Research paper

Full mtDNA genome sequencing of Brazilian admixed populations: A forensic-focused evaluation of a MPS application as an alternative to Sanger sequencing methods

E. Avila[a,b,c,*], P. Graebin[b], G. Chemale[a], J. Freitas[d], A. Kahmann[e], C.S. Alho[b,c]

[a] Setor Técnico-Científico, Superintendência Regional do Rio Grande do Sul, Polícia Federal, Porto Alegre, Brazil
[b] Escola de Ciências, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, Brazil
[c] Instituto Nacional de Ciência e Tecnologia INCT Ciências Forenses, Porto Alegre, Brazil
[d] Instituto Nacional de Criminalística, Polícia Federal, Brasília, DF, Brazil
[e] Instituto de Matemática, Estatística e Física, Universidade Federal de Rio Grande, Rio Grande, Brazil

ABSTRACT

The use of Massive Parallel Sequencing (MPS) techniques have been proposed by the forensic community as an alternative to Sanger sequencing methods in routine forensic casework analysis regarding mitochondrial DNA (mtDNA). Interesting features of MPS include high throughput, ability to simultaneously genotype a significant number of samples by barcoding techniques, processing automation, reduced time and costs, among others. Advantages include the capability of generating full mtDNA genome sequences versus usual techniques, usually limited to hypervariable or control regions exclusively. In this work, 96 reference single-source samples from three different Brazilian cities were subjected to full mtDNA genome sequencing by MPS techniques using an early-access version of Precision ID mtDNA Whole Genome Panel on an Ion Torrent PGM platform (Thermo Fisher Scientific, Waltham, MA, USA). Complete, high-quality sequences were obtained and sequencing performance was evaluated via four different metrics. As a subset of evaluated samples have been previously submitted for Sanger sequencing of the control region, a comparative analysis of both methods' results was conducted in order to compare technique adequacy within a forensic context. Even though this study is one of the first to report full mtDNA genome sequences for Brazilian admixed populations, the observed haplotypes exhibit a predominance of Native American and African maternal lineages in the studied sample set, reproducing results described in the literature for control regions only. Interpopulation analysis among Brazilian and 26 worldwide populations was also carried out. The results indicate that MPS-generated full mtDNA genome sequences may have great utility in forensic real casework applications, with a pronounced gain of genetic information and discrimination power provided by coding region evaluation and the enhanced capacity of heteroplasmies determination. Database construction and other relevant factors concerning implementation of such techniques in Brazilian forensic laboratories are also discussed.

1. Introduction

The investigative use of mitochondrial DNA (mtDNA) genome sequencing has been widely employed worldwide in forensic laboratories, especially in cases involving biological evidence presenting critical conditions, including extremely low quantities, high degradation levels or ancient DNA, where short tandem repeat (STR) analysis might have limited efficiency [1]. Owing to the fact that human cells have multiple mtDNA genome copies, generally surpassing nuclear DNA quantities by

a factor of thousands, evaluation of maternal lineage markers is sometimes the only possible technical approach to obtain genetic information from biological evidence [2]. However, limitations to this method include a lower individualization power compared to autosomal markers, taking into consideration individuals from a single maternal ancestry line may share a unique DNA haplotype. The usual analytical approach to forensic mtDNA analysis consists of Sanger sequencing (SS) capillary electrophoresis (CE) applications, aimed at identifying nucleotide polymorphism events in any of the three highly

variable regions comprising most nucleotide polymorphisms, located inside a non-coding portion of the mitochondrial genome close to the replication origin, called the control region (CR) [3]. Hyper-variable regions I and II (HV-I and HV-II, respectively) encompass most of the individual variation among samples, even those belonging to the same haplotype, and as such are considered the most informative region of the mtDNA genome [4]. However, its complete sequence comprises over 16k base pairs, and ˜75% of total reported polymorphic variants are observed in nucleotide positions located in the coding region, usually not included in forensic mtDNA examination based on limitations inherent to SS applications [5,6].

Massive Parallel Sequencing (MPS) applications have been proposed in order to circumvent SS and CE technical restrictions in mtDNA full-genome analysis [7]. As a consequence of its high-throughput capacity and workflow automation solutions, this technique is able to provide simultaneous sequencing of a large number of samples through bar-coding use with consequent reduced analysis time and significant gains in efficiency and the amount of information obtained from a criminal sample [7]. Efficient MPS of full mtDNA genomes has been achieved with various experimental designs [8–14]. However, it has been put forth that panel designs for forensic applications should account for the possibility of high degradation levels presented by certain forensic samples [15]. Therefore, among proposed solutions is the use of a large number of smaller amplicons, an analytical strategy which should provide efficient amplification of degraded DNA. The Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific, Waltham, MA, USA) commercial solution offers simultaneous amplification of 162 small, overlapping amplicons comprising a full mtDNA genome. It was designed to generate small-sized amplicons and successfully allow processing of highly degraded, low-input and other challenging forensic samples. Validation studies show that robust, sensitive, efficient and reliable typing of forensic samples can be accomplished with different versions of this product, allowing its use in forensic individual identification or parentage testing [5,15–17].

The modern Brazilian population features a high degree of genetic admixture, products of recurrent miscegenation processes occurring throughout national history [18,19]. European (EUR) settler (from the Iberian Peninsula mainly) influx has occurred since the XVI century, with a notable gender bias toward masculine contributions. Such migrant men have interbred with local Native American (NA) populations, especially during the initial colonization phases. Later in Brazilian history, the slavery work force (then widely accepted in Portuguese society) adoption motivated forced transport of large contingents of African (AFR) individuals, with special relevance of Western African Portuguese colonies, settlements or commercial outposts (mostly located in Western Africa, with some minor Eastern Africa participation). Late migratory processes took place in the 19th and 20th centuries, as governmental policies for territory occupation and national economic growth promoted arrival of significant amounts of Asian (essentially from Japan and Mid-Western countries, as Lebanon and Syria, part of the then-existing Ottoman Empire) and Central and Southern European (predominantly Germans, Italians, Portuguese and Spaniards) individuals. Such diverse biogeographical groups shaped the current Brazilian population, forging a genetic and culturally diverse, highly admixed and multiethnic country [20].

The significant Brazilian territorial extension was consolidated through several distinct inland migratory flows, with expansion from coastal areas to continental interior direction following distinct patterns in spatial or temporal territorial occupancy. Multidirectional introgression was guided by social and historical conditions, which significantly varied for each distinct geographical region [20]. As a result of heterogeneous migratory flow dynamics, ethnic composition of regional or local groups may display distinct proportions of parental populations contribution (NA, EUR and AFR) among geopolitical regions. [21]. This phenomenon can be noticed from a genetic standpoint, but is also evidenced through contrasting individual self-declared ethnic or ancestry ratios among regional groups, with a relatively higher NA and AFR influence in Northern and Northeastern Brazil, respectively, while Southern Brazil exhibits a prevalent EUR ancestry [22]. Parental lineage markers clearly display a more markedly asymmetrical distribution of ancestral populations contribution to the current genetic pool [23–25]. Historical and cultural factors motivating admixture mechanisms in Brazil yielded current overall preponderance of the EUR Y-chromosome paternal contribution in contrast to predominance of AFR and NA mitochondrial DNA (mtDNA) maternal lineages [23–25], even when regional idiosyncrasies are considered.

In this work, an evaluation of Precision ID mtDNA Whole Genome Panel sequencing efficiency in 96 Brazilian samples was put forth. This commercially available product was designed for forensic (among other) applications, and the present analysis discusses not only technical considerations (including sequence performance) related to the product, but also some aspects involving convenience in this commercial solution adoption by Brazilian forensic laboratories. A comparative analysis with the current widely used mtDNA technique in Brazil (SS followed by CE of mitochondrial HVRs or CR) is also presented. The current work has a strict focus on single-source reference samples for the Brazilian population. Adoption of proposed analytical alternatives in current forensic casework also demands protocol validation with real evidence obtained from crime scenes. Several studies [5,15–17,26] have already evaluated present commercial solution (as well as posterior versions of the same product) efficiency in mock or real forensic cases with remarkable success. Some particular issues intrinsically linked to the very nature of biological evidence in criminal DNA analysis have also been addressed by several research groups regarding mtDNA evaluation, as mixture occurrence or samples presenting highly degraded DNA or DNA template low-copy numbers [5,15–17,26]. The reported results suggest this panel is a viable alternative for forensic mtDNA full-genome sequencing, with reliability and efficiency in accordance with forensic quality standards.

2. Materials and methods

2.1. Ethical statement

DNA samples included in the present study were obtained from voluntary adults following written informed consent. All selected subjects are part of a research-purposed human samples biobank available in PUC/RS comprised of Brazilian samples. This work was approved by the Pontifical Catholic University of Rio Grande do Sul Institutional Review Board under CAAE 52113715.9.0000.5336 number, and follows the ethical principles stated in the Helsinki Declaration [27] of the World Medical Association.

2.2. Samples description

Oral swabs or peripheral blood were obtained from 96 unrelated volunteer donors from three different Brazilian cities located in distinct geopolitical regions. Subjects from Porto Alegre (Southern Brazil, n = 9), Rio de Janeiro (Southeastern Brazil, n = 3) and Distrito Federal (Center-Western Brazil, n = 84) were evaluated. Sample size for individuals from Distrito Federal was chosen to be larger than other regions based on specific features presented by this population [22]. Besides, this particular sample set from Distrito Federal was thoroughly investigated for a significant number of genetic markers [23,28–30], including a series of forensically relevant polymorphisms, and therefore can be used for a comprehensive characterization of Brazilian genetic complexion. Aiming to contrast sequencing performance and forensic-focused data analysis between gold standard techniques (Sanger method) and MPS results, our sample set includes 65 individuals from Distrito Federal (of a total of 84) who had their mtDNA genome CR sequences described in a precedent report [28].

