Research paper

# Testing the Ion AmpliSeq™ HID Y-SNP Research Panel v1 for performance and resolution in admixed South Americans of haplogroup Q

Zehra Köksal [a,*], Germán Burgos [b,c], Elizeu Carvalho [d], Silvia Loiola [d], María Laura Parolin [e], Alfredo Quiroz [f], Ândrea Ribeiro dos Santos [g,h], Ulises Toscanini [i], Carlos Vullo [j], Claus Børsting [a], Leonor Gusmão [d], Vania Pereira [a]

[a] Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[b] Escuela de Medicina, Facultad de Ciencias de la Salud, Universidad de Las Américas (UDLA), Quito, Ecuador
[c] Grupo de Medicina Xenómica, Universidad de Santiago de Compostela, Santiago de Compostela, Spain
[d] DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro 20550-013, Brazil
[e] Instituto de Diversidad y Evolución Austral (IDEAus), Centro Nacional Patagónico, CONICET, Puerto Madryn, Argentina
[f]Instituto de Previsión Social, Asunción 100153, Paraguay
[g] Human and Medical Genetics Laboratory, Institute of Biological Sciences, Federal University of Pará, Belém, Pará, Brazil
[h] Center for Oncology Research, Federal University of Pará, Belém, Pará, Brazil
[i] Primer Centro Argentino de Inmunogenética (PRICAI), Fundación Favaloro, Buenos Aires, Argentina
[j] DNA Forensic Laboratory, Argentine Forensic Anthropology Team (EAAF), Córdoba, Argentina

## ARTICLE INFO

## ABSTRACT

Y haplogroups, defined by Y-SNPs, allow the reconstruction of the human Y chromosome genealogy, which is important for population, evolutionary and forensic genetics. In this study, Y-SNPs were typed and haplogroups inferred with the MPS Ion AmpliSeq™ HID Y-SNP Research Panel v1, as a high-throughput approach. Firstly, the performance of the panel was evaluated with different DNA input amounts, reagent volumes and cycle numbers. DNA-inputs from 0.5 to 1 ng generated the most balanced read depth. Combined with full reagent and 19 cycles, this offered the highest number of amplicons with a sequencing read depth of at least 20 reads. Secondly, the sub-haplogroups of 182 admixed South Americans and Greenlanders belonging to haplogroup Q were inferred and tested for potential improvement in resolution. Most samples were assigned to lineage Q-M3 with some samples assigned to lineages upstream (Q-M346, L56, L57; Q-L331, L53; Q-L54; Q-CTS11969, CTS11970) or parallel (Q-L330, L334; Q-Z780/M971) to Q-M3. Only one sample was assigned to a downstream lineage (Q-Z35615, Z35616). Most individuals of haplogroup Q with NAM ancestry could neither be distinguished from each other, nor from half of the Greenlandic samples. Typing additional, known SNPs within lineage Q-M3, Z19483 and SA05, increased the resolution of predicted haplogroups. The search for novel variants in the sequenced regions allowed the detection of 42 variants and the subdivision of lineage Q-M3 into new subclades. The variants found in six of these subclades were exclusive to certain South American countries. In light of the limited differentiation of haplogroup Q samples, the additional information on known or novel SNPs disclosed in this study when using MPS Ion AmpliSeq™ HID Y-SNP Research Panel v1 should be included in the Yleaf software, to increase the differentiation of lineage Q-M3.

## 1. Introduction

Y haplogroups, which are defined by Y-SNPs, harbor crucial information for evolutionary, population and forensic genetic studies [1]. The biallelic Y-SNPs with low mutation rates, lack of recombination and paternal inheritance allow reconstruction of the human Y chromosome genealogy. Y-SNP haplogroup analyses enable estimation of human origin, migration patterns of male ancestors and dating of haplogroup branch points. Furthermore, Y haplogroup frequencies in different populations can be used to differentiate male lineages [1–3].

For the determination of Y haplogroups, the Single Base Extension (SBE) technology or minisequencing is widely used in forensic laboratories. It is usually comprised of a multiplex PCR and one or more multiplex SBE reaction(s). Here, the primer adjacent to the SNP is extended with a fluorescently labelled ddNTP and the SBE products are detected via Capillary Electrophoresis (CE) [4]. This technology is robust and sensitive, making it suitable for forensic and evolutionary purposes. However, it has low throughput and targeting multiple SNPs requires several multiplex reactions [5–11].

In the last decades, the development of Massively Parallel Sequencing (MPS) offered reproducible and simultaneous sequencing of thousands of SNPs [12]. However, the highly repetitive nature of the male-specific regions of the Y chromosome complicates the sequencing process. The unique regions make up 8.97 Mb in total and are distributed across the Y chromosome and intermitted by non-unique regions [13]. Targeted sequencing [14,15] or whole genome sequencing with subsequent extraction of relevant positions [13,16,17] enabled large-scale typing of SNPs. The phylogenetic relationships between these SNPs were determined and represented in phylogenetic trees [14, 18–20]. SNPs at specific positions within the phylogenetic tree can be selected for sequencing panels that could either cover as many broad haplogroups as possible or focus on high-resolution sub-branches of certain haplogroups [21,22].

The Ion AmpliSeq™ HID Y-SNP Research Panel v1 is a large-scale Y-SNP typing panel that targets 602 amplicons with 884 SNPs. The majority of the sequenced bases (82%) lie within unique regions of the Y chromosome [13]. The panel include SNPs, which are variable in most human populations. The dominating markers included in the panel target haplogroups in R, E and I branches [21].

Haplogroup Q has a wide geographic distribution ranging from parts of Europe, the Middle East, Asia and the Americas [23,24]. The major lineage is Q1, which subdivides into Q1a and Q1b. Haplogroup Q1a branches into lineage Q-NWT01, which is mainly found in regions around the Arctic Ocean including Northern Canada and Greenland [25, 26]. Q1b is a widespread haplogroup found in the Middle East, China, Mongolia, Siberia and the Americas [27–29]. Its subclade Q-L54 is more represented in Northeast Siberia, and the Americas [9,26,30,31].

In the Americas most of the carriers of haplogroup Q are individuals with patrilineal Native American (NAM) ancestry [31]. NAM Y chromosomes are separated in only ten sub-branches [30,32] with very limited haplogroup inference resolution. The majority (61–75%) of the NAM Y chromosomes are assigned to the sub-lineage Q-M3, carrying a C to T transition within the DYS199 locus [31,33–35]. NAM ancestry can be found to a small extent in admixed modern-day South Americans. Due to the European occupation in the late 15th century [36] and the African slave trade from the middle of the 16th until the middle of the 19th century [37], the genetic profiles of modern South American populations consist mainly of European, African and/or NAM ancestries [38,39].

The present study aimed to evaluate the general performance of the Ion AmpliSeq™ HID Y-SNP Research Panel v1 by testing different DNA input amounts, reagent volumes and amplification cycle numbers. In the second part, we assessed the sub-haplogroup inference of admixed South Americans as well as Greenlanders belonging to haplogroup Q, when using known Y-SNPs and new variation detected in the targeted sequences.

