Forensic Science International: Genetics 40 (2019) 74–84

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Forensic characterization of Brazilian regional populations through massive parallel sequencing of 124 SNPs included in HID ion Ampliseq Identity Panel

Eduardo Avila[a,b,c,*], Aline Brugnera Felkl[b], Pietra Graebin[b], Cláudia Paiva Nunes[b], Clarice Sampaio Alho[b,c]

[a] Setor Técnico-Científico, Superintendência Regional de Polícia Federal do RS, 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

ARTICLE INFO

Keywords:
Massive Parallel Sequencing
SNP
Brazilian population
Y-haplotypes
HID Ion Ampliseq Identity Panel
Forensics

ABSTRACT

Use of Massive Parallel Sequencing (MPS) techniques has been investigated by forensic community aiming introduction of such methods in routine forensic casework analyses. Interesting features presented by MPS include high-throughput, ability to simultaneous genotyping of significant number of samples and forensic markers, workflow automation, among others. Emergence of single nucleotide polymorphism (SNP) as forensic relevant markers was facilitated in this process, since concurrent typing of larger marker sets is necessary for obtaining same levels of individual discrimination provided by other marker categories. In this context, HID Ion Ampliseq Identity Panel is a commercial solution with forensic purposes comprising simultaneous analysis of 90 highly informative autosomal SNPs and 34 Y –chromosome superior clade SNPs for male lineage haplotyping. SNP typing can be obtained with smaller amplicons, and this panel was designed for efficient processing of critical or challenging forensic samples. In this work, a sample of 432 individuals from all five Brazilian geopolitical regions was evaluated with this panel, in order to access feasibility of this panel use in a national basis. Results obtained for all five regions, including forensic parameters, show that this marker set can be efficiently employed for Brazilian nationals in human identification or kinship determination applications, due to high levels of genetic discriminative information content displayed by Brazilians. Interpopulation comparison studies were executed among Brazilian regional populations and 26 worldwide populations, in order to access genetic stratification occurrence. Some levels of population structure were identified, and impact on database design was discussed. Y-chromosome haplotyping of Brazilian samples revealed high levels of European ancestry in Brazilian male lineages, and utility of haplotyping in real forensic casework is addressed. Finally, genotyping and sequencing efficiency with this panel were addressed, as an effort to appraise the adequacy of this panel use in Brazilian national forensic demands.

1. Introduction

Short Tandem Repeats (STR)-based length polymorphism analysis, associated with Capillary Electrophoresis (CE) techniques, has been the gold standard in forensic genetics for the past decades [1]. STR markers are highly polymorphic and informative, and as such were chosen as the core of most criminal genetic databases established worldwide [2]. Despite that, some limitations are still associated with these methods [3]. Relatively large amplicons are generated, posing a range of problems in successfully obtaining genetic profiles for a variety of biological and criminal samples. Low template DNA copy numbers, highly degraded DNA or environmental exposure can considerably impact genetic profile quality, with prominent effects on larger amplicons [4].

In order to circumvent such technical issues, the use of different genetic markers has been proposed [5]. Single Nucleotide Polymorphisms (SNPs) constitute genetic variations where a unique, specific nucleotide position in the genome can be found in multiple forms within a population. SNPs are widespread in human genome, and its use in forensic applications present some advantages over commonly used STRs, since SNPs have a lower mutation rate, can be typed with smaller amplicons and can also provide additional genetic information, as parental lineage determination, biogeographical ancestry or phenotypical traits assessment [6]. Since SNPs are usually found in biallelic forms in different populations, a larger number of different markers

* Corresponding author at: Laboratório de Genética Humana e Molecular, Escola de Ciências, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga, 6681 – Prédio 12C, Sala 231, 90619-900 Porto Alegre, Brazil.
E-mail address: e.avila@edu.pucrs.br (E. Avila).

https://doi.org/10.1016/j.fsigen.2019.02.012
Received 1 October 2018; Received in revised form 11 February 2019; Accepted 11 February 2019
Available online 13 February 2019
1872-4973/

must be analyzed simultaneously to achieve similar power of discrimination or random match probability, when compared to their more polymorphic STRs counterparts [7]. Therefore, SNPs use in forensic DNA analysis for identification purposes have been limited due to the restricted capacity of simultaneous typing of different markers in a single run presented by CE method [8].