2.3. DNA extraction and quantitation

Total DNA was extracted using the automated platform Automate Express Forensic DNA Extraction System (Thermo Fisher Scientific Inc.) with the Prepfiler Automated Forensic DNA Extraction System (Thermo Fisher Scientific Inc.), or manually via the DNA IQ System (Promega Corp., Madison, WI, USA). Autosomal DNA was quantified with the Quantifiler Human DNA Quantification Kit (Thermo Fisher Scientific Inc.) on a 7500 Real-Time PCR System (Thermo Fisher Scientific Inc.). MtDNA quantities were estimated using a 1:100 genomic to mitochondrial DNA copies ratio according to the product manual. All procedures followed the manufacturer´s instructions.

2.4. Library preparation, emulsion PCR, and sequencing

DNA libraries were constructed using the Ion AmpliSeq Library Kit 2.0 (Thermo Fisher Scientific Inc.) combined with an early-access version of Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific Inc.) v.2.2 following the manufacturer´s suggested protocol (Early Access Ampliseq Mitochondrial Sequencing User Manual, May 2015 Revision), applying the suggested Conservative Library Preparation Method. Briefly, PCR amplification was performed in two different reactions per sample, each one using 5 μL of Ion Precision primer pool A or B (Mitochondrial Sequencing Panel) at a final volume of 10 μL containing 1 ng of template DNA and 2 μL of 5x Ion AmpliSeq HiFi Mix. Total reaction volumes were adjusted with nuclease-free water. PCR was performed in a Veriti 96-well Thermal Cycler (Thermo Fisher Scientific Inc.) under the following conditions: enzyme activation for 2 min at 99 °C, 21 cycles of 15 s at 99 °C and 4 min at 60 °C followed by a 10 °C hold. At this point, both amplified mitochondrial PCR reactions for each sample were combined in a single 20 μL pool, where PCR amplicons were partially digested with 2 μL FuPa reagent following incubation at 50 °C for 10 min, 55 °C for 10 min and 60 °C for 20 min, all followed by a 1 h hold at 10 °C. Adaptor ligation to the libraries was executed according to the manufacturer´s instructions using different barcodes for each sample in a single run (Ion Xpress Barcode Adaptors 1–96 Kit or IonCode Barcode Adapters 1–384 Kit) (Thermo Fisher Scientific Inc.), and the resultant products were purified manually with Agencourt AMPure XP reagents (Beckman Coulter Inc., Brea, CA, USA) along with a DynaMag 96-Side Magnetic Rack according to the manufacturer's protocol. After purification, libraries were quantified employing a 7500 Real-Time PCR System (Thermo Fisher Scientific Inc.) with the Ion Library TaqMan™ Quantitation Kit (Thermo Fisher Scientific Inc.) as described in product manual. Libraries were pooled to a final concentration of 20 pM. Emulsion PCR (emPCR) was carried out on the Ion OneTouch 2 instrument (Thermo Fisher Scientific Inc.) with the Ion PGM Hi-Q Template Kit (Thermo Fisher Scientific Inc.) following the manufacturer's protocol (Ion PGM Hi-Q Template Kit, Revision A.0). The emPCR products were enriched on the Ion OneTouch Enrichment System (Thermo Fisher Scientific Inc.). A final volume of 30 μl was loaded per chip. Sequencing was performed on the Ion Torrent PGM™ (Thermo Fisher Scientific Inc.) Sequencer with the Ion PGM Hi-Q Sequencing Kit (Thermo Fisher Scientific Inc.) following the manufacturer´s instructions (Ion PGM Hi-Q Sequencing Kit, Revision B.0) and using Ion 318 Chip Kit v2 (Thermo Fisher Scientific Inc.). Positive and negative controls were evaluated in this study with the results also utilized for internal validation of reagents. A total of three chips were used in distinct runs for complete sample-set genotyping.

2.5. Data analysis and mtDNA haplogrouping

Sequencing data were analyzed using the Torrent Suite Software v5.0 (Thermo Fisher). A modified version of the Revised Cambridge reference sequence (rCRS + 80, based on Genbank NC_012920.1) [31], including repetition of the first 80 nucleotides at the mtDNA sequence end, was employed as reference genome data for alignment. The

number of mapped reads was calculated by the Torrent Coverage Analysis v5.0 plugin (Thermo Fisher Scientific Inc.). Individual differences from the reference sequences were defined as called by Ion Torrent Variant Caller v5.0.0.7 plugin with the target regions file (rCRS_plus_80nts_mt_targets) and no designated hotspot regions file under default analysis settings. Minimum read number was set for six reads per base position. As the plugins used did not follow current nomenclature adopted by the forensic community [32–34], corrections were applied to variant reports, including revisions of the proposed alignment in identified polymorphisms (where differences from the reference sequence were aligned as close to the 3´ end of DNA as possible, rather than the usually reported 5´-end placement obtained with the employed plugins). MtDNA sequences were independently reviewed by two different collaborators through BAM file visualization with IGV (Integrative Genomes Viewer v2.3.80) software [35] and manual corrections were applied. mtDNA haplogroups were determined using Haplogrep 2 v2.1.13 [36] according to mtDNA phylotree Build 17 (available online at http://www.phylotree.org/) [37] and confirmed through EMPOP database v4/R11 [38]. The identified length heteroplasmies were not reported. Point heteroplasmy threshold was set to 15%, and variants above this number were reported in the final results. Sequence quality control was performed by EMPOP tools [38], and the generated data was submitted to the EMPOP database as rCRS variations, deposited under reference number EMP00750.

2.6. Sequencing performance and statistical data analysis

Precision ID mtDNA Whole Genome Panel sequencing performance was evaluated with four different statistical parameters: Read Depth (RD), equivalent to obtained number of reads per amplicon; Amplicon Balance (AB), which assesses individual amplification efficiency and is calculated as the number of reads for each amplicon, divided by the mean RD of all amplicons, per sample; Strand Balance (SB) that measures the balance between forward and reverse reads and was calculated as follows: most prevalent strand read number (forward or reverse) was divided by total read number for allele reads only. Expected values indicating no strand bias was equal to 0.5. The results are presented as deviations from this ideal value. Noise Level (NL) represents the amount of unspecific base call, calculated as the ratio of non-allele calls at a nucleotide position divided by total number of reads at that same nucleotide position. SB and NL were evaluated for reported variants only. Statistical evaluation of read-depth differences among distinct amplicons or primer pools was executed using Microsoft Excel 2016 (Microsoft Corporation, Redmond, USA) or R statistical packages. Additional information on the employed formulae is provided in Supplementary Material M1.

2.7. Interpopulation analysis

Haplotype frequencies, molecular diversity indices and other relevant population parameters were calculated for the sample set in question using Arlequin v3.5.2.2 [39] software. The same program was made use of to perform Analysis of Molecular Variance (AMOVA) and pairwise F_{ST} testing among Brazilian population and 26 worldwide populations. Populations employed in the present study were obtained from the 1000 Genomes Project (1kGP - Phase 3; available online at http://www.internationalgenome.org/data/) as VCF variant files [40]. Samples with previously described parental or fraternal relations sharing a unique haplotype had only one individual from each familial group included in this study. Variants were converted to full mitosequence FASTA files using Haplogrep 2 online software, and realigned against the Brazilian sample set using MAFFT v7 multiple sequence online alignment tool (available online at https://mafft.cbrc.jp/alignment/server/) under standard parameters [41]. Multidimensional Scaling Analysis (MDS) based on pairwise F_{ST} distances was executed using IBM SPSS Statistics Software v.17.0 (IBM Analytics
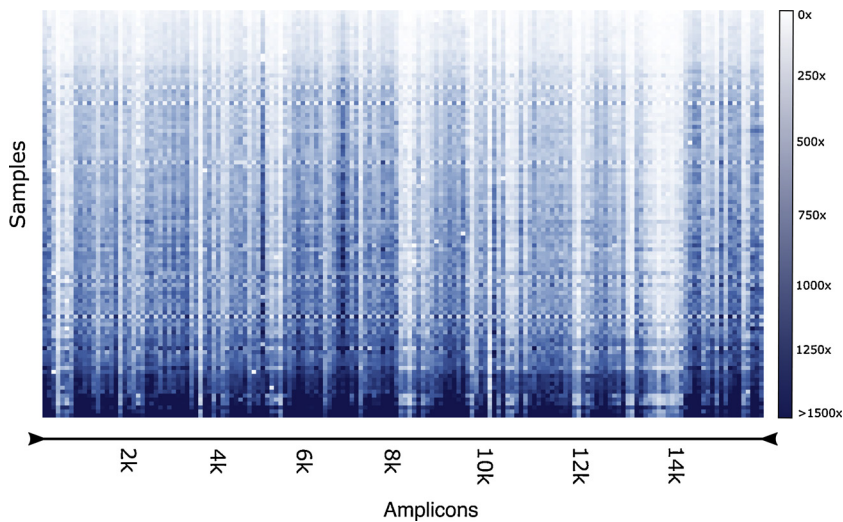
Fig. 1. Read Depth heat map of 162 amplicons included in Precision ID mtDNA Whole Genome Panel for 96 Brazilian samples. Samples were arranged (top to bottom) according to mean number of reads for all amplicons, in ascending order. Amplicons are distributed (left to right) in same order as mtDNA genome covered region (as displayed in Supplementary Table S2).

Inc., Chicago, USA). Populations used in this study, their metadata and employed abbreviations are detailed in Supplementary Table S1. Pairwise F_{ST} distances using Brazilian and worldwide data were also utilized to build cladogram graphics using the Molecular Evolutionary Genetics Analysis v7.0 (MEGA v.7.0) software [42], applying Neighbor-Joining (N-J) methods.