## 2. Material and methods

### 2.1. Samples, DNA extraction and quantification

The performance of the Ion AmpliSeq™ HID Y-SNP Research Panel v1 was evaluated with the DNA 007 sample (Applied Biosystems, Foster City, CA, USA) to determine the most adequate DNA input, cycle number, and reagent volume (Supplementary Figure 1).

For testing haplogroup Q inference, 167 samples from South America and 15 from Greenland belonging to haplogroup Q were selected. The samples were received and selected on the basis of previous studies, in which more than 2000 samples were collected under informed consent and were approved to be used for this study [25,38,40–44]. The study was conducted according to the guidelines of the declaration of Helsinki [45]. The blood samples were collected on FTA cards (Whatman Inc., Clifton, NJ, USA). The DNA was extracted using standard Chelex and phenol-chloroform extraction methods, or with the BioRobot EZ1 Workstation (Qiagen, Hilden, Germany) following the manufacturer's recommendations. DNA concentrations were measured by the Qubit™ dsDNA High Sensitivity assay and the Qubit® 3.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) following the manufacturer's recommendations.

### 2.2. Sample selection

The samples from unrelated men from different South American regions and Greenland had previously been typed for different sets of Y-STRs in connection to other studies [25,38,40–44]. Depending on the STR kit used, data were partly or completely available for 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, DYF387S1. The STR profiles were further investigated using the Y-DNA Haplogroup Predictor NevGen (www.nevgen.org) to determine the most likely haplogroup.

A total of 59 Y-SNPs were typed in 961 of the South Americans using PCR-SBE-CE as previously described [43,46–50]. Ten of the targeted SNPs belonged to haplogroup Q -M242, P36.2, M346, M3, M19, Z19319, SA01, Z19483, M557, SA05 [51] (Supplementary Figure 2).

After sample screening, 167 samples that belonged to the Native American haplogroup Q were selected for this study from different regions in Argentina (N = 38), Bolivia (N = 52), Brazil (N = 17), Ecuador (N = 36) and Paraguay (N = 24). For an overview of collection sites, please see Supplementary Figure 3. Fifteen individuals from Greenland belonging to haplogroup Q were included in this study as possible outliers to the South American samples.

### 2.3. Ion S5 sequencing

The library preparation was conducted using the Ion AmpliSeq™ Library Kit 2.0 (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's manual (Precision ID Library Kit on the Ion-Torrent system by Thermo Fisher Scientific).

For the performance testing, different DNA input amounts (1 ng, 0.5 ng, 0.2 ng, 0.1 ng, 0.05 ng, 0.025 ng and 0.0125 ng), cycle numbers during target amplification (19 and 21 cycles) and reagent volumes during amplification and library preparation were used (volume recommended by the manufacturer and half the recommended volume). For all conditions tested, samples were amplified and sequenced in duplicates. The library preparation of samples from haplogroup Q was conducted using 1 ng of DNA when possible and either 19 cycles with full reagent volume or 21 cycles with half reagent volume (Supplementary Figure 1).

The libraries were purified manually (Precision ID SNP Panels with the HID Ion S5™/HID Ion GeneStudio™ S5 System Application Guide)

or on a Biomek® 3000 Laboratory Automation Workstation (Beckman Coulter Inc., CA, USA) [52,53] by adding 45 µl of (1.5x) Agencourt® AMPure® XP Beads (Beckman Coulter, Indianapolis, IN, USA) and following the manufacturer's manual. The eluted libraries were quantified using the Library TaqMan™ Quantitation Kit (Thermo Fisher Scientific, Waltham, MA, USA) and pooled to a final concentration of 50 pM. For the South American and Greenlandic samples, libraries with concentrations below 50 pM (range: 13–46 pM) were re-amplified with 8 cycles according to the manufacturer's manual (Precision ID Library Kit on the IonTorrent system by Thermo Fisher Scientific). For the libraries in the performance test with concentrations below 50 pM (range: 6–48 pM), re-amplifications were omitted in order to present the influence of the chosen experiment conditions. The libraries were pooled with equal volumes.

Sequencing was carried out using the Ion S5™ Precision ID Chef & Sequencing Kit (Thermo Fisher Scientific, Waltham, MA, USA) and 32 samples were loaded on each Ion 530™ Chip Kit (Thermo Fisher Scientific, Waltham, MA, USA). Samples were sequenced on the Ion S5™ System with 650 run flows.

### 2.4. Sequence data analysis, Y-SNP calling and Y haplogroup inference

Sequence analysis was initially performed on the Torrent Suite Server v.5.10.1 (Thermo Fisher Scientific, Waltham, MA, USA) including alignment and base calling. BAM-files were generated using the adequate target and hotspot BED-files in the Torrent Suite™ Software and aligned to the hg19/GRCh37 reference.

The plugin coverageAnalysis v5.10.0.3 (Thermo Fisher Scientific, Waltham, MA, USA) was used to observe the sequence coverage of the targeted regions.

For variant calling, the plugin variantCaller v5.10.1.20 was used and a minimum variant frequency of 98% was applied. BAM- and CSV-files generated by the plugin were used for all further analyses.

The software Yleaf version 2.2 [54] was applied for Y-SNP calling and Y haplogroup inference using bam and index files. The following acceptance criteria were applied: minimum of 20 reads for each base, quality threshold of 20 (1% error rate) for each read and majority base threshold 95% (at least 95% of the reads of a base have to be identical). The results of the haplogroup prediction in the Yleaf software were based on the most downstream SNPs that passed these quality criteria. For haplogroups that were represented by several alternative SNPs in the panel, the software checks if at least one of the SNP positions carried an ancestral allele, and if so, any additional information from the alternative SNP locations is disregarded. SNP data that did not meet the quality thresholds were manually inspected.

For haplogroup inference, a position file is needed stating the location of the SNP as well as the possible alleles. The Yleaf v.2.2 software has different position files of targeted SNPs available: firstly, the 884 Y-SNPs targeted by the panel and secondly SNPs sequenced by whole genome sequencing. The latter contained 64812 unique SNP locations, 905 of which lie within the 602 amplicons and 56 of these are variable in haplogroup Q. The Y-SNP distributions in scatter maps were generated using the scattergeo function in the package plotly v4.14.3 in python v3.6.8.

### 2.5. Verification of SNP variants

Novel variants found using the Torrent Suite Server's variant Caller plugin were validated using Yleaf and checked using the Integrative Genomics Viewer (IGV). Variants that lie close to the amplicon start or end and those in repetitive regions were confirmed by singleplex PCR-SBE-CE analysis. DNA was amplified using the Qiagen Multiplex PCR kit (Qiagen, Hilden, Germany) and a final concentration of 0.2 µM of the forward and reverse PCR primers, respectively. In a thermal cycler, DNA denatured at 95 °C for 15 min, followed by 35 cycles of 94 °C for 30 s, the annealing temperature of the respective PCR primers (Supplementary Table 1) for 30 s and 72 °C for 30 s, and a final extension at 72 °C for 10 min. For the enzymatic clean-up 5 µl PCR product were combined with 2 µl Exo-SAP-IT™ (Thermo Fisher Scientific, Waltham, MA, USA) and incubated at 37 °C for 60 min and 75 °C for 15 min. The SBE was conducted by adding 2 µl SNaPshot™ Multiplex Ready Reaction Mix to 1 µl of cleaned PCR product with the SBE primer at a final concentration of 0.2 µM. Thermal cycling was composed of 30 cycles of 96 °C for 10 s, the corresponding SBE primer annealing temperature (Supplementary Table 1) for 5 s and 60 °C for 30 s. The SBE products were treated with 1 µl SAP (Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C for 30 min and 75 °C for 15 min

For capillary electrophoresis, 0.5 µl of the product was combined with 0.5 µl GeneScan™ 120 LIZ® Size Standard and 9 µl Hi-Di formamide. The SBE fragments were separated and detected using the Applied Biosystems® 3500 Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA, USA) and the FragmentAnalysis36_POPxl run module (POP-4™ polymer, 36 cm capillary, Dye Set E5). The analysis of results was conducted using GeneMapper® ID-X Software v.1.4 (Thermo Fisher Scientific, Waltham, MA, USA).