As an alternative, Massive Parallel Sequencing (MPS) techniques can be used for simultaneous typing of a large number of markers, with high throughput and consequent reduced analysis time [9]. Sample barcoding allows concurrent evaluation of significant number of samples, and process automation can be incorporated to minimize handling of criminal samples and improve workflow efficiency [10]. MPS achieves DNA typing on a single base resolution, and experimental design can be planned to assure high coverage. The HID-Ion AmpliSeq Identity Panel (Thermo Fischer Scientific, Waltham, MA, USA) commercial solution offers simultaneous amplification of 90 autosomal SNP markers and 34 Y upper-clade SNPs, designed to generate small size amplicons and successfully allow processing of highly degraded, low input and other forensic challenging samples [11]. Validation studies show that robust, efficient and reliable typing of forensic samples could be accomplished with different versions of this product, allowing its use in forensic individual identification or parentage testing [12–15]. Casework applications, however, demand proper knowledge regarding distribution and behavior of individual SNPs included in this panel in target populations, and several studies have been conducted to assess such features in specific populations [12,16–20].

Modern Brazilian population is a product of recurrent miscegenation processes occurred throughout national history, resulting in presently existent genetic structure where a high degree of genetic admixture is observed [21,22]. From the 16th century on, significant influxes of European (EUR) settlers (mainly from Iberian Peninsula) occurred, who interbred with Native American (NA) populations already present in Brazilian territory. Soon after, a large contingent of Africans (AFR), mostly from Western Africa Portuguese occupied territories, were forcedly brought to Brazil as slaves. Finally, late migratory movements took place in 19th and 20th centuries, with arrival of European (predominantly Germans, Italians, Portuguese and Spaniards) and Asian (essentially from Japan, Lebanon, Syria and other Mid-Western countries, then part of Ottoman Empire) migrants, giving rise to the highly admixed and multiethnic Brazilian population [23].

Brazilian territorial occupation and inland expansion were unevenly distributed over national history, following various patterns of multidirectional introgression according to social and historical conditions, and significantly vary for each distinct geographical region [23]. Heterogeneous processes of migratory flows led to marked differences in regional ethnical composition, and distinctive proportions of parental populations (NA, EUR and AFR) contribution in present-day geopolitical regions are noticeable [24]. Distribution of ethnic groups based on self-declared ancestry proportions differs considerably among regions, with relative higher Amerindian and African influence in Northern and Northeastern Brazil, respectively, while Southern Brazil displays prevalent European ancestry [25]. Asymmetrical contributions of ancestral populations to current genetic pool can also be observed in lineage markers, with predominant EUR Y-chromosome paternal contribution, and mitochondrial DNA (mtDNA) maternal lineages usually associated with AFR and NA origins [26,27]. Therefore, the possible occurrence of genetic structure concerning the specific SNP markers analyzed in this study in Brazilian populations must be investigated, with its inherent impact in forensic databases design, in order to avoid bias introduction in result interpretation. In the present study, Brazilian regional populations were compared to 26 different worldwide reference populations based on 88 ID-oriented SNPs included in HID-Ion AmpliSeq Identity Panel, aiming to scrutinize genetic diversity, similarity levels and population stratification for all investigated groups.

2. Materials and methods

2.1. Ethical statement

All samples analyzed in this study were voluntarily obtained following written informed consent from donors, and refer to a research-oriented human samples biobank available in PUC/RS. Anonymous samples were randomly selected from this bank, based on geographical origin only. This work follows the ethical principles stated in the Helsinki Declaration [28] of the World Medical Association, and was approved by Pontifical Catholic University of Rio Grande do Sul Institutional Review Board, under CAAE 52113715.9.0000.5336 number.

