maisano Delser P., Wetton J. H., Arroyo-Pardo E., Cavalleri G. L., de Knijff P., Destro Bisol G., Dupuy B. M., Eriksen H. A., Jorde L. B., King T. E., Larmuseau M. H., López de Munain A., López-Parra A. M., Loutradis A., Milasin J., Novelletto A., Pamjav H., Sajantila A., Schempp W., Sears M., Tolun A., Tyler-Smith C., Van Geystelen A., Watkins S., Winney B., and Jobling M. A., The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades, *Molecular Biology and Evolution*. (2015) **32**, no. 3, 661–673, **https://doi.org/10.1093/molbev/msu327**, 2-s2.0-84927738893, **25468874**.

CAS | **PubMed** | **Web of Science®** | **Google Scholar**

---

20   Cruz-Dávalos D. I., Nieves-Colón M. A., Sockell A., Poznik G. D., Schroeder H., Stone A. C., Bustamante C. D., Malaspinas A.-S., and Ávila-Arcos M. C., In-solution Y-chromosome capture-enrichment on ancient DNA libraries, *BMC Genomics*. (2018) **19**, no. 1, **https://doi.org/10.1186/s12864-018-4945-x**, 2-s2.0-85051558996, **30107783**.

PubMed | **Google Scholar**

---

21   Pilli E., Morelli S., Poggiali B., and Alladio E., Biogeographical ancestry, variable selection, and PLS-DA method: a new panel to assess ancestry in forensic samples via MPS technology, *Forensic Science International: Genetics*. (2023) **62, article 102806**, **https://doi.org/10.1016/j.fsigen.2022.102806**, **36399972**.

PubMed | **Google Scholar**

---

22   Karmin M., Saag L., Vicente M., Sayres M. A. W., Järve M., Talas U. G., Rootsi S., Ilumäe A.-M., Mägi R., Mitt M., Pagani L., Puurand T., Faltyskova Z., Clemente F., Cardona A., Metspalu E., Sahakyan H.,

Lachance J., Tishkoff S., Momynaliev K., Ricaut F.-X., Kusuma P., Razafindrazaka H., Pierron D., Cox M. P., Sultana G. N. N., Willerslev R., Muller C., Westaway M., Lambert D., Skaro V., Kovačevic L., Turdikulova S., Dalimova D., Khusainova R., Trofimova N., Akhmetova V., Khidiyatova I., Lichman D. V., Isakova J., Pocheshkhova E., Sabitov Z., Barashkov N. A., Nymadawa P., Mihailov E., Seng J. W. T., Evseeva I., Migliano A. B., Abdullah S., Andriadze G., Primorac D., Atramentova L., Utevska O., Yepiskoposyan L., Marjanovic D., Kushniarevich A., Behar D. M., Gilissen C., Vissers L., Veltman J. A., Balanovska E., Derenko M., Malyarchuk B., Metspalu A., Fedorova S., Eriksson A., Manica A., Mendez F. L., Karafet T. M., Veeramah K. R., Bradman N., Hammer M. F., Osipova L. P., Balanovsky O., Khusnutdinova E. K., Johnsen K., Remm M., Thomas M. G., Tyler-Smith C., Underhill P. A., Willerslev E., Nielsen R., Metspalu M., Villems R., and Kivisild T., A recent bottleneck of Y chromosome diversity coincides with a global change in culture, *Genome Research*. (2015) **25**, no. 4, 459–466, https://doi.org/10.1101/gr.186684.114, 2-s2.0-84927713385, 25770088.

CAS | PubMed | Web of Science® | Google Scholar

23 Bortolini M.-C., Salzano F. M., Thomas M. G., Stuart S., Nasanen S. P. K., Bau C. H. D., Hutz M. H., Layrisse Z., Petzl-Erler M. L., Tsuneto L. T., Hill K., Hurtado A. M., Castro-de-Guerra D., Torres M. M., Groot H., Michalski R., Nymadawa P., Bedoya G., Bradman N., Labuda D., and Ruiz-Linares A., Y-chromosome evidence for differing ancient demographic histories in the Americas, *American Journal of Human Genetics*. (2003) **73**, no. 3, 524–539, https://doi.org/10.1086/377588, 2-s2.0-0042763545, 12900798.