All detected positions were checked in other sources to investigate if they had previously been reported, namely in the ISOGG and in the dbSNP databases and the Y-Phylotree [55], as well as via literature research [9,30,51,56]. Additionally, VCF and CRAM files of publicly available data have been investigated for the 42 variants (Supplementary Table 2). Novel SNPs not previously described have been submitted to the dbSNP database.

### 2.6. Median-joining network analysis

Median-joining networks of Y-STR profiles including the loci DYS19, DYS385a, DYS385b, DYS389I, DYS389II-I, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and YGATAH4 were built using the Network version 10.2.0.0 (http://www.fluxus-engineering.com/). The STR loci were weighted proportionally to the reciprocal locus variance among all 182 samples as suggested by [57] and rounded to a whole number.

## 3. Results and Discussion

### 3.1. Performance test

The performance of the panel was analysed by testing three criteria: DNA input, cycle number during amplification, and reagent volume for target amplification and library preparation.

As expected, a combination of high DNA input, higher PCR cycle number and full reagent volume led to higher library concentrations (Supplementary Figure 4).

For DNA input below 0.1 ng, all library concentrations were below 50 pM. Even for inputs above 0.2 ng, not all amplicons were amplified and those that were amplified had differences in coverage. The normalized amplicon balance (Supplementary Figure 5) showed preferential amplification of some amplicons. This is most likely due to differences in the PCR efficiency [58]. For DNA inputs of 1 ng or 0.5 ng, the four experimental set-ups (19 or 21 cycles, full or half volume of reagents) had very similar amplicon balance distributions whereas the read depth variation was larger for libraries with a DNA input of 0.2 ng (Supplementary Figure 5).

Looking at the amplicon read depth and its relation to the amplicon size, it was observed that longer amplicons were more likely to have read counts below the threshold of 20 reads, although a few exceptions were detected. Amplicon length was significantly correlated with the percentage of samples with coverage of > 20 reads (Kruskal-Wallis test H=91.61, p = 3.07e-18) (Supplementary Figure 6).

Supplementary Figure 7 presents the number of amplicons per sample with less than 20 reads. As expected, the number of underperforming amplicons increased with a decrease in the DNA input

amount. In the case of 0.2 ng DNA input or lower, 21 cycles and full reagent volume resulted in higher read depths and lower number of amplicons with less than 20 reads. For a DNA input of 0.5 ng or more, the choice of cycle number and reagent volume was less critical. The best results were achieved for set ups with at least 0.5 ng of DNA, full volume of reagents and 19 cycles. However, for economic reasons, using half volume of reagents and 21 cycles is a good alternative with only a minimal decrease in performance.

Some amplicons failed to be amplified even under optimal conditions (1 ng DNA input, 19 cycles, full volume of reagents). Six amplicons were covered by less than 20 reads in all samples (Supplementary Table 3). These make up 1% of all amplicons in the panel and vary in size from 38 to 127 bp. Sixteen SNPs are positioned on these amplicons and they define branches of haplogroups A (2), B (1), C (1), E (3), G (1), H (1), I (1), J (1), O (1), R (3), and Q (1); the latter, L529, belongs to a parallel branch of Q-M3. Five of these SNPs have at least one alternative SNP of the same sub-branch included in the panel. These alternative SNPs are included in amplicons which are sequenced with more than 20 reads in at least half of the samples (Supplementary Table 3).

### 3.2. Application of the Y panel on samples of haplogroup Q

Y-STRs and/or Y-SNPs were typed to identify admixed South Americans that were likely to carry a Y chromosome within haplogroup Q [38,40–44]. Of the 2041 South Americans, 167 were predicted to belong to haplogroup Q. Furthermore, fifteen samples from Greenland belonging to haplogroup Q were added as an outlier group to the South American samples. All 182 samples were typed with the Ion AmpliSeq™ HID Y-SNP Research Panel v1 and the Y-SNP results confirmed that these individuals belonged to haplogroup Q.

The haplogroup predictions based on the 56 Q-SNPs targeted by the Y-MPS panel are presented in Fig. 1. The predictions were based on the most downstream SNPs that met the quality control criteria (minimum of 20 reads per base, quality threshold of 20 for each read, majority base threshold of 95%). When SNPs from the position files of Yleaf (based on the phylogenetic tree) were identified, additional SNPs of further downstream haplogroups were analysed to achieve higher resolution. If all alternative SNPs of a haplogroup failed the quality criteria, the software assigned the upstream haplogroup as the predicted haplogroup. Therefore, the software does not necessarily report the true haplogroup resolution of the sample, and the prediction is greatly influenced by the quality of the data generated.

The distribution of Q haplogroups showed that the subclade Q-M3, found in 145 out of 167 South American samples, was the most prominent in all South American countries (Fig. 1). Nine additional samples were predicted to belong to a haplogroup upstream of lineage Q-M3, due to a lack of typed markers CTS7779 and M3. As shown in Supplementary Table 4, eight of these samples belonged to lineage Q-M3 according to SNaPshot results. For the remaining sample, SNaPshot data were not available.

The only lineage downstream of Q-M3 detected was Q-Z35616, which was found in one individual in Bolivia.

The two lineages parallel to the Q-M3 branch, Q-L330, L334 and Q-M971/Z780, were found in Northern Brazil and in samples from Ecuador over the Bolivian Andean region to Southeast Argentina, respectively. Haplogroup Q-M971/Z780 has not been described in populations outside South America to date, and its presence in 12 individuals in our dataset is in accordance with findings in the literature, where Q-Z780/M971 is suggested to be associated to Native Americans [9,57,59].

Out of the 15 Greenlanders, eight had the same lineage Q-M3 as the majority of the South American samples. The remaining seven Greenlanders carried the haplogroup Q-NWT01/F746 from a parallel branch (Fig. 1), which was in accordance with previous studies [25].

Overall, the haplogroup Q-SNPs in the Y-MPS panel were evenly spread mainly over parallel and upstream branches of Q-M3. Therefore,

the strength of the Ion AmpliSeq™ HID Y-SNP Research Panel v1 lies in distinguishing samples from phylogenetically distant Q-branches, rather than samples from phylogenetically close sub-branches of Q-M3.