2.2. DNA samples, extraction and quantification

Oral swabs or peripheral blood were obtained from 432 unrelated volunteer donors from all five different Brazilian geopolitical regions. Male contributors were preferred, in order to assure a significant sample size for investigation of SNPs located in the Y chromosome included in this study. Number of samples obtained from each Brazilian region and its States are presented in Supplementary Figure S1. Genomic DNA was extracted using the automated platform Automate Express Forensic DNA Extraction System (Thermo Fischer Scientific Inc.) with the Prepfiler Automated Forensic DNA Extraction System (Thermo Fischer Scientific Inc.), or manually using DNA IQ System (Promega Corp., Madison, WI, USA). DNA was quantified with the Quantifiler Human DNA Quantification Kit (Thermo Fischer Scientific Inc.) on a 7500 Real-Time PCR System (Thermo Fischer Scientific Inc.) or using Qubit 2.0 Fluorometer with Qubit dsDNA HS Assay Kit (Thermo Fischer Scientific Inc.). All procedures followed manufacturer´s instructions.

2.3. Library preparation, emulsion PCR, and sequencing

DNA libraries were constructed using the Ion AmpliSeq Library Kit 2.0 (Thermo Fisher Scientific Inc.) combined with the HID-Ion AmpliSeq Identity Panel (Thermo Fisher Scientific Inc.), following manufacturer´s suggested protocol (HID-Ion Ampliseq Library Preparation, Revision C.0). Briefly, PCR amplification was performed in a final volume of 20 µL that contained 1 ng of template DNA, 4 µL of 5x Ion AmpliSeq™ HiFi Mix, and 10 µL of 2x Ion AmpliSeq™ primer pool (Identity Panel). Final volumes were adjusted with nuclease-free water. PCR was performed in a Veriti 96-well Thermal Cycler (Thermo Fischer 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. 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, followed by a 1 h hold at 10 °C. Ligation of adaptors to the libraries were executed according to the manufacturer´s instructions, using a different barcodes for each sample in a same run (Ion Xpress Barcode Adaptors 1–96 Kit or IonCode Barcode Adapters 1-384 Kit) (Thermo Fisher Scientific Inc.), and resulting products were purified using Agencourt AMPure XP reagents (Beckman Coulter Inc., Brea, CA, USA), according to the manufacturer´s instructions. After purification, libraries were quantified using a 7500 Real-Time PCR System (Thermo Fischer Scientific Inc.) with the Ion Library TaqMan™ Quantitation Kit (Thermo Fischer Scientific Inc.). Libraries were pooled to a final concentration of 20 pM. Emulsion PCR (emPCR) was performed 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 PGM™ (Thermo Fischer Scientific Inc.) Sequencer with the Ion PGM Hi-Q Sequencing Kit (Thermo Fisher Scientific Inc.) following

manufacturer´s instructions (Ion PGM Hi-Q Sequencing Kit, Revision B.0) and using Ion 318 Chip Kit v2 (Thermo Fisher Scientific Inc.). A total of six chips were used in distinct runs for complete samples genotyping.

2.4. Data analysis and Y-chromosome haplogrouping

Sequencing data were analyzed using the Torrent Suite Software v5.0 (Thermo Fischer). Hg19 was used as reference genome data. The number of mapped reads was calculated by the Torrent Coverage Analysis v5.0 plugin (Thermo Fischer Scientific Inc.). SNP genotypes were called by the HID SNP Genotyper v4.3.1 plugin, with target regions file (iiSNPv3_FP.hotspot) and the hotspot regions file (IISNPv2.20140429.Designed) under default analysis settings. Minimum coverage was set for 6 reads per base position. All SNPs genotypes were independently reviewed by two different collaborators, and manual corrections applied. The Y-chromosome haplogroups were determined automatically by the plugin using Y-SNP combination, and confirmed according to the International Society of Genetic Genealogy (ISOGG) Haplogroup Tree 2018 version 13.227 (available online at https://isogg.org/tree/). Y markers nomenclature is presented as depicted by the plugin. Equivalency between reference SNP ID numbers (according to NCBI dbSNP) [29] and marker names as classified by HID SNP Genotyper plugin is listed in Supplementary Table S1. Only variations present in haplogroups identified in this study were listed. A comprehensive tree of all possible Y-chromosome superior-clade haplogroups generated by the 34 Y-SNP markers included in this panel can be found in [30]. Generated data was submitted to NCBI SRA database as raw sequence reads, under reference number SRP155858.