CAS | PubMed | Web of Science® | Google Scholar

24 Geppert M., Baeta M., Núñez C., Martínez-Jarreta B., Zweynert S., Cruz O. W. V., González-Andrade F., González-Solorzano J., Nagy M., and Roewer L., Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America, *Forensic Science International: Genetics*. (2011) **5**, no. 2, 100–104, https://doi.org/10.1016/j.fsigen.2010.08.016, 2-s2.0-79952706056, 20932815.

CAS | PubMed | Web of Science® | Google Scholar

25 Zegura S. L., Karafet T. M., Zhivotovsky L. A., and Hammer M. F., High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas, *Molecular Biology and Evolution*. (2004) **21**, no. 1, 164–175, https://doi.org/10.1093/molbev/msh009, 2-s2.0-1242284634, 14595095.

CAS | PubMed | Web of Science® | Google Scholar

26 Seielstad M., Yuldasheva N., Singh N., Underhill P., Oefner P., Shen P., and Wells R. S., A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas, *American Journal*

1/8/24, 10:24   Application of Targeted Y-Chromosomal Capture Enrichment to Increase the Resolution of Native American Haplogroup Q - Köksal - 20…

⟨ Back

| CAS | PubMed | Web of Science® | Google Scholar |

27   Grugni V., Raveane A., Ongaro L., Battaglia V., Trombetta B., Colombo G., Capodiferro M. R., Olivieri A., Achilli A., Perego U. A., Motta J., Tribaldos M., Woodward S. R., Ferretti L., Cruciani F., Torroni A., and Semino O., Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas, *BMC Biology*. (2019) **17**, no. 1, https://doi.org/10.1186/s12915-018-0622-4, 2-s2.0-85060378188, 30674303.

| PubMed | Google Scholar |

28   Pinotti T., Bergström A., Geppert M., Bawn M., Ohasi D., Shi W., Lacerda D. R., Solli A., Norstedt J., Reed K., Dawtry K., González-Andrade F., Paz-y-Miño C., Revollo S., Cuellar C., Jota M. S., Santos J. E., Ayub Q., Kivisild T., Sandoval J. R., Fujita R., Xue Y., Roewer L., Santos F. R., and Tyler-Smith C., Y Chromosome sequences reveal a short Beringian standstill, rapid expansion, and early population structure of Native American founders, *Current Biology*. (2019) **29**, no. 1, 149–157.e3, https://doi.org/10.1016/j.cub.2018.11.029, 2-s2.0-85059608002, 30581024.

| CAS | PubMed | Web of Science® | Google Scholar |

29   Roewer L., Nothnagel M., Gusmão L., Gomes V., González M., Corach D., Sala A., Alechine E., Palha T., Santos N., Ribeiro-Dos-Santos A., Geppert M., Willuweit S., Nagy M., Zweynert S., Baeta M., Núñez C., Martínez-Jarreta B., González-Andrade F., Fagundes de Carvalho E., da Silva D. A., Builes J. J., Turbón D., Lopez Parra A. M., Arroyo-Pardo E., Toscanini U., Borjas L., Barletta C., Ewart E., Santos S., and Krawczak M., Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans, *PLOS Genetics*. (2013) **9**, no. 4, article e1003460, https://doi.org/10.1371/journal.pgen.1003460, 2-s2.0-84876864572, 23593040.

| PubMed | Google Scholar |

30   Mezzavilla M., Geppert M., Tyler-Smith C., Roewer L., and Xue Y., Insights into the origin of rare haplogroup C3∗ Y chromosomes in South America from high-density autosomal SNP genotyping, *Forensic Science International: Genetics*. (2015) **15**, 115–120, https://doi.org/10.1016/j.fsigen.2014.11.005, 2-s2.0-84921346073, 25435155.

| CAS | PubMed | Web of Science® | Google Scholar |

31   Lell J. T., Brown M. D., Schurr T. G., Sukernik R. I., Starikovskaya Y. B., Torroni A., Moore L. G., Troup G. M., and Wallace D. C., Y chromosome polymorphisms in native American and Siberian populations:

1/8/24, 10:24 Application of Targeted Y-Chromosomal Capture Enrichment to Increase the Resolution of Native American Haplogroup Q - Köksal - 20…

<   Back

CAS | PubMed | Web of Science® | Google Scholar |

32 Lell J. T., Sukernik R. I., Starikovskaya Y. B., Su B., Jin L., Schurr T. G., Underhill P. A., and Wallace D. C., The dual origin and Siberian affinities of Native American Y chromosomes, *American Journal of Human Genetics*. (2002) **70**, no. 1, 192–206, https://doi.org/10.1086/338457, 2-s2.0-0036137029, 11731934.

CAS | PubMed | Web of Science® | Google Scholar |

33 Karafet T., Zegura S. L., Vuturo-Brady J., Posukh O., Osipova L., Wiebe V., Romero F., Long J. C., Harihara S., Jin F., Dashnyam B., Gerelsaikhan T., Omoto K., and Hammer M. F., Y chromosome markers and trans-Bering Strait dispersals, *American Journal of Physical Anthropology*. (1997) **102**, no. 3, 301–314, https://doi.org/10.1002/(SICI)1096-8644(199703)102:3%3C301::AID-AJPA1%3E3.0.CO;2-Y.

CAS | PubMed | Web of Science® | Google Scholar |

34 Jannuzzi J., Ribeiro J., Alho C., de Oliveira Lázaro e Arão G., Cicarelli R., Simões Dutra Corrêa H., Ferreira S., Fridman C., Gomes V., Loiola S., da Mota M. F., Ribeiro-dos-Santos Â., de Souza C. A., de Sousa Azulay R. S., Carvalho E. F., and Gusmão L., Male lineages in Brazilian populations and performance of haplogroup prediction tools, *Forensic Science International: Genetics*. (2020) **44, article 102163**, https://doi.org/10.1016/j.fsigen.2019.102163, 31704485.

PubMed | Google Scholar |

35 Toscanini U., Brisighelli F., Llull C., Berardi G., Gómez A., Andreatta F., Pardo-Seco J., Gómez-Carballa A., Martinón-Torres F., Álvarez-Iglesias V., and Salas A., Charting the Y-chromosome ancestry of present-day Argentinean Mennonites, *Journal of Human Genetics*. (2016) **61**, no. 6, 507–513, https://doi.org/10.1038/jhg.2016.3, 2-s2.0-84976604170, 26841831.

CAS | PubMed | Web of Science® | Google Scholar |

36 Ambrosio I. B., Braganholi D. F., Orlando L. B. M., Andrekenas N. C., da Mota P. I., da Silva D. A., Astolfi-Filho S., de Carvalho E. F., Cicarelli R. M. B., and Gusmão L., Mutational data and population profiling of 23 Y-STRs in three Brazilian populations, *Forensic Science International: Genetics*. (2020) **48, article 102348**, https://doi.org/10.1016/j.fsigen.2020.102348, 32707472.

PubMed | Google Scholar |

Association between Y haplogroups and autosomal aims reveals intra-population substructure in Bolivian populations, *International Journal of Legal Medicine*. (2015) **129**, no. 4, 673–680, https://doi.org/10.1007/s00414-014-1025-x, 2-s2.0-84931563246, 24878616.

PubMed │ Web of Science® │ Google Scholar

38   Simão F., Ribeiro J., Vullo C., Catelli M., Gomes V., Xavier C., Huber G., Bodner M., Quiroz A., Ferreira A., Carvalho E., Parson W., and Gusmão L., The ancestry of eastern Paraguay: a typical south American profile with a unique pattern of admixture, *Genes*. (2021) **12**, no. 11, https://doi.org/10.3390/genes12111788.

PubMed │ Google Scholar

39   World Medical Association, Declaration of Helsinki. Ethical principles for medical research involving human subjects, *Jahrbuch Für Wissenschaft Und Ethik*. (2009) **14**, 233–238, https://doi.org/10.1515/9783110208856.233.