The information based on Y-STR data showed that the 182 haplogroup Q samples had 167 unique Y-STR haplotypes, based on 17 Y-STR data available for all samples (Supplementary Figure 8). The Y-STR profiles of samples within lineage Q-M3 were very diverse most probably due to the ancient origin of this lineage [32]. The most distinct Y-STR profiles were those within lineage Q-F746/NWT01, which presents an earlier separation from other branches leading to Q-M3, Q-M971/Z780 and Q-L330, L334. The two samples of lineage Q-L330, L334 found in Brazil, had very similar Y-STR profiles. Most likely this is due to a recent common ancestor of the two individuals. Alternatively, this could indicate a more recent origin of the variation, when compared to Q-M3 and Q-M971/Z780. Interestingly, as for Q-M3, STR profiles inside haplogroup Q-M971/Z780 were very diverse, being in distant branches of the network. This result hints to potential variation hidden within these haplogroups that has not been disclosed by the SNP markers analysed in this work. Further, the lack of clear separation of the STR profiles between some Q-M3 and Q-M971/Z780 branches might be due to recurrent microsatellite mutations.

### 3.2.1. Additional Y-SNPs improved the resolution of haplogroup Q

Out of the 167 South Americans, 133 samples were previously typed for ten haplogroup Q SNPs using PCR-SBE-CE [43]. This information resulted in higher resolution of six samples (Supplementary Figure 9). Four samples from South Bolivia, North and East Argentina, and East Paraguay had the Q-Z19483 variant and two samples from North Brazil had the Q-SA05 variant. Overall, these markers increase the resolution within lineage Q-M3 and therefore can be used to differentiate individuals of this lineage.

### 3.2.2. Novel variations identified using the Ion AmpliSeq™ HID Y-SNP Research Panel v1

The sequenced amplicons of South American and Greenlandic samples were further investigated with the variant caller plugin for potential new variation not identified by Yleaf. The plugin reported 46 additional variants. All variants were manually inspected using IGV and searched for in public databases (ISOGG Y-DNA Haplogroup Trees 2016–2020, dbSNP (last accessed in February 2022), Y-Phylotree [55]).

Two of these variants were annotated SNPs (rs1005041 at position 7570822 and rs7892914 at position 22293981 according to GRCh37) that are not specific for haplogroup Q. The SNP rs1005041 specifies a sub-clade of haplogroup R (the haplogroup of the reference sequence), while rs7892914 was found in the derived allelic state in samples of haplogroups Q, R and I and therefore, is not specific of haplogroup Q. Two other variants (in positions 15702828 and 22505839) were sequencing artifacts located in highly repetitive regions.

Of the remaining 42 variants, eleven were previously reported in the dbSNP database [60] and 31 were, to our knowledge, not reported in public databases. Further information on the alternative alleles of the 42 variants is presented in Supplementary Table 5. Six of the 31 variants (in positions 7842069, 14125686, 17434355, 17922077, 22505875 and 22741690) were located close to the ends of an amplicon and/or in repetitive regions. These variants were successfully confirmed by singleplex PCR-SBE-CE analysis (data not shown).

In order to refine the location of these new variants within the haplogroup Q tree, information was collected from publicly available Y-chromosomal data for South American individuals [9,17,30,51,55,56,61–63].

Among the eleven previously reported SNPs, one SNP (rs4252209, MEH2) was found in all haplogroup Q samples. Two SNPs (rs541403360, rs564763483) were only absent in samples assigned to haplogroups Q1a1-F746/NWT01, Q1b1a2-M971/Z780 and Q1b1a3-L330,L334. These two SNPs (rs541403360, rs564763483) were previously suggested to have the same position in the tree as the marker Q-M3
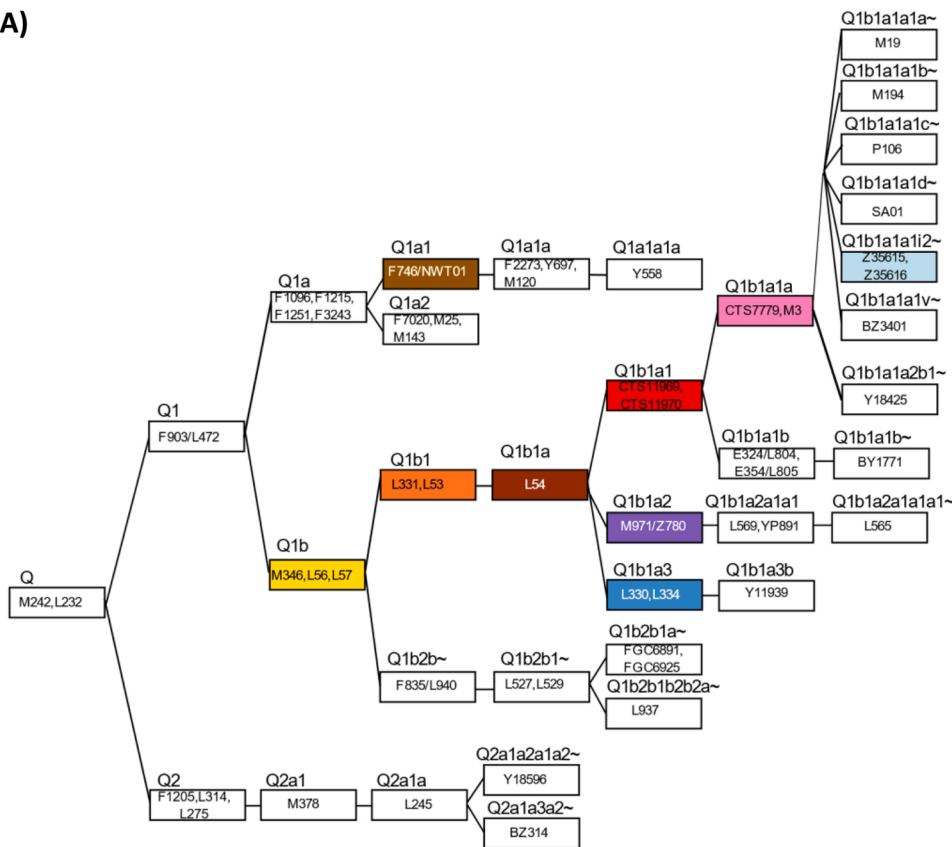
**A)**