2.5. Statistical data analysis

Statistical data analysis and results report were performed according to previously reported [19]. HID-Ion AmpliSeq™ Identity Panel Sequencing performance was evaluated using four different statistical parameters [14]: Locus Balance (LB) assesses individual locus amplification efficiency, and is calculated as the coverage of a locus divided by the mean coverage of all locus, per sample; Locus Strand Balance (LSB) measures the balance between forward and reverse reads, and is shown as the number of forward reads divided by the total reads number. Heterozygote Balance (HB) estimates the ratio between two distinct alleles read counts in heterozygous samples, and it was plotted as the number of reads of the first base call in alphabetical order (A, C or G) divided by the number of reads of the alternate allele (C, G or T, respectively). Noise Level (NL) represents the amount of unspecific base call, calculated as the ratio of non-allele calls divided by total coverage.

Several forensic parameters were calculated for the sample set in question, using STRAF online software (available online at http://cmpg.unibe.ch/shiny/STRAF/), as well as allele frequencies [31]. Forensic parameters included observed heterozygosity (Ho), expected heterozygosity (He), probability matching (PM), probability of exclusion (PE), polymorphism information content (PIC), power of discrimination (PD) and typical paternity index (TPI). Markers linkage disequilibrium (LD), Hardy-Weinberg equilibrium (HWE), Analysis of Molecular Variance (AMOVA) and pairwise F_{ST} among five Brazilian regions alone or with 26 worldwide populations were performed with the software Arlequin 3.5.2.2 [32]. Same software was used to compute exact test of population differentiation among Brazilian regions, with number of steps for dememorization and Markov Chain equals to 10,000 and 100,000, respectively. Populations employed in the present study, obtained from 1000 Genomes Project (Phases 1 to 3, available online at https://www.ensembl.org/Homo_sapiens/Info/Index) along with abbreviations used here, are listed in Supplementary Table S2. Multidimensional Scaling Analysis (MDS) based on pairwise F_{ST} distances and Principal Components Analysis (PCA) of worldwide populations allele frequencies were executed using IBM SPSS Statistics Software v.17.0 (IBM Analytics Inc.) [33]. Individual genotypes PCA was performed using the above-cited STRAF online software.

Occurrence of genetic structure among 26 worldwide populations [34] and Brazilian dataset was assessed with Structure v.2.3.4 software [35], with six independent runs for each K value, ranging from 2 to 7. A 100,000 length Burn-In Period was applied, followed by a 100,000 steps Monte Carlo Markov Chain (MCMC) procedure, using the standard admixed model. Results were then processed through CLUMPAK SERVER online tool (available online at http://clumpak.tau.ac.il) [36], in order to generate representative graphics and estimate best K value fitting the data. Pairwise F_{ST} distances using complete Brazilian data or regional subsets were also used to build cladogram graphics using the Molecular Evolutionary Genetics Analysis v.7.0 (MEGA v.7.0) software [37], applying Neighbor-Joining (N-J) methods.

Y-Haplogroup frequencies were determined by direct counting. Y network tree graphics were designed by construction of median-joining [38] networks based on 14 Y-SNPs based haplotypes found in Brazilian population using NETWORK 5.0.0.1 software (available online at http://www.fluxus-engineering.com/sharenet.htm).