Google Scholar

40   Olofsson J. K., Pereira V., Børsting C., and Morling N., Peopling of the north circumpolar region–insights from Y chromosome STR and SNP typing of Greenlanders, *PLoS One*. (2015) **10**, no. 1, article e0116573, https://doi.org/10.1371/journal.pone.0116573, 2-s2.0-84922154849, 25635810.

PubMed │ Google Scholar

41   Brion M., Sobrino B., Blanco-Verea A., Lareu M. V., and Carracedo A., Hierarchical analysis of 30 Y-chromosome SNPs in European populations, *International Journal of Legal Medicine*. (2005) **119**, no. 1, 10–15, https://doi.org/10.1007/s00414-004-0439-2, 2-s2.0-12444293829, 15095093.

CAS │ PubMed │ Web of Science® │ Google Scholar

42   van Oven M., Toscani K., van den Tempel N., Ralf A., and Kayser M., Multiplex genotyping assays for fine-resolution subtyping of the major human Y-chromosome haplogroups E, G, I, J, and R in anthropological, genealogical, and forensic investigations, *Electrophoresis*. (2013) **34**, no. 20-21, 3029–3038, https://doi.org/10.1002/elps.201300210, 2-s2.0-84887401536, 23893838.

PubMed │ Google Scholar

43   Gomes V., Sánchez-Diz P., Amorim A., Carracedo A., and Gusmão L., Digging deeper into East African human Y chromosome lineages, *Human Genetics*. (2010) **127**, no. 5, 603–613, https://doi.org/10.

‹ Back

44   Resque R., Gusmão L., Geppert M., Roewer L., Palha T., Alvarez L., Ribeiro-dos-Santos Â., and Santos S., Male lineages in Brazil: intercontinental admixture and stratification of the European background, *PLoS One*. (2016) **11**, no. 4, article e0152573, https://doi.org/10.1371/journal.pone.0152573, 2-s2.0-84962841543.

PubMed │ Google Scholar

45   González M., Gomes V., López-Parra A. M., Amorim A., Carracedo A., Sánchez-Diz P., Arroyo-Pardo E., and Gusmão L., The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88, *European Journal of Human Genetics*. (2013) **21**, no. 3, 324–331, https://doi.org/10.1038/ejhg.2012.167, 2-s2.0-84874118100, 22892526.

CAS │ PubMed │ Web of Science® │ Google Scholar

46   Jota M. S., Lacerda D. R., Sandoval J. R., Vieira P. P. R., Ohasi D., Santos-Jnior J. E., Acosta O., Cuellar C., Revollo S., Paz-y-Mi'o C., Fujita R., Vallejo G. A., Schurr T. G., Tarazona-Santos E. M., Pena S. D. J., Ayub Q., Tyler-Smith C., and Santos F. R., New native South American Y chromosome lineages, *Journal of Human Genetics*. (2016) **61**, no. 7, 593–603, https://doi.org/10.1038/jhg.2016.26, 2-s2.0-84979633665.

PubMed │ Web of Science® │ Google Scholar

47   Paz Sepúlveda P. B., Mayordomo A. C., Sala C., Sosa E. J., Zaiat J. J., Cuello M., Schwab M., Rodríguez Golpe D., Aquilano E., Santos M. R., Dipierri J. E., Alfaro Gómez E. L., Bravi C. M., Muzzio M., and Bailliet G., Human Y chromosome sequences from Q haplogroup reveal a South American settlement pre-18,000 years ago and a profound genomic impact during the Younger Dryas, *PLoS One*. (2022) **17**, no. 8, article e0271971, https://doi.org/10.1371/journal.pone.0271971, 35976870.

PubMed │ Google Scholar

48   Popović D., Molak M., Ziółkowski M., Vranich A., Sobczyk M., Vidaurre D. U., Agresti G., Skrzypczak M., Ginalski K., Lamnidis T. C., Nakatsuka N., Mallick S., and Baca M., Ancient genomes reveal long-range influence of the pre-Columbian culture and site of Tiwanaku, *Science Advances*. (2021) **7**, no. 39, article eabg7261, https://doi.org/10.1126/sciadv.abg7261, 34559567.