**B)**



**Fig. 1.** (A) Phylogenetic tree with the typed Q-SNPs in the Ion AmpliSeq™ HID Y-SNP Research Panel v1. Different names for the same SNP position are separated by '/'. Haplogroups in white cells (A) were not found in any of the typed samples. The other haplogroups were found in at least one individual. Among the detected lineages, Q-NWT01/F746 was only found in the typed samples from Greenland and was absent in the samples from South America. (B) The haplogroups that were predicted based on the most downstream SNPs that met the quality control criteria were illustrated in the map for South American samples with the colour code according to the legend presented in (B).
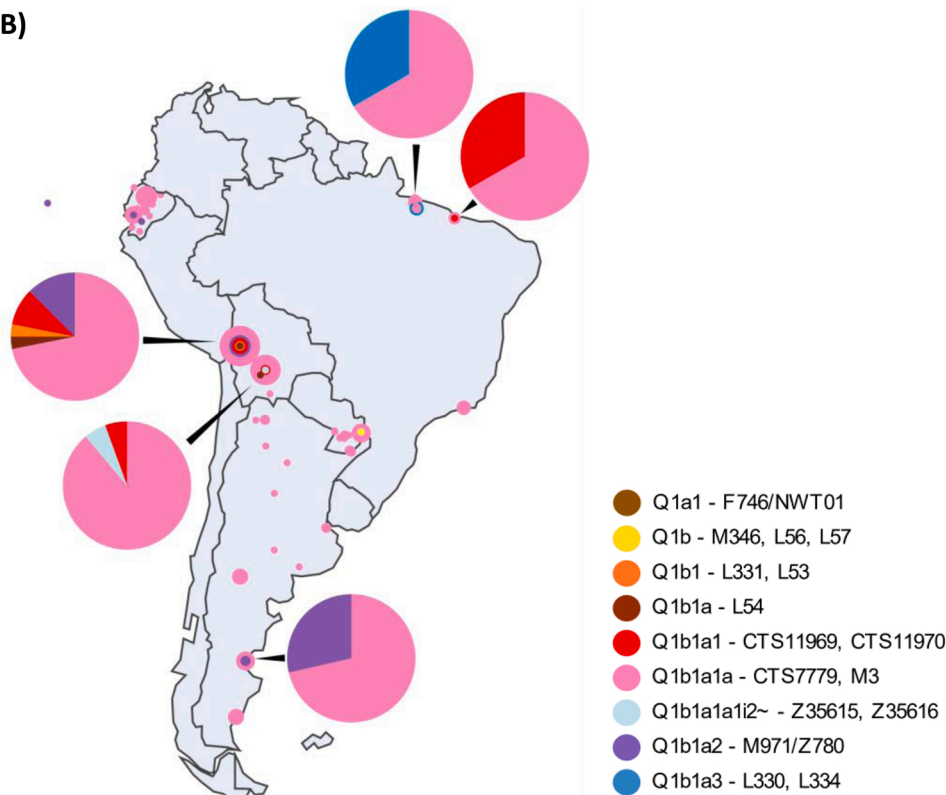
[63].

For variants downstream of lineage Q-M3, our findings suggest that SNP rs779825282 is downstream of marker Z35616 (with missing information on Z35615). In another sample, Z35615 was suggested to be positioned downstream of Z35616 (without carrying SNP rs779825282) as shown in Supplementary Figure 10. SNaPshot results suggest that marker SA05 forms a parallel branch to all novel variants downstream of lineage Q-M3. Among the three samples belonging to lineage Q-Z19483, one carried the variants rs1603040369, rs779238512 and rs1603557500. The resulting phylogenetic tree including all 42 variants and relevant Q-SNPs is shown in Supplementary Figure 10.

In the 1000 Genomes dataset, one Peruvian sample (HG01967) of haplogroup Q1b1a1a1k1a1~ (according to ISOGG Haplogroup Tree 2019–2020) carried the SNP rs757579581. The SNP rs757579581 was found in the same sample as marker Q-Z35764 [61,63]. Among the 11 individuals analysed by Karmin et al. (2015), one of the South American samples from Colla, Argentina (GS000019960-ASM) carried the variant at position 7192964. The individual with the variant at position 7192964 also carried the marker Q-B46 and was therefore assigned to haplogroup Q1b1a1a1k2~. A more detailed tree for the variants downstream of lineage Q-M3 based on findings in the literature is presented in Supplementary Figure 11.

In order to understand whether some variants could be allocated to certain geographic regions, nine variants found in more than one individual were investigated in more detail. The phylogenetic relationships of the nine variants, which are downstream of Q-CTS7779, M3, are depicted in Fig. 2. Most of the nine variants were exclusive to samples from Ecuador. Eleven samples from Ecuador had a derived allele at position 15508183 (G→A), five of which additionally showed the derived variant at location 22505875 (C→G) and one at both 19179327 (G→A) and 22505832 (G→T). Further derived variants were observed at

position 17434355 (A→T) in three Ecuadorian samples, 7192964 (T→C) in two Ecuadorian samples, 22797693 (A→G) in two Bolivian samples and at both locations 7842069 (G→A) and 14495293 (T→A) in two Paraguayan samples.

To evaluate if this new underlying SNP variation was reflected in the Y-STR haplotypes, a median-joining network based on Y-STR profiles for all Q-M3 individuals was constructed (Supplementary Figure 12).

Individuals with variants in the position 22797693 shared the same Y-STR haplotype. Related profiles were observed in individuals carrying variant 17434355. Similar was observed for individuals with variants in positions 7842069/14495293. This points towards a more recent origin of each of these variants. For individuals with variants in position 15508183 partial clustering of their Y-STR patterns was observed. The two individuals that presented variation in position 7192964 showed Y-STR profiles differing in 20 mutational steps.

In summary, the comparison of the data generated in this work with data from other sources allowed the placement of six variants (MEH2, rs541403360, rs564763483, rs757579581, rs779825282 and 7192964) within the phylogenetic Q-tree. For the remaining 36 variants information on the phylogeny downstream of lineage Q-M3 is still limited. The geographic distribution of nine SNPs present in more than one individual did not provide enough resolution to show clear patterns in the samples from different South American populations. Sequencing of a bigger region of the Y chromosome and increasing the number of analysed samples [13] could enable the discovery of additional variants that are specific for certain sub-populations, and help to complete the phylogenetic relationships between known and novel SNPs.

## 4. Conclusions

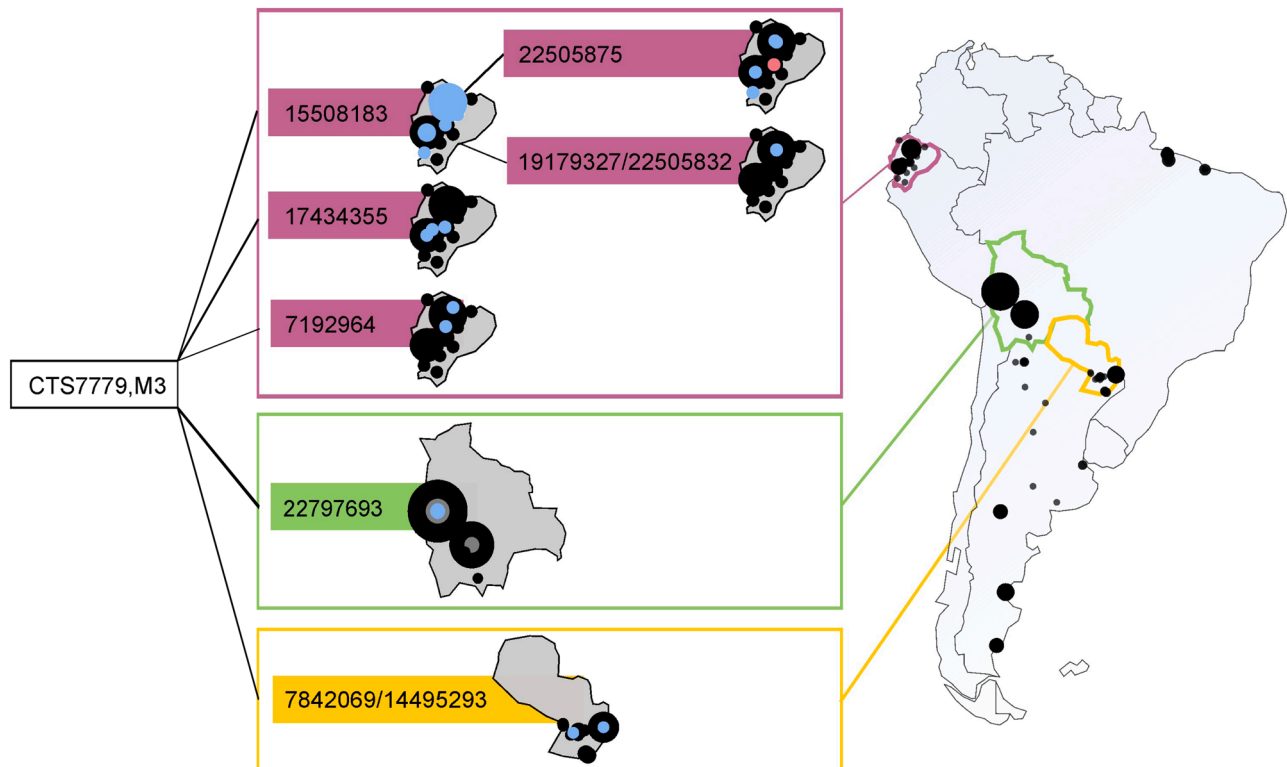Here, we evaluated the performance of a customized Y-SNP panel