3. Results

The HID Ion Ampliseq Identity Panel was successfully used to generate genetic profiles for 432 Brazilian samples, from all five geopolitical regions. This commercial kit comprises 90 autosomal biallelic SNPs specially selected to provide enough discrimination power for individual identification, achieving match probabilities equivalent to current commercial STR-based genotyping solutions. SNP markers were selected from previous studies (43 SNPs proposed in Kidd´s panel [39] and 48 from SNPforID effort [40]) due to presenting high heterozygosity levels in different worldwide populations. In addition to autosomal markers, 34 Y-chromosome SNPs were included in the panel, which allow an upper-clade haplogroup designation for major Y-chromosome tree groups. The kit was designed to properly handle critical or difficult forensic samples, with smaller amplicons sizes than usual STR markers (average read size for this study was 106 bp). Brazilian population stratification, forensic parameters, allele frequencies, Y-chromosome haplotype distribution and other relevant descriptive indexes associated with included SNPs, along with sequencing performance, were investigated in the present study, in order to assess the adequacy of present panel introduction in Brazilian routine forensic casework.

3.1. Forensic parameters of 124 SNPs for Brazilian population

Complete SNP genotypes of all 432 Brazilian individuals, for 124 SNPs comprised in this panel, are presented in Supplementary Table S3 (Supplementary Material). Obtained genetic profiles were employed to further characterize regional populations, with individuals from South, Southeast, Northeast, North and Center-West geopolitical regions (sample sizes of 108, 83, 81, 80 and 80 individuals, respectively) included in this study. Resulting allele frequencies for 90 autosomal SNPs are listed in Supplementary Table S4, including national (all regions combined) and regional data. Pairwise F_{ST} and exact test of population differentiation based on genotype frequencies were performed for all five Brazilian regions, and results are presented in Supplementary Table S5. No evidence of significant differences among Brazilian regions was found after Bonferroni correction, based solely on allele frequencies of 90 autosomal SNPs evaluated in the present study. Forensic parameters estimations for Brazilian national and regional data, including Ho, He, PM, PE, PD, PIC, TPI are presented in Supplementary Table S6, as well as p-values for HWE tests for all loci. No locus presenting significant deviations of HWE were found in investigated regional or national populations, after Bonferroni correction. Despite that, four autosomal SNP marker pairs displayed significant linkage equilibrium, even after Bonferroni correction, for Brazilian combined dataset (pairs rs4288409-rs12997453; rs964681-rs159606; rs1736442-rs1454361 and rs221956-

Table 1
Forensic parameters for Brazilian national and region populations, based on 90 autosomal SNPs included in HID Ion Ampliseq Identity Panel.

Population	Combined PM	Combined PE
Br-S	2.97143×10^{-36}	0.999999951474895
Br-SE	2.21577×10^{-36}	0.999999928310930
Br-NE	1.62475×10^{-36}	0.999999815034317
Br-N	2.73451×10^{-36}	0.999999926102221
Br-CW	3.15661×10^{-35}	0.999999992066246
BRA (all regions included)	9.93154×10^{-37}	0.999999917134468