3. Results

The Precision ID mtDNA Whole Genome Panel was successfully used to generate full genome sequences for 96 Brazilian samples from three different cities located in distinct geopolitical regions. This commercial kit comprises 162 primer pairs amplified in two distinct multiplex reactions (A and B pools), generating a series of short, overlapping mtDNA amplicons covering the full mitochondrial genome. Supplementary Table S2 details mtDNA regions associated with each amplicon. Such a sequencing strategy has been demonstrated to be useful for successful amplification of ancient, highly degraded, mixed and other challenging forensic samples [5,15–17,26]. Therefore, the present kit was designed to properly handle critical or difficult forensic samples, although such an approach may be fraught with setbacks particularly related to nonspecific amplification and detection of nuclear copies of mtDNA sequences (also known as numts) [6,15,43,44]. The obtained results show an average amplicon size for all amplicons equivalent to 105 bp. Brazilian population haplotype distribution and frequencies, as well as other relevant descriptive indices connected with forensic applications of mtDNA analysis and sequencing performance, were investigated in the present study in order to assess distribution of mtDNA haplotype frequencies for the Brazilian population and evaluate the adequacy of the present panel introduction in Brazilian routine forensic casework.

3.1. MPS mtDNA sequencing performance

In order to properly assess sequencing efficiency using Precision ID mtDNA Whole Genome Panel, a series of analytical metrics were employed as described in the Material and Methods section. All samples in the present study were genotyped through Ion PGM Sequencer, employing native data interpretation solutions provided by the manufacturer. The Variant Caller plugin was the main pipeline for data calling. Additionally, the Coverage Analysis plugin was utilized to access or calculate sequencing performance metrics for all genotyping procedures. As the maximum number of samples suggested by the manufacturer in a single run is smaller than total investigated sample size, all 96 samples were analyzed in three independent runs along with

other samples not included in this study, totaling no more than 48 samples per chip. Data regarding each individual experiment is presented in Supplementary Table S3. The best results were observed for chip mt3, where the slightly smaller number of samples included in the run and superior number of usable reads resulted in improved overall mean depth per sample. The number of usable reads obtained in first two experiments was severely affected by an unusually high polyclonal ISPs ratio, which was later corrected during the third run's execution. Performance improvements were achieved by using freshly opened reagents in library purification and ISP enrichment steps. In order to assess technique reproducibility, two specific samples were included in all three runs, and the resulting sequences showed no differences in sequence obtained for these samples across all three distinct experiments. In a similar manner, the employed positive controls also displayed complete concordance for all three distinct runs with reference sequences. Considering overall sequencing performance, features presented by control samples were found to be similar to other Brazilian samples, including read depth and other relevant elements.

3.1.1. Amplicon-based metrics evaluation

For the adopted protocol, inclusion of 48 samples in a single run using 318v2 model chips is proposed to ensure a mean depth of at least 400x per locus. Despite that, mean depth per sample ranged from 40x to 2,587x, with an average value of 430x (s.d. 392x), which can be considered satisfactory for reliable genotyping of single-source reference samples, where the proposed depth of as low as 10 reads might be considered sufficient for genotype determination [17]. Similar mean depth results were obtained with the same commercial kit for heat degraded samples [5], suggesting reliable use of real forensic casework evidence might be possible. Still, regarding RD metrics, when mean read depth is evaluated per amplicon instead of on a per sample basis, the observed values vary from 91x to 1032x, with average values of 430x (s.d. 185x). Supplementary Figure S1 depicts the mean read depth results obtained for each individual amplicon. This significant variance in amplification performance for different primer pairs has been previously reported, and was demonstrated to be sequence-dependent, where high CG content or homopolymeric regions performed comparatively worse than other regions [5]. Fig. 1 illustrates a heat map graphical representation of the complete sample set's (96 individuals) number of reads for each evaluated amplicon. Differences in RD values can be easily observed among distinct samples (depicted with horizontal lines), and relative amplification performance of each amplicon (vertically represented) can also be evaluated. Low-performance regions and amplicons, identified with vertical lines featuring lighter color shades, display smaller RD numbers, and detected variation

roughly reproduces results obtained with this reagent by other research groups [5,6,15–17,45]. To further explore these findings, less efficient amplicons were characterized using a previously proposed value of 100x for average read depth as a threshold for low-performing amplicons [5]. In our analysis, only three amplicons were found to have amplification levels under this value: amplicons mt_4, mt_101 and mt_139. All three amplicons were reported as challenging, with the former associated with inconsistencies during base calling across two different MPS systems [6] and the latter two identified as endowed with low amplification efficiency in previous studies [5,46]. It is important to note that some samples present a peculiar checkerboard pattern, with intercalated high and low depth values across the entire mtDNA genome. This feature is explained by discrepancies in PCR efficiency attained by the two different primer pools, A and B, employed, where uneven amplification performance leads to higher RD values for amplicons included in a specific pool when compared to its counterpart. This variation in efficacy between both pools comprising this kit have been previously described [6], and it was even considered to be instrument-dependent (where different equipment models or manufacturers consistently display preferential amplification of pool A or B in detriment of the other. However, even though variations between both primer pool´s quality were observed for certain samples in our results, no evidence of directional favoring of any of such pools was noted (0.46 *p-value* on a two-tailed binomial test). Thus, observed fluctuations in depth numbers may be explained by stochastic factors or inconsistencies in library preparation, as this latter step was carried out manually, considering no automation alternative was available in our laboratory.

Concerning AB values, a simple visual exam of Fig. 1 and Supplementary Fig. 1 allows for the conclusion that a significant variation in PCR efficiency among amplicons is observed. Supplementary Fig. 2 (A and B) presents AB values for all 162 amplicons, divided by primer pair pools included in the kit (A and B, respectively). As shown in Supplementary Fig. 2, AB median values for amplicons included in pool A ranged from 0.18 at mt_7 and 2.6 at mt_135 amplicons. For pool B, this statistic varies from 0.16 at mt_116 amplicon to 1.8 at mt_132. The overall average of AB values for each sample, considering all 162 amplicons, is 1.00. Thirty-one amplicons in total (mt_7, 11, 13, 35, 71, 103, 107, 143 and 161 for pool A and mt_2, 4, 10, 32, 40, 48, 50, 52, 56, 78, 80, 102, 104, 112, 114, 116, 118, 120, 122, 124, 126 and 154) exhibited poor AB performance with values under the 0.50 threshold, meaning the number of reads for these 31 regions are below half the average RD per sample. This low performance issue is minimized by overlapping design of primers included in both pools, where sequences located in both extremities for underachieving amplicons can be partially compensated by flanking amplicons in an alternative primer pool. However, a specific region comprising amplicons mt_102 to mt_104 (bases 10481 to 10775 in rCRS) was found to be particularly challenging as low RD values were obtained for tandem located amplicons. Homopolymeric regions have been proposed as the main cause for lower efficiency in Ion PGM and other platforms [14,45], possibly because of inaccurate flow-call and pH fluctuation detection in proton-based systems. As no significant homopolymeric stretches are found in this particular sequence, other causes might be implicated. Additional factors, such as sample quality and G/C content, have been demonstrated to have effects on base-calling accuracy [46], and some of them might be directly impacting amplification efficiency in this particular region. Therefore, a careful evaluation of suggested base calls and meticulous interpretation of detected variants is particularly important for regions described as problematic or demonstrating lower quality results under this analytic protocol, especially in forensic cases where critical biological evidence is involved, including challenging samples presenting degraded or a low quantity of DNA. Efforts to achieve improvement in primer balance and efficiency might also be considered by manufacturers in updated kit versions in order to obtain products suitable for reliable and trustworthy sequencing of mtDNA in real forensic

casework.

3.1.2. Polymorphism-based metrics evaluation

Regarding SB results, overall performance of this panel was robust, with 58% of amplicons (94 total) displaying SB median values under the 0.1 threshold. However, roughly 14.2% (23 total) of primer pairs resulted in amplified fragments where at least one identified variation within such regions exhibits high imbalance, with observed SB median values above 0.4. Supplementary Figure S3A depicts a boxplot representation of results obtained for SB values for all detected variants identified in the evaluated samples organized by amplicon. Not only median values exhibit high heterogeneity among amplicons, but also variance presents a very broad distribution among amplicons (with standard deviation values ranging from 0.011 to mt_133 to 0.242 to mt_79 for amplicons with at least five variations detected for all samples). The impact of amplicons presenting low performance for SB metrics in forensic cases may be significant, especially in distinguishing real occurring variations in overlapped regions amplified by more than a single primer pair. Base call errors have been reported as a consequence of extreme strand imbalance, even though strand bias is suggested to be inherent to DNA sequence properties [5]. Therefore, visual analysis of BAM files for rCRS differences presenting significant strand imbalance is important for evaluating if such observations are real or an artifact originating from panel design. Finally, Supplementary Figure S3B exhibits NL measures distribution for all amplicons containing variants in our sample set. Overall, NL can be considered low with all median results under 5% noise-to-signal ratio. This result is consistent with previous studies [5,6,15–17,45] and consistent with high-quality demands involved in forensic DNA genotyping or sequencing. Only two amplicons (mt_15 and mt_86) presented median NL levels greater than 4%. Noise reads can be resulting from sequencing errors, polymerase errors in PCR, alignment mistakes and contamination by alternative DNA or numts sequences [17]. Analysis of NL not only in these two particular regions, but for most of the observed individual variants with increased base miscalling ratios, indicates that most of this problem is caused by the presence of nearby small or medium-sized homopolymeric regions (presenting three- to six-nucleotide stretches of a unique base) [17]. Even though a proportion of this observation might be caused by real-length heteroplasmic events, it is more likely that inaccurate flow-call or pH fluctuation detection, a feature inherent to Ion PGM system technology, might be responsible, at least partially, for reported NL. Visual inspection of BAM files allowed for identification of a significant number of small sequences (usually less than 40 nucleotides long) inside mtDNA, thereby presenting high noise levels (not depicted in Supplementary Figure S3B as such polymorphisms were not identified as rCRS variants) and usually distributed in several different concentrated areas. Such regions, which were found to be present in the same overall mtDNA locations throughout all evaluated samples, are widely distributed all over the mtDNA genome. Despite no specific assessments being performed to fully characterize these regions, such polymorphisms are probably associated with autosomal numt sequences, which are also amplified along with actual mtDNA as a consequence of a multiple, short, overlapping amplicon-sequencing strategy adopted in this panel. However, our data analysis suggests that these deviant polymorphisms were amplified (when compared to actual mitochondrial sequences) at a rate insufficient to significantly affect automatic base-calling performed by algorithms included in employed plugins. Previous works [6,15,43,44] have posited alternative approaches to verify the presence of numts impact in the present panel's sequencing efficiency as employing alignment of MPS sequences against the full human genome to confirm numts amplification by this kit in nuclear DNA. However, the same reports also indicate that numts co-amplification has a limited impact in rCRS variant identification. Yet, special attention to such regions is necessary when forensic samples are being evaluated, and this considers that common occurrence of DNA mixtures, environmental