**Fig. 2.** Combined phylogenetic tree of 9 novel variants downstream of Q-CTS7779, M3 and the geographic origin of the 20 samples carrying at least one of the 9 novel variants. Six variants were exclusively found in Ecuador (magenta) and one in Bolivia (green) and Paraguay (yellow). Variant positions separated by '/' were observed in the same allelic state within all samples. The distribution and abundance of the samples carrying the respective variants are shown in maps of the individual countries. The colour coding characterizes the sequencing quality: black – ancestral variant, above threshold (>20 reads); grey – ancestral variant, below threshold (<20 reads); blue – derived variant, above threshold (>20 reads), red – derived variant, below threshold (<20 reads).

developed for MPS. The panel was also used to type 182 samples belonging to haplogroup Q (admixed South Americans and Greenlanders) in order to assess the sub-haplogroup inference within haplogroup Q.

The best performance was achieved with 1 ng DNA input, full volume of reagents and 19 cycles. Using half volume of reagents and increasing the number of cycles to 21 did not greatly compromise the panel performance.

The Q-SNPs included in the panel covered mainly sub-haplogroups upstream and parallel to Q-M3, rather than downstream. Thus, the haplogroup resolution of the typed South American and Greenlandic samples was somewhat limited, since 85% of the samples (154 out of 182 samples) were assigned to lineage Q-M3.

Nine samples were assigned to haplogroups upstream of lineage Q-M3 based on low quality data generated with the MPS panel. Eight of these samples had SNaPshot data available and were confirmed to belong to lineage Q-M3.

Due to its broad haplogroup coverage, this MPS panel can serve as a first-tier screening tool and additional and more specific panels can be designed to achieve higher resolution when needed.

For the specific case of Q-M3, it would be relevant to include known downstream SNPs, such as Z19483 and SA05; and the novel variations detected already targeted in the sequenced amplicons should be included in the position file of the Yleaf analysis software. The latter would allow further subdivision of the lineage Q-M3 into seven sub-branches that so far have been exclusively found in samples from Ecuador, Bolivia and Paraguay. This regional variation may hint to wider continental patterns of diversity. Although not yet fully disclosed, a broader and more exhaustive study of these lineages may reveal information on the ancestral paths of the Native Americans into and throughout the continent.

## Acknowledgments

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2022.102708.

## References

[1] P.A. Underhill, L. Jin, A.A. Lin, S.Q. Mehdi, T. Jenkins, D. Vollrath, et al., Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography, Genome Res 7 (1997) 996–1005, https://doi.org/10.1101/gr.7.10.996.

[2] M.A. Jobling, C. Tyler-Smith, Fathers and sons: the Y chromosome and human evolution, Trends Genet 11 (1995) 449–456, https://doi.org/10.1016/s0168-9525(00)89144-1.

[3] P. de Knijff, Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome, Am. J. Hum. Genet. 67 (2000) 1055–1061, https://doi.org/10.1016/S0002-9297(07)62935-8.

[4] M. Fondevila, C. Børsting, C. Phillips, M. de la Puente, E.-N. Consortium, A. Carracedo, et al., Forensic SNP genotyping with SNaPshot: technical considerations for the development and optimization of multiplexed SNP assays, Forensic Sci. Rev. 29 (2017) 57–76.

[5] S. Inagaki, Y. Yamamoto, Y. Doi, T. Takata, T. Ishikawa, K. Yoshitome, et al., Typing of Y chromosome single nucleotide polymorphisms in a Japanese population by a multiplexed single nucleotide primer extension reaction, Leg. Med (Tokyo) 4 (2002) 202–206. https://doi.org/10.1016/s1344-6223(02)00032-9.

[6] J.J. Sanchez, C. Børsting, C. Hallenberg, A. Buchard, A. Hernandez, N. Morling, Multiplex PCR and minisequencing of SNPs–a model with 35 Y chromosome SNPs, Forensic Sci. Int 137 (2003) 74–84, https://doi.org/10.1016/s0379-0738(03)00299-8.

[7] V. Onofri, F. Alessandrini, C. Turchi, M. Pesaresi, L. Buscemi, A. Tagliabracci, Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs, Forensic Sci. Int 157 (2006) 23–35, https://doi.org/10.1016/j.forsciint.2005.03.014.

[8] A. Blanco Verea, J. Jaime, M. Brión, A. Carracedo, Y-chromosome lineages in native South American population, Forensic Sci. Int. Genet. 4 (2010) 187–193, https://doi.org/10.1016/j.fsigen.2009.08.008.

[9] M. Geppert, Q. Ayub, Y. Xue, S. Santos, Â. Ribeiro-dos-Santos, M. Baeta, et al., Identification of new SNPs in native South American populations by resequencing the Y chromosome, Forensic Sci. Int Genet 15 (2015) 111–114, https://doi.org/10.1016/j.fsigen.2014.09.014.

[10] B. Sobrino, M. Brión, A. Carracedo, SNPs in forensic genetics: a review on SNP typing methodologies, Forensic Sci. Int 154 (2005) 181–194, https://doi.org/10.1016/j.forsciint.2004.10.020.

[11] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724, https://doi.org/10.1002/elps.200500671.

[12] M.A. Jobling, C. Tyler-Smith, Human Y-chromosome variation in the genome-sequencing era, Nat. Rev. Genet 18 (2017) 485–497, https://doi.org/10.1038/nrg.2017.36.

[13] W. Wei, Q. Ayub, Y. Chen, S. McCarthy, Y. Hou, I. Carbone, et al., A calibrated human Y-chromosomal phylogeny based on resequencing, Genome Res 23 (2013) 388–395, https://doi.org/10.1101/gr.143198.112.