rs2016276). Pairwise LD test p-values for Brazilian national population are detailed in Supplementary Table S7. Average Ho for each Brazilian region is estimated as follows: 0.473 (South), 0.465 (Southeast and North), 0.45 (Northeast) and 0.49 (Center-West). Maximum and minimum Ho values for the complete dataset were found both in Northeastern Brazil (0.679 at rs6811238 and 0.235 at rs1357617, respectively). Marker rs938283 was found to have the lowest discrimination power for Southern, Southeastern and Northern samples (MP = 0.527, 0.566 and 0.563, respectively), while markers rs7704770 (MP = 0.351), rs729172 (MP = 0.342) and rs214955 (MP = 0.34) were the most discriminative for the aforementioned regions, in same order. At the same time, a single locus (rs2016276) possesses smallest power of discrimination values for both Northeastern (MP = 0.556) and Center-Western (MP = 0.573) populations. Although this specific marker is listed among SNPs presenting linkage disequilibrium in Brazilian population, its linked counterpart, rs221956, displays intermediate discrimination values for the same groups (MP = 0.432 and 0.401, respectively). Still regarding Northeast and Center-West regions, SNPs rs1335873 (MP = 0.342) and rs1498553 (MP = 0.354) are the most discriminatory for above-named Brazilian regions. Table 1 exhibits combined matching probability (CMP) and combined power of exclusion (CPE) for each Brazilian region and for the country as a whole. As a comparison parameter for such exquisite discrimination power presented by the present panel, CMP for a 15 STR markers panel (including all 13 CODIS core markers) in Brazilian population was estimated to be around 18 orders of magnitude smaller than CMP provided by Ion Ampliseq Identity Panel for Brazilian population [41]. Above results suggest that autosomal SNP set included in Ion HID Ampliseq Identity Panel shows enough polymorphism and informative value for Brazilian regional and national populations, and therefore can also be incorporated in forensic analytical repertoire, as a useful tool in kinship testing or individual identification exams.

3.2. Y-haplotype frequencies distribution in Brazilian regional populations

34 Y-chromosomes SNPs included in Ion HID Ampliseq Identity Panel were used to determine superior-clade haplotype frequencies for all five Brazilian regions, in a total of 388 male samples. Haplotype classification was automatically provided by HID SNP Genotyper v4.3.1 plugin, and outputs were confirmed according to International Society of Genetic Genealogy (ISOGG) Haplogroup Tree 2017 version 12.334. In fact, a single misclassification was observed, but it was due a Y-tree restructuration implemented after ISOGG Haplogroup Tree 2014. This older Y-tree version used C to T transition in Y-SNP P198 to categorize samples belonging to then-existing O3 subclade. Since 2015 Y-tree version, P198 has been used as a characteristic marker for O2 Y-haplogroup, and corrected, current classification is presented in this paper.

A total of 14 different Y-chromosome haplogroups were observed among Brazilians in this study. Regional populations presented different distributions of identified haplotypes, with numbers ranging from nine distinct Y-clades found in Southeastern and Center-Western Brazil to only six haplotypes detected in Northeastern Brazil. Both Southern and Northern regions featured eight variants each for Y-chromosome. European haplogroups showed higher prevalence for all regions, with a significant predominance of R1b haplotypes (relative frequencies ranging from 42.5% to 57.34% among all five Brazilian subpopulations) and representing around half Y-haplogroup observations in the whole country. In addition to R1b, only four other male clades were detected with an overall frequency over 5% in national population, here listed according to their prevalence: E, I, J and G. Unique, single occurrence haplotypes are present in Southern, Southeastern and Center-Western samples only. Countrywide Y-haplogroup frequencies ranged from 0.26% (B, DE, L, N and O2 groups) to 50.26% for the most prevalent R1b. Supplementary Figure S2 shows a Y-haplogroups tree for clades observed in this study only, with Y SNPs markers whose mutations are relevant in respective classification displayed. Branch ends feature a pie plot, with area proportional to haplogroup frequency, and relative contribution of each region to number of observed cases. Table 2 details complete Y-haplogroup diversity, including distribution among Brazilian regional and national populations. Relative contribution of E haplogroup is higher in Northeastern and Center-Western populations than other Brazilian regions. Besides that, the number of samples associated with typical NA haplogroups (Q) were found to have a low prevalence (1.55%), limited to Northern and Center-Western individuals. This results confirms previous observations regarding differential contributions of ancestral populations for different Brazilian regions, with AFR or NA ancestry ratios more substantial in specific Brazilian areas, as well as uneven male and female

Table 2
Y-chromosome haplotypes frequencies found in 388 male Brazilian samples, based on 34 superior-clade Y-SNPs included in Ion HID Ampliseq Identity Panel.