contamination, low DNA quantity and/or quality or DNA degradation levels can be mistakenly identified as numts or have its genetic features marked or influenced by this very same numts presence in evaluated samples. The results indicated that sequencing performance of the Precision ID mtDNA Whole Genome Panel was usually favorable, with accuracy, reproducibility and efficiency adequate for forensic applications. Certain problems observed in specific amplicons have, in a general manner, no major influence in proper mtDNA full sequencing and rCRS variant determination. However, questions that must be addressed by the manufacturer include continuous development and revision of plugins along with base-calling algorithms in order to provide reliable, easy-to-use solutions that can be incorporated into analytical resources available in forensic genetics laboratories. Analytical tools and computational resources considering the nature of forensic analysis, including but not limited to uncertain sample quality or conditions, mixture occurrence, automation or need for high throughput and idiosyncrasies related to forensic data reports need to be provided, in order to maximize result reliability and facilitate data interpretation by analysts.

3.2. Genetic diversity and mtDNA haplogroup distribution in a Brazilian sample set

Successful full mtDNA genome sequencing was attained for 96 Brazilian individuals, and our experiment design showed that simultaneous amplification of a significant number of samples can be achieved with barcoding strategies. The capability of concurrent analysis of multiple sample can be a determining factor for technique adoption by forensic laboratories as reports of casework backlog and high-throughput needs are common [47]. Obtained haplotypes and haplogroups for all 96 samples are presented in Supplementary Table S4. When considering all 96 obtained Brazilian full mtDNA genome sequences, 94 different sequences were found, though we observed two distinct sample pairs presenting a shared, unique haplotype (belonging to B2 and L2a1a2c groups). The resulting haplotype diversity for all 96 samples is 0.9996 ± 0.0016, which is shown to be slightly superior than previously reported haplotype diversity for 306 individuals from Distrito Federal [28] for CR sequence only, but significantly less variable (0.9988 ± 0.7471). The number of polymorphic sites for the sample set is 737, with 668 transitions and 50 transversions observed. Mean number of pairwise differences among samples is 57.65 ± 25.12. Table 1 presents coarse haplotype frequency distribution for all Brazilian samples, including haplotype biogeographic origin. Complete refined, low-clade haplotype classification for each sample is reported in Supplementary Table S4.

Table 1 presents coarse haplogroup frequencies found in 96 unrelated individuals from the Brazilian population. When considering haplogroup frequencies in our sample set, the results obtained in this study reflects the general distribution of maternal lineages in the Brazilian population with a significant preponderance of AFR and NA mitochondrial DNA [25,28,48–55]. A significant gender bias and contribution imbalance has been reported for parental linages in the Brazilian population caused by a peculiar and uneven presence of EUR paternal lineages, along with elevated AFR and NA relative maternal contribution [21,23–25]. The two more frequent coarse haplogroups found in our experiment possess NA origin, specifically C1 (14.58%) followed by A2 (13.54%). In total, samples with NA precedence comprise 38.54% of all Brazilian samples evaluated. Africa was found to be the most frequent biogeographical origin of Brazilian maternal lineages, with 40.62% of the complete sample set. The most common coarse haplogroups of African ancestry were L1, L2 and L3, each with 12 identified samples. This finding agrees with historical events of the African slave trade and forced introduction of individuals from this continent in Brazil during the XVI and XIX centuries [54,55]. The remaining samples belong to groups with presumed biogeographical origin from other continents and worldwide populations. For instance, haplogroup H was the most common haplotype in the Brazilian population (11.47%) not belonging to Native or African lines. This clade is also the most frequent in Europe, especially in the Iberian Peninsula [56], also in agreement with Brazilian territorial occupation dynamics. Other unusual haplogroups with a single observation, as M5 or X2b4a, might reflect recent immigration (from XIX century on) of Asian (mainly Japanese) or Arabic (from then Ottoman Empire regions, like Lebanon and Syria) to Brazil. In this respect, full mtDNA genome sequencing provides an interesting tool to thoroughly analyze samples. As an example, we describe findings for sample RS09 classified in haplogroup J1b2. This lineage has a proposed semite origin with higher frequencies in the Middle East, Indo-Persian and other Arabian populations [57,58]. However, this specific sample presents mutations 6719C and 14927 G, which despite not being yet recognized as diagnostic for a distinct clade, have been closely associated to individuals with a Spanish or Iberian ancestry background [57,59], rather believed to have a European origin. Fernandes and collaborators [60] propose that these mutations can be characteristic of a new J1b2 subclade, called J1b2b. Therefore, despite a proposed Near-East origin for J1b2, this particular sample seems to belong to a new, undefined subclade with European origin that may be associated with populational groups in the Iberian Peninsula and their worldwide dispersal during the Great Navigations Age (from XV-XVI centuries on). This example is presented to illustrate the relevance of phylogenetic knowledge about worldwide mtDNA genome variations, and how biogeographical origin of already characterized or yet unknown subclades might be related to specific mutations, some of which are located in the mitochondrial coding region exclusively. Therefore, full mtDNA sequencing as a forensic tool can supply additional information with special consideration of highly ethnically diversified and admixed populations as Brazilian.

Most studies concentrating on Brazilian mitochondrial lines are limited to HV-I or HV-II sequence analysis, and therefore a more exhaustive analysis of results obtained in this study cannot be performed, especially concerning comparative evaluation of low-level mtDNA clades. However, the approximate frequencies of coarse haplogroup distribution mirrors previously reported results, with major AFR and NA contributions [25,28,48–55]. Supplementary Figure S4 features a complete phylogenetic tree including all samples and variants observed in the present study. Several elements supporting forensic casework analysis are based on consistently accepted phylogenetic data of worldwide mtDNA genome samples, including database construction and quality control procedures [61]. With that in mind, construction of a database of full mtDNA sequences based on local Brazilian populations is very relevant owing to the existence of unknown (or not yet

Table 1
mtDNA coarse haplotype frequencies based on 96 full mtDNA genomes from Brazilian population.

Coarse haplogroup classification	Haplotype Biogeographic Origin	Observed samples	Frequency (%)
A2	Native American	13	13.54
B2	Native American	5	5.21
C1	Native American	14	14.58
D	Native American	5	5.21
H	Indo-European	11	11.47
J1	Indo-European	1	1.04
K	Indo-European	1	1.04
L0	African	2	2.08
L1	African	12	12.5
L2	African	12	12.5
L3	African	12	12.5
L4	African	1	1.04
M5	Asian	1	1.04
U	Indo-European	4	4.17
V9	Indo-European	1	1.04
X2	Indo-European	1	1.04
Total	–	**96**	**100**

acknowledged as a unique clade) lineages. A large sample size for database composition is necessary in order to fully represent Brazilian mtDNA genetic diversity and assess any possible population substructure occurrence [62,63].

3.3. Comparative performance of MPS and Sanger sequencing in mtDNA genome haplotype determination and forensic resolution between CR x full mitogenome

A subset of total samples analyzed in the present study comprised of 65 individuals from Distrito Federal has been previously included in a Brazilian mitochondrial CR sequence database [28]. In order to evaluate overall sequencing performance and information gain provided by full sequence analysis, we compared the employed MPS technique results with current methods widely available in Brazilian forensic laboratories (SS). As described earlier, the additional information provided by inclusion of mitochondrial genome coding sequences in contrast to CR probing only was able to provide a higher haplotype diversity even with a significantly smaller sample size utilized in the present study (less than a third of the original total sample number). Even more significant than that, inclusion of extra genetic information was sufficient to discriminate two pairs of samples (DF05/DF84 and DF44/DF53) presenting distinct haplotypes, which were previously classified as sharing the same haplotype (inside each pair) considering CR genetic data exclusively. As the number of randomly selected samples included in this comparative analysis is relatively small (65), addition of mtDNA coding sequences to haplotype comparison was shown to increase relative discrimination capability significantly, therefore justifying its adoption in real forensic casework routine.

Haplotype classification differences observed between Sanger sequencing of CR and full mtDNA MPS were observed for 36 of 65 samples (55.4%). Such differences are expected and can be easily explained by additional information provided by coding region sequencing. As previously reported [64], changes in haplogroup assignment based on addition of polymorphisms located in coding regions, which are essential for haplogroup designation, have been observed. Such changes included not only reassignment to more refined, low-clade haplogroups, but also revisions of clade denomination based on the absence or presence of typical, diagnostic mutations for expected haplogroups.