[14] R. Scozzari, A. Massaia, B. Trombetta, G. Bellusci, N.M. Myres, A. Novelletto, et al., An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa, Genome Res 24 (2014) 535–544, https://doi.org/10.1101/gr.160788.113.

[15] S. Lippold, H. Xu, A. Ko, M. Li, G. Renaud, A. Butthof, et al., Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences, Invest. Genet. 5 (2014) 13, https://doi.org/10.1186/2041-2223-5-13.

[16] S. Yan, C.-C. Wang, H.-X. Zheng, W. Wang, Z.-D. Qin, L.-H. Wei, et al., Y chromosomes of 40% Chinese descend from three neolithic super-grandfathers, PLoS One 9 (2014), e105691, https://doi.org/10.1371/journal.pone.0105691.

[17] M. Karmin, L. Saag, M. Vicente, M.A.W. Sayres, M. Järve, U.G. Talas, et al., A recent bottleneck of Y chromosome diversity coincides with a global change in culture, Genome Res 25 (2015) 459–466, https://doi.org/10.1101/gr.186684.114.

[18] R.M. Durbin, D. Altshuler, R.M. Durbin, G.R. Abecasis, D.R. Bentley, A. Chakravarti, et al., A map of human genome variation from population-scale sequencing, Nature 467 (2010) 1061–1073, https://doi.org/10.1038/nature09534.

[19] P. Francalacci, L. Morelli, A. Angius, R. Berutti, F. Reinier, R. Atzeni, et al., Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny, Science 341 (2013) 565–569, https://doi.org/10.1126/science.1237947.

[20] W. Wei, Q. Ayub, Y. Xue, C. Tyler-Smith, A comparison of Y-chromosomal lineage dating using either resequencing or Y-SNP plus Y-STR genotyping, Forensic Sci. Int Genet 7 (2013) 568–572, https://doi.org/10.1016/j.fsigen.2013.03.014.

[21] A. Ralf, M. van Oven, D. Montiel González, P. de Knijff, K. van der Beek, S. Wootton, et al., 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 Sci. Int Genet 41 (2019) 93–106, https://doi.org/10.1016/j.fsigen.2019.04.001.

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

[23] A. Varzari, V. Kharkov, A.G. Nikitin, F. Raicu, K. Simonova, W. Stephan, et al., Paleo-Balkan and Slavic contributions to the genetic pool of moldavians: insights from the Y chromosome, PLOS ONE 8 (2013), e53731, https://doi.org/10.1371/journal.pone.0053731.

[24] Y.-Z. Huang, H. Pamjav, P. Flegontov, V. Stenzl, S.-Q. Wen, X.-Z. Tong, et al., Dispersals of the Siberian Y-chromosome haplogroup Q in Eurasia, Mol. Genet Genom. 293 (2018) 107–117, https://doi.org/10.1007/s00438-017-1363-8.

[25] J.K. Olofsson, V. Pereira, C. Børsting, N. Morling, Peopling of the North Circumpolar Region – Insights from Y Chromosome STR and SNP Typing of Greenlanders, PLOS ONE 10 (2015), e0116573, https://doi.org/10.1371/journal.pone.0116573.

[26] M.C. Dulik, A.C. Owings, J.B. Gaieski, M.G. Vilar, A. Andre, C. Lennie, et al., Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athapaskan- and Eskimoan-speaking populations, PNAS 109 (2012) 8471–8476, https://doi.org/10.1073/pnas.1118760109.

[27] B. Malyarchuk, M. Derenko, G. Denisova, A. Maksimov, M. Wozniak, T. Grzybowski, et al., Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a, J. Hum. Genet 56 (2011) 583–588, https://doi.org/10.1038/jhg.2011.64.

[28] J.R. Sandoval, D.R. Lacerda, M.S.A. Jota, A. Salazar-Granara, P.P.R. Vieira, O. Acosta, et al., The genetic history of indigenous populations of the peruvian and bolivian altiplano: the legacy of the uros, PLOS ONE 8 (2013), e73006, https://doi.org/10.1371/journal.pone.0073006.

[29] H. Lacau, T. Gayden, M. Regueiro, S. Chennakrishnaiah, A. Bukhari, P.A. Underhill, et al., Afghanistan from a Y-chromosome perspective, Eur. J. Hum. Genet 20 (2012) 1063–1070, https://doi.org/10.1038/ejhg.2012.59.

[30] V. Battaglia, V. Grugni, U.A. Perego, N. Angerhofer, J.E. Gomez-Palmieri, S. R. Woodward, et al., The first peopling of South America: new evidence from Y-chromosome haplogroup Q, PLOS ONE 8 (2013), e71390, https://doi.org/10.1371/journal.pone.0071390.

[31] S.L. Zegura, T.M. Karafet, L.A. Zhivotovsky, M.F. Hammer, High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of native american Y

chromosomes into the Americas, Mol. Biol. Evol. 21 (2004) 164–175, https://doi.org/10.1093/molbev/msh009.

[32] L.S. Jurado Medina, P.B. Paz Sepúlveda, V. Ramallo, C. Sala, J. Beltramo, M. Schwab, et al., Patrilineages in Argentina and Paraguay (Continental Origin for Q Haplogroup), Hum. Biol. 92 (2021) 63–80, https://doi.org/10.13110/humanbiology.92.2.01.

[33] J.T. Lell, M.D. Brown, T.G. Schurr, R.I. Sukernik, Y.B. Starikovskaya, A. Torroni, et al., Y chromosome polymorphisms in native American and Siberian populations: identification of native American Y chromosome haplotypes, Hum. Genet 100 (1997) 536–543, https://doi.org/10.1007/s004390050548.

[34] J.T. Lell, R.I. Sukernik, Y.B. Starikovskaya, B. Su, L. Jin, T.G. Schurr, et al., The Dual Origin and Siberian Affinities of Native American Y Chromosomes, Am. J. Hum. Genet 70 (2002) 192–206, https://doi.org/10.1086/338457.

[35] T. Karafet, S.L. Zegura, J. Vuturo-Brady, O. Posukh, L. Osipova, V. Wiebe, et al., Y chromosome markers and Trans-Bering Strait dispersals, Am. J. Phys. Anthr. 102 (1997) 301–314, https://doi.org/10.1002/(SICI)1096-8644(199703)102:3<301::AID-AJPA1>3.0.CO;2-Y.

[36] C. Núñez, M. Geppert, M. Baeta, L. Roewer, B. Martínez-Jarreta, Y chromosome haplogroup diversity in a Mestizo population of Nicaragua, Forensic Sci. Int.: Genet. 6 (2012) e192–e195, https://doi.org/10.1016/j.fsigen.2012.06.011.

[37] P.D. Curtin, The atlantic slave trade: a census, UW Press (1972), https://doi.org/10.2307/1159486.

[38] J. Jannuzzi, J. Ribeiro, C. Alho, G. de O.L. e Arão, R. Cicarelli, H.S.D. Corrêa, et al., Male lineages in Brazilian populations and performance of haplogroup prediction tools, Forensic Sci. Int.: Genet. (2020) 44, https://doi.org/10.1016/j.fsigen.2019.102163.