PubMed │ Google Scholar

⟨ Back

…, …, Blanche M., Delozie J. F., Garh M., Ma… …B., Reich D., Sahara M. B., Skoglund M., Sealy M.,
Xue Y., Durbin R., and Tyler-Smith C., Insights into human genetic variation and population history from
929 diverse genomes, *Science*. (2020) **367**, no. 6484, article eaay5012, **https://doi.org/10.1126/science.a**
ay5012, **32193295**.

| **PubMed** | **Google Scholar** |

50   Nakatsuka N., Lazaridis I., Barbieri C., Skoglund P., Rohland N., Mallick S., Posth C., Harkins-Kinkaid
K., Ferry M., Harney É., Michel M., Stewardson K., Novak-Forst J., Capriles J. M., Durruty M. A., Álvarez K.
A., Beresford-Jones D., Burger R., Cadwallader L., Fujita R., Isla J., Lau G., Aguirre C. L., Le Blanc S.,
Maldonado S. C., Meddens F., Messineo P. G., Culleton B. J., Harper T. K., Quilter J., Politis G., Rademaker
K., Reindel M., Rivera M., Salazar L., Sandoval J. R., Santoro C. M., Scheifler N., Standen V., Barreto M. I.,
Espinoza I. F., Tomasto-Cagigao E., Valverde G., Kennett D. J., Cooper A., Krause J., Haak W., Llamas B.,
Reich D., and Fehren-Schmitz L., A paleogenomic reconstruction of the deep population history of the
Andes, *Cell*. (2020) **181**, no. 5, 1131–1145.e21, **https://doi.org/10.1016/j.cell.2020.04.015**, **32386546**.

| **CAS** | **PubMed** | **Web of Science®** | **Google Scholar** |

51   Nakatsuka N., Luisi P., Motti J. M. B., Salemme M., Santiago F., D'Angelo del Campo M., Vecchi R. J.,
Espinosa-Parrilla Y., Prieto A., Adamski N., Lawson A. M., Harper T. K., Culleton B. J., Kennett D. J.,
Lalueza-Fox C., Mallick S., Rohland N., Guichón R. A., Cabana G. S., Nores R., and Reich D., Ancient
genomes in South Patagonia reveal population movements associated with technological shifts and
geography, *Nature Communications*. (2020) **11**, no. 1, **https://doi.org/10.1038/s41467-020-17656-w**,
32747648.

| **PubMed** | **Google Scholar** |

52   Moreno-Mayar J. V., Vinner L., de Barros D. P., de la Fuente C., Chan J., Spence J. P., Allentoft M. E.,
Vimala T., Racimo F., Pinotti T., Rasmussen S., Margaryan A., Iraeta Orbegozo M., Mylopotamitaki D.,
Wooller M., Bataille C., Becerra-Valdivia L., Chivall D., Comeskey D., Devièse T., Grayson D. K., George L.,
Harry H., Alexandersen V., Primeau C., Erlandson J., Rodrigues-Carvalho C., Reis S., Bastos M. Q. R.,
Cybulski J., Vullo C., Morello F., Vilar M., Wells S., Gregersen K., Hansen K. L., Lynnerup N., Mirazón Lahr
M., Kjær K., Strauss A., Alfonso-Durruty M., Salas A., Schroeder H., Higham T., Malhi R. S., Rasic J. T.,
Souza L., Santos F. R., Malaspinas A.-S., Sikora M., Nielsen R., Song Y. S., Meltzer D. J., and Willerslev E.,
Early human dispersals within the Americas, *Science*. (2018) **362**, no. 6419, article eaav2621, **https://doi.**
org/10.1126/science.aav2621, 2-s2.0-85058604407.

| **PubMed** | **Google Scholar** |

53   Fernandes D. M., Sirak K. A., Ringbauer H., Sedig J., Rohland N., Cheronet O., Mah M., Mallick S.,
Olalde I., Culleton B. J., Adamski N., Bernardos R., Bravo G., Broomandkhoshbacht N., Callan K., Candilio
F., Demetz L., Carlson K. S. D., Eccles L., Freilich S., George R. J., Lawson A. M., Mandl K., Marzaioli F.,

‹   Back

M. V., Marcheco-Teruel B., Maria C. T., Martínez C., ~~s I., Pateman M., Simms T. M., Sivoli C. G., Vilar M., Kennett D. J., Keegan W. F., Coppa A., Lipson M., Pinhasi R., and Reich D., A genetic history of the pre-contact Caribbean, *Nature*. (2021) **590**, 103–110, https://doi.org/10.1038/s41586-020-03053-2.