Regarding sequencing consistency between SS and MPS results, we have identified five inconsistencies between two employed techniques for this sample set, just as presented in Supplementary Table S5. Three of the observed differences were shown to consist of heteroplasmic variations, unidentified by SS and evidenced with MPS techniques. This kind of observation is actually expected as MPS methods have a markedly superior ability to detect heteroplasmies versus traditional SS [43]. The remaining discrepancies refer to two mutations (belonging to a single sample, DF45, at positions 297 and 416) identified by Sanger sequencing and undetected in full mtDNA sequencing. The reasons for such discrepancies are unclear: they were at first thought as possibly related to algorithms used in base-calling as both mutations are located in primer-binding regions of the evaluated panel; thus, results would have been affected during data-processing steps, like primer trimming or other automated procedures. Manual review of data with IGV visualization software, however, did not reveal the presence of both polymorphisms in the DF45 sample as the overlapping strategy used with this panel design allowed for sequence evaluation in these regions by amplification with another primer pair. Therefore, reasons for the two mentioned inconsistencies remain uncertain, since raw data for SS was not available for additional review. However, the careful performed analysis in MPS data for this particular sample suggests that discrepancies are likely a result of misinterpreted results in SS evaluation. As a consequence, total rCRS variations detected inside the CR is equivalent to 1024, and the total disagreement ratio (including unidentified heteroplasmies) was less than 0.5% between both methods. In addition, no inconsistencies were detected for samples evaluated by

MPS only and sequenced in duplicate across two distinct runs. This result suggests that MPS-obtained sequences might be as reliable as currently employed techniques in forensic laboratories, providing a highly informative and trustworthy source of genetic data with possible forensic applications, as stated by other research groups [26].

It has been proposed that the actual frequency of heteroplasmic polymorphisms in mtDNA samples is higher than previously estimated as MPS techniques have an increased ability to recognize and identify such occurrences besides Sanger sequencing, especially at lower levels (between 10–20% heteroplasmy thresholds) [8,65]. In this study, all three Sanger-identified point heteroplasmies were also reported by MPS processing. In addition to that, two heteroplasmic polymorphisms were identified with MPS inside CR and not previously reported as such by Sanger analysis. Considering the adopted threshold of 15% and minimal coverage of 40x for HP determination (as proposed by [44]), our analyses identified a total number of 20 HP observations with 15 of them located inside mtDNA CR. It is important to note that certain low-level heteroplasmies might have been undetected in our analysis, since they present low-quality scores in our analysis (as a result of marked strand imbalance, for instance) or were identified as noise (caused by a low number of reads within these particular nucleotides, insufficient for variant identification). Supplementary Table S6 details all identified PHP and relative contributions for each variant.

3.4. Inter-population analysis based on mtDNA haplotype frequencies

A series of analyses were conducted to evaluate the differences and genetic structure among Brazilian mtDNA haplotype frequencies and 26 worldwide populations included in 1kGP. Pairwise F_{ST} values were calculated based on haplotype frequency results that are displayed as a heat map in Fig. 2. For the Brazilian population as a whole, the results ranged from 0.06518 (for ASW group, comprised of Americans with AFR ancestry) to 0.1693 (EUR from Finland). Brazilians showed higher similarity levels with admixed populations where the AFR and/or NA contribution to the pool of maternal lineages is higher (in addition to ASW, Puerto Ricans (PUR), Mexicans (MXL) and African Caribbean (ACB) populations), followed by samples from African countries and remaining admixed Latin Americans. Further, EUR samples were found to have a higher pairwise distance from the BRA population. Both Asian groups presented intermediate genetic resemblance with the present sample set, with South Asians (SAS) displaying an overall genetic similarity with investigated regions slightly higher than East Asians (EAS) samples (average pairwise F_{ST} values to Brazilian population equivalent to 0.0926 and 0.1018, respectively). Complete pairwise F_{ST} values are presented in Supplementary Table S7 (with values obtained based on populations) and S8 (displaying differences based on biogeographic continental origin groups).

In order to further examine the aforementioned results and clarify the eventual genetic relationship between 26 worldwide populations and Brazilian regions, a MDS plot based on pairwise F_{ST} values was drawn for the obtained data and is presented in Fig. 3A. A clear pattern can be observed with negative values for Dimension 1 as a characteristic feature for non-AFR populations. Both Asians groups (EAS and SAS) are grouped approximately together around the axis center with an apparently distinct cluster separation between both geographical origins. As expected, AMR and Asians groups have a close distribution occupying the first quadrant. However, while EUR populations seem to be grouped relatively close to each other, AMR and Asians have a broader dispersion. The BRA population seem to be located in a very particular location, with a clear directional tendency towards positive values in Dimension 1 (trending to AFR cluster). PUR seems to be an outlier in the AMR group, exhibiting a similar tendency to the BRA population and close to the AFR cluster. A wider dispersion displayed by AMR populations seems to reflect their admixed nature, and how distinct relative contributions of parental populations affect modern population frequency distribution. A cladogram was also drawn
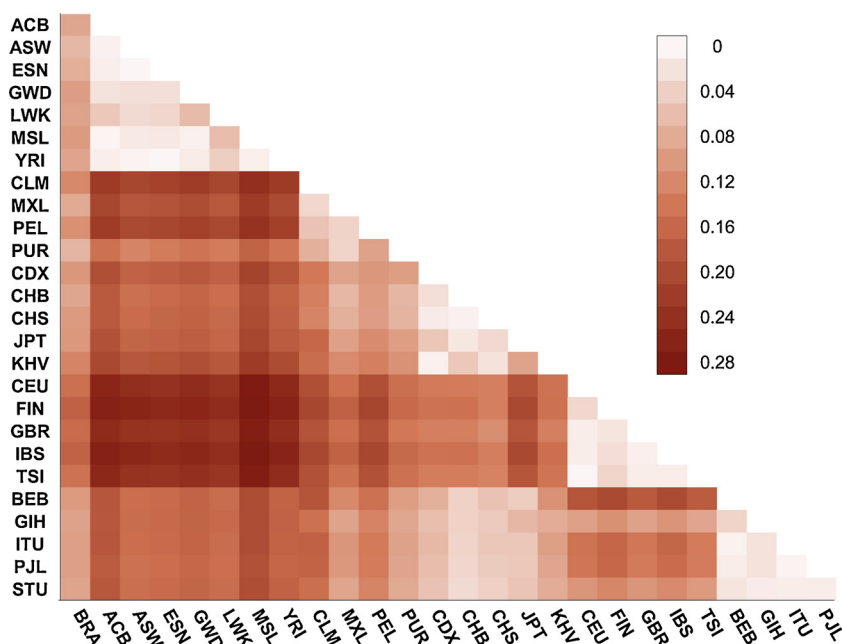
Fig. 2. Heat map of pairwise F_{ST} values based on mtDNA haplotype frequencies for 26 worldwide populations and Brazilian data set. Population full names described in Supplementary Table S1.

applying N-J methods, and is presented in Fig. 3B. The resulting phylogenetic tree shows the BRA population positioned between the AFR and AMR branches, reflecting higher parental contributions of individuals with such biogeographical origins for maternal lines. EUR populations are located at the treetop, featuring the largest distance to the Brazilian sample set. It is important to highlight that these findings are supported by previously reported genetic, historical or ethnographic information, especially for the Brazilian case [18–25,48–55].

Finally, as the occurrence of genetic stratification regarding mtDNA haplotype frequencies seems to be clearly defined for Brazilian and all other worldwide populations, statistical tests were performed to evaluate the amount of genetic variance associated exclusively with the verified genetic structure. Global AMOVA tests were executed with BRA and 26 worldwide populations, arranged according to geographical origin (BRA samples were categorized in a separate group from remaining AMR populations), and the results are detailed in Table 2. Among-group variation accounts for an estimate of roughly 85% of total genetic variability with less than 1% of total genetic differences explained by divergence found among populations within same groups. These results indicate that population genetic structure associated with mtDNA haplotypes are highly dependent on sample biogeographical origin, and that populations within the same continent present very similar haplotype distributions. Therefore, maternal lineage ancestral origins can be traced back to their estimated parental population with a certain degree of confidence. In this regard, even though high admixture levels are evidenced in BRA populations, uneven contribution levels of distinct parental populations in maternal and paternal lineages (assessed by mtDNA and Y-chromosome haplotypes) limit global diversity presented by mtDNA alone. Owing to historical interbreeding dynamics in BRA population formation, AFR and AMR possess relatively higher contributions to BRA mtDNA genetic pool, which does not necessarily reflect the overall contribution of parental populations from distinct biogeographical origins to the modern Brazilian population [18,19,21,54]. This fact introduces a certain degree of genetic structure and bias that should be considered when forensic applications or database constructions are designed. Till the present moment, genetic structure in Brazilian populations regarding mtDNA haplotype distribution is admitted, although very limited in magnitude, considering both spatial distribution in regional populations (estimated on a state or

geopolitical region basis) or in distinct ethnic groups in the Brazilian population (based on self-declared assessment) [28,66,67]. As a final note, it is important to account for the fact that small, isolated groups within the national territory were observed to display higher levels of mtDNA genetic structure when compared to the national population [55,68–70]. This fact must be taken into consideration during forensic analysis, especially when criminal evidence is believed to originate from individuals belonging to these local communities.