[39] T. Palha, L. Gusmão, E. Ribeiro-Rodrigues, J.F. Guerreiro, Â. Ribeiro-dos-Santos, S. Santos, Disclosing the genetic structure of brazil through analysis of male lineages with highly discriminating haplotypes, PLOS ONE 7 (2012), e40007, https://doi.org/10.1371/journal.pone.0040007.

[40] U. Toscanini, F. Brisighelli, C. Llull, G. Berardi, A. Gómez, F. Andreatta, et al., Charting the Y-chromosome ancestry of present-day Argentinean Mennonites, J. Hum. Genet. 61 (2016) 507–513, https://doi.org/10.1038/jhg.2016.3.

[41] L. Roewer, M. Nothnagel, L. Gusmão, V. Gomes, M. González, D. Corach, et al., Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in native South Americans, PLoS Genet 9 (2013), e1003460, https://doi.org/10.1371/journal.pgen.1003460.

[42] I.B. Ambrosio, D.F. Braganholi, L.B.M. Orlando, N.C. Andrekenas, I. da Mota Pontes, D.A. da Silva, et al., Mutational data and population profiling of 23 Y-STRs in three Brazilian populations, Forensic Sci. Int Genet 48 (2020), 102348, https://doi.org/10.1016/j.fsigen.2020.102348.

[43] C. Vullo, V. Gomes, C. Romanini, A.M. Oliveira, O. Rocabado, J. Aquino, et al., Association between Y haplogroups and autosomal AIMs reveals intra-population substructure in Bolivian populations, Int J. Leg. Med 129 (2015) 673–680, https://doi.org/10.1007/s00414-014-1025-x.

[44] F. Simão, J. Ribeiro, C. Vullo, M. Catelli, V. Gomes, C. Xavier, et al., The ancestry of Eastern Paraguay: a typical South American profile with a unique pattern of admixture, Genes 12 (2021) 1788, https://doi.org/10.3390/genes12111788.

[45] Association (WMA) WM. Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. Jahrbuch Für Wissenschaft Und Ethik 2009; 14:233–8. https://doi.org/10.1515/9783110208856.233.

[46] M. Brion, B. Sobrino, A. Blanco-Verea, M.V. Lareu, A. Carracedo, Hierarchical analysis of 30 Y-chromosome SNPs in European populations, Int J. Leg. Med 119 (2005) 10–15, https://doi.org/10.1007/s00414-004-0439-2.

[47] M. van Oven, K. Toscani, N. van den Tempel, A. Ralf, M. Kayser, 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 34 (2013) 3029–3038, https://doi.org/10.1002/elps.201300210.

[48] V. Gomes, P. Sánchez-Diz, A. Amorim, A. Carracedo, L. Gusmão, Digging deeper into East African human Y chromosome lineages, Hum. Genet 127 (2010) 603–613, https://doi.org/10.1007/s00439-010-0808-5.

[49] R. Resque, L. Gusmão, M. Geppert, L. Roewer, T. Palha, L. Alvarez, et al., Male Lineages in Brazil: intercontinental admixture and stratification of the European background, PLoS One 11 (2016), e0152573, https://doi.org/10.1371/journal.pone.0152573.

[50] M. González, V. Gomes, A.M. López-Parra, A. Amorim, A. Carracedo, P. Sánchez-Diz, et al., The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88, Eur. J. Hum. Genet 21 (2013) 324–331, https://doi.org/10.1038/ejhg.2012.167.

[51] M.S. Jota, D.R. Lacerda, J.R. Sandoval, P.P.R. Vieira, D. Ohasi, J.E. Santos-Jnior, et al., New native South American Y chromosome lineages. Jinrui Idengaku Zasshi The Japanese, J. Hum. Genet. 61 (2016) 593–603, https://doi.org/10.1038/jhg.2016.26.

[52] M.S. Farzad, B.M. Pedersen, H.S. Mogensen, C. Børsting, Development of an automated AmpliSeqTM library building workflow for biological stain samples on the Biomek® 3000, Biotechniques 68 (2020) 342–344, https://doi.org/10.2144/btn-2019-0156.

[53] S. van der Heijden, S.J. de Oliveira, M.-L. Kampmann, C. Børsting, N. Morling, Comparison of manual and automated AmpliSeqTM workflows in the typing of a Somali population with the Precision ID Identity Panel, Forensic Sci. Int.: Genet. 31 (2017) 118–125, https://doi.org/10.1016/j.fsigen.2017.09.009.

[54] A. Ralf, D. Montiel González, K. Zhong, M. Kayser, Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data, Mol. Biol. Evol. 35 (2018) 1291–1294, https://doi.org/10.1093/molbev/msy032.

[55] M. van Oven, A. Van Geystelen, M. Kayser, R. Decorte, M.H.D. Larmuseau, Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome, Hum. Mutat. 35 (2014) 187–191, https://doi.org/10.1002/humu.22468.

[56] N. Rivera Franco, Y. Braga, M. Espitia Fajardo, G. Barreto, Identifying new lineages in the Y chromosome of Colombian Amazon indigenous populations, Am. J. Phys. Anthr. 172 (2020) 165–175, https://doi.org/10.1002/ajpa.24039.

[57] M. Regueiro, J. Alvarez, D. Rowold, R.J. Herrera, On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists, Am. J. Phys. Anthr. 150 (2013) 333–348, https://doi.org/10.1002/ajpa.22207.

[58] P. Gampawar, Y. Saba, U. Werner, R. Schmidt, B. Müller-Myhsok, H. Schmidt, Evaluation of the performance of AmpliSeq and SureSelect exome sequencing libraries for ion proton, Front. Genet. 10 (2019) 856, https://doi.org/10.3389/fgene.2019.00856.

[59] C. Posth, N. Nakatsuka, I. Lazaridis, P. Skoglund, S. Mallick, T.C. Lamnidis, et al., Reconstructing the deep population history of Central and South America, Cell 175 (1185–1197) (2018), e22, https://doi.org/10.1016/j.cell.2018.10.027.

[60] S.T. Sherry, M.H. Ward, M. Kholodov, J. Baker, L. Phan, E.M. Smigielski, et al., dbSNP: the NCBI database of genetic variation, Nucleic Acids Res 29 (2001) 308–311, https://doi.org/10.1093/nar/29.1.308.

[61] G.D. Poznik, Y. Xue, F.L. Mendez, T.F. Willems, A. Massaia, M.A. Wilson Sayres, et al., Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences, Nat. Genet 48 (2016) 593–599, https://doi.org/10.1038/ng.3559.

[62] P. Hallast, C. Batini, D. Zadik, P. Maisano Delser, J.H. Wetton, E. Arroyo-Pardo, et al., The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades, Mol. Biol. Evol. 32 (2015) 661–673, https://doi.org/10.1093/molbev/msu327.

[63] T. Pinotti, A. Bergström, M. Geppert, M. Bawn, D. Ohasi, W. Shi, et al., Y chromosome sequences reveal a short beringian standstill, rapid expansion, and early population structure of Native American Founders, Curr. Biol. 29 (149–157) (2019), e3, https://doi.org/10.1016/j.cub.2018.11.029.