CAS | PubMed | Web of Science® | Google Scholar

54   Posth C., Nakatsuka N., Lazaridis I., Skoglund P., Mallick S., Lamnidis T. C., Rohland N., Nägele K., Adamski N., Bertolini E., Broomandkhoshbacht N., Cooper A., Culleton B. J., Ferraz T., Ferry M., Furtwängler A., Haak W., Harkins K., Harper T. K., Hünemeier T., Lawson A. M., Llamas B., Michel M., Nelson E., Oppenheimer J., Patterson N., Schiffels S., Sedig J., Stewardson K., Talamo S., Wang C.-C., Hublin J.-J., Hubbe M., Harvati K., Nuevo Delaunay A., Beier J., Francken M., Kaulicke P., Reyes-Centeno H., Rademaker K., Trask W. R., Robinson M., Gutierrez S. M., Prufer K. M., Salazar-García D. C., Chim E. N., Müller Plumm Gomes L., Alves M. L., Liryo A., Inglez M., Oliveira R. E., Bernardo D. V., Barioni A., Wesolowski V., Scheifler N. A., Rivera M. A., Plens C. R., Messineo P. G., Figuti L., Corach D., Scabuzzo C., Eggers S., De Blasis P., Reindel M., Méndez C., Politis G., Tomasto-Cagigao E., Kennett D. J., Strauss A., Fehren-Schmitz L., Krause J., and Reich D., Reconstructing the deep population history of Central and South America, *Cell*. (2018) **175**, no. 5, 1185–1197.e22, https://doi.org/10.1016/j.cell.2018.10.027, 2-s2.0-85056451563, 30415837.

CAS | PubMed | Web of Science® | Google Scholar

55   Danecek P., Bonfield J. K., Liddle J., Marshall J., Ohan V., Pollard M. O., Whitwham A., Keane T., McCarthy S. A., Davies R. M., and Li H., Twelve years of SAMtools and BCFtools, *GigaScience*. (2021) **10**, no. 2, article giab008, https://doi.org/10.1093/gigascience/giab008, 33590861.

PubMed | Web of Science® | Google Scholar

56   Quinlan A. R. and Hall I. M., BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics*. (2010) **26**, no. 6, 841–842, https://doi.org/10.1093/bioinformatics/btq033, 2-s2.0-77951770756, 20110278.

CAS | PubMed | Web of Science® | Google Scholar

57   Rivera Franco N., Braga Y., Espitia Fajardo M., and Barreto G., Identifying new lineages in the Y chromosome of Colombian Amazon indigenous populations, *American Journal of Physical Anthropology*. (2020) **172**, no. 2, 165–175, https://doi.org/10.1002/ajpa.24039, 32141615.

PubMed | Google Scholar

‹ Back

non-recombining DNA, Cells. (2023) 14, no. 10, https://doi.org/10.3390/genes14101887, 37895937.

PubMed   Google Scholar

59   Y Chromosome Consortium, A nomenclature system for the tree of human Y-chromosomal binary haplogroups, *Genome Research*. (2002) **12**, no. 2, 339–348, https://doi.org/10.1101/gr.217602, 2-s2.0-0036178949, 11827954.

CAS   PubMed   Web of Science®   Google Scholar

60   van Oven M., Van Geystelen A., Kayser M., Decorte R., and Larmuseau M. H. D., Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome, *Human Mutation*. (2014) **35**, no. 2, 187–191, https://doi.org/10.1002/humu.22468, 2-s2.0-84891946246, 24166809.

CAS   PubMed   Google Scholar

Download PDF

Back