4. Conclusion

The present study sought to evaluate the utility of the Precision ID mtDNA Whole Genome Panel in Brazilian populations in order to start the development of a frequency database that allows for use of this commercial solution in real forensic casework. A total of 96 samples collected from three Brazilian cities located in distinct geopolitical regions were evaluated. This is one of the first studies reporting complete mitochondrial genome sequences for Brazilian samples, and additional genetic information provided by coding region inclusion during analysis can be a determinant in correct haplotype classification and identification of new or undescribed subclades. Diversity indices and other relevant forensic statistical parameters obtained for this sample set exhibit high polymorphism levels that are suitable for human identification purposes. Full mtDNA haplotypes were shown to add valuable amounts of genetic information available to forensic investigators, and in a limited sample size (n = 65) were observed to properly discriminate two distinct pairs of samples, previously undistinguishable via traditional Sanger sequencing of CR techniques. Therefore, from a forensic perspective, inclusion of MPS analytical tools for mtDNA analysis can be demonstrated to be a very useful resource for law enforcement agencies not only because of full mtDNA sequencing capability, but also to high-throughput performance. Workflow automation should also be considered when casework volumes or existent case backlog is significant. However, some issues still need to be addressed, including technical optimizations in panel design and improvements in raw data processing and results presentation. Design and construction of a representative forensic national database is also essential, with a significant sample number addition in order to appraise evidence of possible genetic regional or ethnical stratification in the

A

Stress = .11659   RSQ = .94901

B
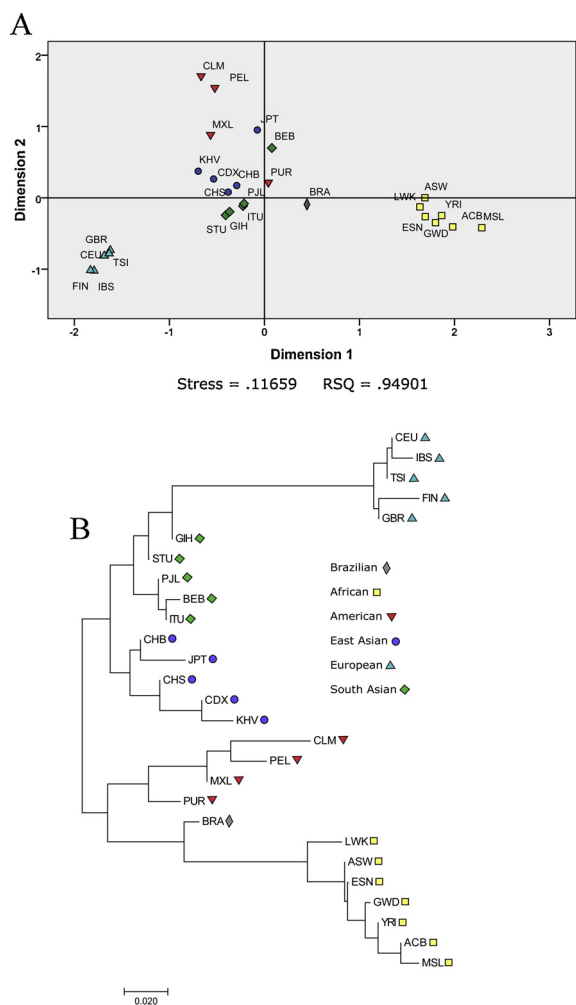
Fig. 3. Genetic distance evaluation for inter population analysis of 26 world-wide populations and Brazilian data ser. (A) MDS plot constructed based on pairwise F_{ST} values for full mtdna haplotype frequencies. (B) Phylogenetic tree build with Neighbor-joining methods, based on pairwise F_{ST} values obtained. Scale in F_{ST} units.

Table 2
Fixation indexes and global AMOVA results for Brazilian and 26 worldwide populations, based on full mtDNA haplotype frequencies.

Source of Variation	Relative Variation (%)	Fixation Indexes
Among groups	84.65	F_{CT}: 0.84645
Among populations within groups	0.32	F_{SC}: 0.02111
Within populations	15.03	F_{ST}: 0.84969

Brazilian population [63].

Proposed MPS metrics evaluated in this work can be used in addition to other computational resources to assess the overall quality of produced mtDNA results [71]. For this particular product version and equipment used, a great deal of effort is still necessary in data review, manual data quality verification and results evaluation as automatically generated reports are not in compliance with suggested forensic standards. In particular, handling and manual revision of primer-derived bias (caused by use of degenerated primers in reagent design) at the time of results interpretation was shown to be a meticulous task where subjective analyst intervention was crucial [72]. As such, a considerable degree of expertise in mtDNA data analysis is still necessary, and such proficiency levels for technical personnel might not be available in every national forensic laboratory. For instance, the recently proposed

possibility of paternal heritage for mtDNA owing to defective elimination of spermatozoid carried organelles [73] in zygotes might be indistinguishable from a mtDNA mixture sample. Although there is insufficient scientific evidence to fully support this hypothesis [74], such a possibility might present a challenge to an inexperienced or inadequately trained lab workforce. Techniques for mixture deconvolution based on mtDNA data have been proposed [17,26,44], but these methods involve manual or visual data inspection and critical results evaluation, where analyst knowledge may be critical in phasing contributing samples and mixture parsing [17,26]. Even though software packages for automatic mtDNA mixture analysis have been released [75], such tools still require some level of operator judgment in data evaluation. Therefore, inclusion of MPS tools in Brazilian forensic routine casework processing would require intensive training and official governmental investment in order to assure uniform, consistent results production across different national forensic laboratories throughout the country. As an alternative, the constitution of a dedicated, specialized central laboratory (or even a small number of regional facilities, aiming to minimize logistical restraints resulting from Brazilian large territorial distribution) responsible for processing mtDNA-related biological evidence could be considered, where sufficient personnel training and expertise would be ensured and financial or human resource allocation could be optimized.

Declaration of Competing Interest

Authors declare they have no conflict of interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.fsigen.2019.07.004.

References

[1] M. Kaur, Mitochondrial DNA (mtDNA) sequence analysis as an additional tool for forensic identification purposes, J. Indian Acad. Forensic Med. (38) (2016).
[2] B. Budowle, M.W. Allard, M.R. Wilson, R. Chakraborty, Forensics and mitochondrial DNA: applications, debates, and foundations, Annu. Rev. Genomics Hum. Genet. (4) (2003) 119–141, https://doi.org/10.1146/annurev.genom.4.070802.110352.
[3] G.N.N. Sultana, M.Z. Sultan, Mitochondrial DNA and methods for forensic identification, J. Forensic Sci. Criminal Investig. 1 (9) (2018) 555755, , https://doi.org/10.19080/JFSCI.2018.09.555755.
[4] T. Melton, C. Holland, M. Holland, Forensic mitochondrial DNA: current practice and future potential, Forensic Sci. Rev. 2 (24) (2012) 110.
[5] K.T. Wai, M. Barash, P. Gunn, Performance of the early access AmpliSeq™ Mitochondrial Panel with degraded DNA samples using the Ion Torrent™ platform, Electrophoresis. 21 (39) (2018) 2776–2784, https://doi.org/10.1002/elps.201700371.
[6] A.E. Woerner, A. Ambers, F.R. Wendt, J.L. King, R.S. Moura-Neto, R. Silva, B. Budowle, Evaluation of the precision ID mtDNA whole genome panel on two massively parallel sequencing systems, Forensic Sci. Int. Genet. (36) (2018) 213–224, https://doi.org/10.1016/j.fsigen.2018.07.015.
[7] J. Butler, The future of forensic DNA analysis, Philos. Trans. B (370) (2015)

20140252, https://doi.org/10.1098/rstb.2014.0252.

[8] M.X. Sosa, I.K. Sivakumar, S. Maragh, V. Veeramachaneni, R. Hariharan, M. Parulekar, K.M. Fredrikson, T.T. Harkins, J. Lin, A.B. Feldman, P. Tata, G.B. Ehret, A. Chakravarti, Next-Generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency, PLoS Comput. Biol. 10 (8) (2012) e1002737, , https://doi.org/10.1371/journal.pcbi.1002737.

[9] W. Parson, C. Strobl, G. Huber, B. Zimmermann, S.M. Gomes, L. Souto, L. Fendt, R. Delport, R. Langit, S. Wootton, R. Lagacé, J. Irwin, Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM), Forensic Sci. Int. Genet. (6) (2013) 632–639, https://doi.org/10.1016/j.fsigen.2013.09.007.

[10] W. Parson, G. Huber, L. Moreno, M.B. Madel, M.D. Brandhagen, S. Nagl, C. Xavier, M. Eduardoff, T.C. Callaghan, J.A. Irwin, Massively parallel sequencing of complete mitochondrial genomes from hair shaft samples, Forensic Sci. Int. Genet. (15) (2015) 8–15, https://doi.org/10.1016/j.fsigen.2014.11.009.

[11] N.E.1 Weiler, G.1 de Vries, T.2 Sijen, Development of a control region-based mtDNA SNaPshot™ selection tool, integrated into a mini amplicon sequencing method, Sci. Justice (56) (2016) 2, https://doi.org/10.1016/j.scijus.2015.11.003.

[12] G.1 Chemale, G.G. Paneto, M.A. Menezes, J.M. de Freitas, G.S. Jacques, R.M. Cicarelli, P.R. Fagundes, Development and validation of a D-loop mtDNA SNP assay for the screening of specimens in forensic casework, Forensic Sci. Int. Genet. 3 (7) (2013) 353–358, https://doi.org/10.1016/j.fsigen.2018.02.005.

[13] S.Y. Shih, N. Bose, A.B.R. Gonçalves, H.A. Erlich, C.D. Calloway, Applications of probe capture enrichment next generation sequencing for whole mitochondrial genome and 426 nuclear SNPs for forensically challenging samples, Genes (Basel) 1 (9) (2018) 49, https://doi.org/10.3390/genes9010049.

[14] Y. Zhou, F. Guo, J. Yu, F. Liu, J. Zhao, H. Shen, B. Zhao, F. Jia, Z. Sun, H. Song, X. Jiang, Strategies for complete mitochondrial genome sequencing on Ion Torrent PGM™ platform in forensic sciences, Forensic Sci. Int. Genet. (22) (2016) 11–21, https://doi.org/10.1016/j.fsigen.2016.01.004.

[15] L. Chaitanya, A. Ralf, M. van Oven, T. Kupiec, J. Chang, R. Lagacé, M. Kayser, Simultaneous whole mitochondrial genome sequencing with short overlapping amplicons suitable for degraded DNA using the Ion Torrent Personal Genome Machine, Hum. Mutat. 12 (36) (2015) 1236–1247, https://doi.org/10.1002/humu.22905.

[16] C. Strobl, M. Eduardoff, M.M. Bus, M. Allen, W. Parson, Evaluation of the precision ID whole MtDNA genome panel for forensic analyses, Forensic Sci. Int. Genet. (35) (2018) 21–25, https://doi.org/10.1016/j.fsigen.2018.03.013.

[17] J.D. Churchill, M. Stoljarova, J.L. King, B. Budowle, Massively parallel sequencing-enabled mixture analysis of mitochondrial DNA samples, Int. J. Legal Med. 5 (132) (2018) 1263–1272, https://doi.org/10.1007/s00414-018-1799-3.

[18] S.M. Callegari-Jacques, D. Grattapaglia, F.M. Salzano, S.P. Salamoni, S.G. Crossetti, M.E. Ferreira, M.H. Hutz, Historical genetics: spatiotemporal analysis of the formation of the Brazilian population, Am. J. Hum. Biol. 6 (15) (2003) 824–834, https://doi.org/10.1002/ajhb.10217.

[19] T.C. Lins, R.G. Vieira, B.S. Abreu, D. Grattapaglia, R.W. Pereira, Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs, Am. J. Hum. Biol. 2 (22) (2010) 187–192, https://doi.org/10.1002/ajhb.20976.

[20] IBGE, Brasil: 500 anos de povoamento, Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, 2007 231 p.

[21] S.D.J. Pena, G. Di Pietro, M. Fuchshuber-Moraes, J.P. Genro, M.H. Hutz, F.S. Kehdy, F. Kohlrausch, L.A. Magno, R.C. Montenegro, M.O. Moraes, M.E. de Moraes, M.R. de Moraes, E.B. Ojopi, J.A. Perini, C. Racciopi, A.K. Ribeiro-Dos-Santos, F. Rios-Santos, M.A. Romano-Silva, V.A. Sortica, G. Suarez-Kurtz, The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected, PLoS One 2 (6) (2011) e17063, https://doi.org/10.1371/journal.pone.0017063.

[22] IBGE. Atlas do Censo Demográico, Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, . (2010) 2013 216.

[23] E. Avila, A.B. Felkl, P. Graebin, C.P. Nunes, C.S. Alho, Forensic characterization of Brazilian regional populations through massive parallel sequencing of 124 SNPs included in HID ion Ampliseq Identity Panel, Forensic Sci. Int. Genet. (40) (2019) 74–84, https://doi.org/10.1016/j.fsigen.2019.02.12.

[24] R. Resque, L. Gusmão, M. Geppert, L. Roewer, T. Palha, L. Alvarez, A. Ribeiro-dos-Santos, S. Santos, Male lineages in Brazil: intercontinental admixture and stratification of the European background, PLoS One 4 (11) (2016) e0152573, , https://doi.org/10.1371/journal.pone.0152573.

[25] J. Alves-Silva, M.S. Santos, P.E. Guimarães, A.C. Ferreira, H.J. Bandelt, S.D. Pena, V.F. Prado, The ancestry of Brazilian mtDNA lineages, Am. J. Hum. Genet. 2 (67) (2000) 444–461, https://doi.org/10.1086/303004.

[26] J.D. Churchill, M. Stoljarova, J.L. King, B. Budowle, Parsing apart the contributors of mitochondrial DNA mixtures with massively parallel sequencing data, Forensic Sci. Int. Genet. Suppl. Ser. (6) (2017) e439–e441, https://doi.org/10.1016/j.fsigss.2017.09.145.

[27] WMA, World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects, J. Am. Med. Assoc. 20 (310) (2013) 2191–2194, https://doi.org/10.1001/jama.2013.281053.

[28] J.M. Freitas, L.H. Fassio, D.F. Braganholi, G. Chemale, Mitochondrial DNA control region haplotypes and haplogroup diversity in a sample from Brasília, Federal District, Brazil, Forensic Sci. Int. Genet. (18) (2019) S1872–4973, https://doi.org/10.1016/j.fsigen.2019.02.006 30514-3.

[29] T.F.P. Palha, E.M.R. Ribeiro-Rodrigues, G.C. Cavalcante, A. Marrero, I.R. Souza, C.J.S. Uehara, C.H.S. Motta, D. Koshikene, D.A. Silva, E.F. Carvalho, G. Chemale, J.M. Freitas, L. Alexandre, R.T.F. Paranaiba, M.P. Soler, S. Santos, Population genetic analysis of insertion–deletion polymorphisms in a Brazilian population using

[30] G. Chemale, J.M. Freitas, J.L. Badaraco, L.S. Rosa, J.A. Martins, A.D. Martins, G.S. Jacques, R.T. Paranaíba, Y-chromosomal STR haplotypes in a sample from Brasília, Federal District, Brazil, Forensic Sci. Int. Genet. (9) (2014) e3–4, https://doi.org/10.1016/j.fsigen.2013.06.005.

the Investigator DIPplex kit, Forensic Sci. Int. Genet. (19) (2015) 10–14, https://doi.org/10.1016/j.fsigen.2015.03.015.

[31] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 2 (23) (1999) 147, https://doi.org/10.1038/13779.

[32] A. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tully, M. Wilson, DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. (2) (2000) 79–85, https://doi.org/10.1016/S0379-0738(00)00161-4.

[33] W. Parson, H.-J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, Forensic Sci. Int. Genet. 1 (1) (2007) 13–19, https://doi.org/10.1016/j.fsigen.2006.11.003.

[34] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, A.R. Mayr, N. Morling, E. Pokorak, M. Prinz, A. Salas, P.M. Schneider, T.J. Parsons, DNA commission of the international society for forensic genetics: revised and extended guidelines for mitochondrial DNA typing, Forensic Sci. Int. Genet. (13) (2014) 134–142, https://doi.org/10.1016/j.fsigen.2014.07.010.

[35] H. Thorvaldsdóttir, J.T. Robinson, J.P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, Brief. Bioinformatics 2 (14) (2013) 178–192, https://doi.org/10.1093/bib/bbs017.

[36] H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt, F. Kronenberg, A. Salas, S. Schönherr, HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing, Nucleic Acids Res. 2016 (15) (2016) w58–w63, https://doi.org/10.1093/nar/gkw233.

[37] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 2 (30) (2009) e386–e394, https://doi.org/10.1002/humu.20921.

[38] N. Huber, W. Parson, A. Dürc, Next generation database search algorithm for forensic mitogenome analyses, Forensic Sci. Int. Genet. (37) (2018) 204–214, https://doi.org/10.1016/j.fsigen.2010.02.008.

[39] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 3 (10) (2010) 564–567, https://doi.org/10.1111/j.1755-0998.2010.02847.x.

[40] 1000 Genomes Project Consortium, et al., An integrated map of genetic variation from 1092 human genomes, Nature (491) (2012) 56–65, https://doi.org/10.1038/nature11632.

[41] K. Katoh, H. Toh, Recent developments in the MAFFT multiple sequence alignment program, Brief. Bioinformatics 4 (9) (2008) 286–298, https://doi.org/10.1093/bib/bbn013.

[42] S. Kumar, G. Stecher, K. Tamura, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 7 (33) (2016) 1870–1874, https://doi.org/10.1093/molbev/msw054.

[43] R.S. Just, A.C. Irwin, W. Parson, Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing, Forensic Sci. Int. Genet. (18) (2015) 131–139, https://doi.org/10.1016/j.fsigen.2015.05.003.

[44] S. Cho, M.Y. Kim, J.H. Lee, S.D. Lee, Assessment of mitochondrial DNA heteroplasmy detected on commercial panel using MPS system with artificial mixture samples, Int. J. Legal Med. 4 (132) (2018) 1049–1056, https://doi.org/10.1007/s00414-017-1755-7.

[45] S. Park, S. Cho, H.J. Seo, J.H. Lee, M.-Y. Kim, S.D. Lee, Entire mitochondrial DNA sequencing on massively parallel sequencing for the Korean population, J. Korean Med. Sci. (32) (2017) 587–592, https://doi.org/10.3346/jkms.2017.32.4.587.

[46] M.G. Ross, C. Russ, M. Costello, A. Hollinger, N.J. Lennon, R. Hegarty, C. Nusbaum, D.B. Jaffe, Characterizing and measuring bias in sequence data, Genome Biol. 5 (14) (2013) r51, https://doi.org/10.1186/gb-2013-14-5-r51.

[47] M. Nelson, R. Chase, L. DePalma, Making Sense of DNA Backlogs, 2012 — Myths vs. Reality, National Institute of Justice, Washington, 2012, p. 2013 20.

[48] C. Fridman, R.S. Gonzalez, A.C. Pereira, M.M. Cardena, Haplotype diversity in mitochondrial DNA hypervariable region in a population of southeastern Brazil, Int. J. Legal Med. (128) (2014) 589–593, https://doi.org/10.1007/s00414-014-1023-z.

[49] S. Bernardo, R. Hermida, M. Desidério, D.A. Silva, E.F. de Carvalho, MtDNA ancestry of Rio de Janeiro population, Brazil, Mol. Biol. Rep. 4 (41) (2014) 1945–1950, https://doi.org/10.1007/s11033-014-3041-9.

[50] A.B. Barbosa, L.A. da Silva, D.A. Azevedo, V.Q. Balbino, L.M. Silva, Mitochondrial DNA control region polymorphism in the population of Alagoas state, north-eastern Brazil, J. Forensic Sci. (53) (2008) 142–146, https://doi.org/10.1111/j.1556-4029.2007.00619.x.

[51] L. Palencia, L. Valverde, A. Alvarez, L.M. Cainé, S. Cardoso, M.A. Alfonso-Sánchez, M.F. Pinheiro, Mitochondrial DNA diversity in a population from Santa Catarina (Brazil): predominance of the European input, Int. J. Legal Med. 4 (124) (2010) 331–336, https://doi.org/10.1007/s00414-010-0464-2.

[52] N.M. Sanches, G.G. Paneto, R.F. Figueiredo, A.O. Mello, R.M. Cicarelli, Mitochondrial DNA control region diversity in a population from Espírito Santo state, Brazil, Mol. Biol. Rep. 10 (41) (2014) 6645–6648, https://doi.org/10.1007/s11033-014-3547-1.

[53] A.P. Schaan, L. Costa, D. Santos, A. Modesto, M. Amador, C. Lopes, S.H. Rabenhorst, R. Montenegro, B.D.A. Souza, T. Lopes, F.K. Yoshioka, G. Pinto, V. Silbiger, Ribeiro-Dos-Santos Â. mtDNA structure: the women who formed the Brazilian Northeast, BMC Evol. Biol. 1 (17) (2017) 185, https://doi.org/10.1186/s12862-017-1027-7.

[54] T. Hünemeier, C. Carvalho, A.R. Marrero, F.M. Salzano, S.D. Pena, M.C. Bortolini,

Niger-Congo speaking populations and the formation of the Brazilian gene pool: mtDNA and Y-chromosome data, Am. J. Phys. Antropol. 2 (133) (2007) 854–867, https://doi.org/10.1002/ajpa.20604.

[55] W.A. Silva, M.C. Bortolini, M.P. Schneider, A. Marrero, J. Elion, R. Krishnamoorthy, M.A. Zago, MtDNA haplogroup analysis of black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade, Hum. Biol. 1 (78) (2006) 29–41, https://doi.org/10.1353/hub.2006.0028.

[56] R. Barral-Arca, S. Pischedda, A. Gómez-Carballa, A. Pastoriza, A. Mosquera-Miguel, M. López-Soto, F. Martinón-Torres, V. Álvarez-Iglesias, A. Salas, Meta-analysis of mitochondrial DNA variation in the Iberian Peninsula, PLoS One 7 (11) (2016) e0159735, , https://doi.org/10.1371/journal.pone.0159735.

[57] M. Pala, A. Olivieri, A. Achilli, M. Accetturo, E. Metspalu, M. Reidla, E. Tamm, M. Karmin, T. Reisberg, B.H. Kashani, U.A. Perego, V. Carossa, F. Gandini, J.B. Pereira, P.A. Soares, Mitochondrial DNA signals of late glacial recolonization of Europe from Near Eastern refugia, Am. J. Hum. Genet. 5 (90) (2019) 915–924, https://doi.org/10.1016/j.ajhg.2012.04.003.

[58] K.K. Abu-Amero, J.M. Larruga, V.M. Cabrera, A.M. González, Mitochondrial DNA structure in the Arabian Peninsula, BMC Evol. Biol. 45 (8) (2008), https://doi.org/10.1186/1471-2148-8-45.

[59] R.S. Just, T.M. Diegoli, J.L. Saunier, J.A. Irwin, T.J. Parsons, Complete mitochondrial genome sequences for 265 African American and U.S "Hispanic" individuals, Forensic Sci. Int. Genet. 3 (2) (2008) e45–e48, https://doi.org/10.1016/j.fsigen.2007.12.001.

[60] V. Fernandes, P. Triska, J.B. Pereira, F. Alshamali, T. Rito, A. Machado, Z. Fajkošová, B. Cavadas, V. Černý, P. Soares, M.B. Richards, L. Pereira, Genetic stratigraphy of key demographic events in Arabia, PLoS One 3 (10) (2015) e0118625, , https://doi.org/10.1371/journal.pone.0118625.

[61] T. Heinz, M. Pala, A. Gómez-Carballa, M.B. Richards, A. Salas, Updating the African human mitochondrial DNA tree: relevance to forensic and population genetics, Forensic Sci. Int. Genet. (39) (2017) 156–159, https://doi.org/10.1016/j.fsigen.2016.12.016.

[62] F.A. Kaestle, R.A. Kittles, A.L. Roth, E.J. Ungvarsky, Database limitations of the evidentiary value of forensic mitochondrial DNA evidence, Am. Crim. Law Rev. 1 (43) (2006) 53–88.

[63] J.A. Irwin, W. Parson, M.D. Coble, Just RS. mtGenome reference population databases and the future of forensic mtDNA analysis, Forensic Sci. Int. Genet. (5) (2011) 222–225, https://doi.org/10.1016/j.fsigen.2010.02.008.

[64] F. Simão, C. Strobl, C. Vullo, L. Catelli, P. Machado, N. Huber, L. Schnaller, G. Huber, C. Xavier, E.F. Carvalho, L. Gusmão, W. Parson, The maternal inheritance of Alto Paraná revealed by full mitogenome sequences, Forensic Sci. Int. Genet. (39) (2019) 66–72, https://doi.org/10.1016/j.fsigen.2018.12.007.

[65] A. Ramos, C. Santos, L. Mateiu, M.M. Gonzalez, L. Alvarez, L. Azevedo, A. Amorim, M.P. Aluja, Frequency and pattern of Heteroplasmy in the complete human mitochondrial genome, PLoS One 10 (8) (2013) e74636, , https://doi.org/10.1371/journal.pone.0074636.

[66] B.R. Ramos, M.P. D'Elia, M.A. Amador, N.P. Santos, S.E. Santos, E.C. Castelli, S.S. Witkin, H.A. Miot, L.D. Miot, M.G. da Silva, Neither self-reported ethnicity nor declared family origin are reliable indicators of genomic ancestry, Genetica 3 (144) (2016) 259–265, https://doi.org/10.1007/s10709-016-9894-1.

[67] M.M. Cardena, A. Ribeiro-Dos-Santos, S. Santos, A.J. Mansur, A.C. Pereira, C. Fridman, Assessment of the relationship between self-declared ethnicity, mitochondrial haplogroups and genomic ancestry in Brazilian individuals, PLoS One 4 (8) (2013) e62005, , https://doi.org/10.1371/journal.pone.0062005.

[68] V.F. Gonçalves, F.C. Parra, H. Gonçalves-Dornelas, C. Rodrigues-Carvalho, H.P. Silva, S.D. Pena, Recovering mitochondrial DNA lineages of extinct Amerindian nations in extant homopatric Brazilian populations, Investig. Genet. 1 (1) (2010) 13, https://doi.org/10.1186/2041-2223-1-13.

[69] B.M. Carvalho, M.C. Bortolini, S.E.B. Sidney Santos, A.K.C. Ribeiro-dos-Santos, Mitochondrial DNA mapping of social-biological interactions in brazilian amazonian african-descendant populations, Genet. Mol. Biol. 1 (31) (2008) 12–22, https://doi.org/10.1590/S1415-47572008000100002.

[70] R.H. Ward, F.M. Salzano, S.L. Bonatto, M.H. Hutz, C.E.A. Coimbra Jr., R.V. Santos, Mitochondrial DNA polymorphism in three Brazilian Indian tribes, Am. J. Hum. Biol. 3 (8) (1996) 317–323, https://doi.org/10.1002/(SICI)1520-6300(1996)8:3<317::AID-AJHB2>3.0.CO;2-X.

[71] J.D. Churchill, J.L. King, R. Chakraborty, B. Budowle, Effects of the Ion PGM™ Hi-Q™ sequencing chemistry on sequence data quality, Int. J. Legal Med. 5 (130) (2016) 1169–1180, https://doi.org/10.1007/s00414-016-1355-y.

[72] T.I. Huszar, J.H. Wetton, M.A. Jobling, Mitigating the effects of reference sequence bias in single-multiplex massively parallel sequencing of the mitochondrial DNA control region, Forensic Sci. Int. Genet. (40) (2019) 9–17, https://doi.org/10.1016/j.fsigen.2019.01.008.

[73] S. Luo, C.A. Valencia, J. Zhang, N.-C. Lee, J. Slone, B. Gui, X. Wang, Z. Li, S. Dell, J. Brown, S.M. Chen, Y.-H. Chien, W.-L. Hwu, P.-C. Fan, L.-J. Wong, P.S. Atwal, T. Huang, Biparental inheritance of mitochondrial DNA in humans, Proc. Natl. Acad. Sci. U. S. A. 51 (115) (2018) 13039–13044, https://doi.org/10.1073/pnas.1810946115.

[74] S. Lutz-Bonengel, W. Parson, No further evidence for paternal leakage of mitochondrial DNA in humans yet, Proc. Natl. Acad. Sci. U. S. A. 6 (116) (2019) 1821–1822, https://doi.org/10.1073/pnas.1820533116.

[75] S.H. Vohr, R. Gordon, J.M. Eizengaa, H.A. Erlich, C.D. Calloway, R.A. Green, A phylogenetic approach for haplotype analysis of sequence data from complex mitochondrial mixtures, Forensic Sci. Int. Genet. (30) (2017) 93–105, https://doi.org/10.1016/j.fsigen.2017.05.007.