Y-haplogroup	Region count (%)					
	South	*Southeast*	*Northeast*	*North*	*Center-West*	*Brazil*
B	–	1 (1.23)	–	–	–	1 (0.26)
D	–	2 (2.47)	–	–	–	2 (0.52)
DE	–	–	–	–	1 (1.25)	1 (0.26)
E	8 (10.67)	13 (16.05)	21 (26.25)	9 (12.5)	18 (22.5)	69 (17.77)
G	6 (8.0)	5 (6.17)	3 (3.75)	5 (6.94)	4 (5.0)	23 (5.93)
I	10 (13.33)	9 (11.11)	11 (13.75)	8 (11.11)	5 (6.25)	43 (11.08)
J	5 (6.67)	4 (4.94)	8 (10.0)	6 (8.33)	10 (12.5)	33 (8.5)
L	1 (1.33)	–	–	–	–	1 (0.26)
T	1 (1.33)	1 (1.23)	3 (3.75)	1 (1.39)	2 (2.5)	8 (2.06)
N	–	–	–	–	1 (1.25)	1(0.26)
O2*	1 (1.33)	–	–	–	–	1 (0.26)
Q	–	–	–	2 (2.78)	4 (5.0)	6 (1.55)
R1a1	–	1 (1.23)	–	3 (4.17)	–	4 (1.03)
R1b	43 (57.34)	45 (55.57)	34 (42.5)	38 (52.78)	35 (43.75)	195 (50.26)
Total	75 (100)	81 (100)	80 (100)	72 (100)	80 (100)	388 (100)

O2 haplogroup marked with * was identified by plugin as O3, and reclassified according 2018 version of ISOGG Y-tree.

parental lineage proportions, with majoritarian EUR Y-chromosome contribution for the country as a whole, in contrast to relative higher AFR and NA-associated mtDNA lineages present in current Brazilian genetic pool [21–27].

3.3. Inter-population analysis and genetic stratification

A series of analyses was conducted to evaluate differences among Brazilian regions distributions in genotype and allele frequencies, in order to assess potential population genetic stratification for autosomal SNPs typed. Pairwise F_{ST} values for 90 overlapped autosomal SNPs for Brazilian regions only (presented in Supplementary Table S5) ranged from 0.00024 (Br-S and Br-N) to 0.00672 (Br-N and Br-NE pair). Overall, Northeastern Brazil was found to be the most genetically distinct Brazilian region, with higher pairwise F_{ST} average observed. Despite that, fixation indexes calculated are relatively low, demonstrating homogeneity in allele frequency distribution of 90 autosomal SNPs included in HID Ion Ampliseq Identity panel throughout Brazilian regions. Observed degree of genetic similarity between Southern and Northern Brazilian regions was not expected, due to very distinct historical colonization processes both regions went through. This unexpected founding can be explained by an overrepresentation of samples coming from Northern state of Acre in this regional subset. This particular Brazilian State, which comprises only 4.65% of Northern Brazilian geopolitical region population [25], represents almost three-quarters of total Northern samples evaluated in this study. Acre State has a unique colonization history in Brazil, with significant migratory waves of Southern region settlers in late 20th century [42]. In this case, expected elevated NA relative contribution in Northern region, which was consistently detected previously [21–27], might have been overlooked in this study due to sampling bias, also responsible for overstating EUR genetic heritage in Northern region habitants. Addition of samples from the two most populated Northern Brazilian States Amazonas and Pará is underway, aiming to improve present frequency database reliability.

Pairwise F_{ST} values were also calculated based on 88 overlapped autosomal SNPs among Brazilian national or regional populations and 26 worldwide populations included in 1000 Genomes Project, and results for regional datasets are displayed as a heat map in Fig. 1 (a heat map for national data can be found in Supplementary Figure S3). For Brazilian population as a whole, results ranged from 0.00283 (Puerto Ricans) to 0.072 (Mende from Sierra Leone). Brazilians showed higher similarity levels with EUR populations, followed by samples from AMR countries. Considering Brazilian populations, Br-NE presented markedly lower F_{ST} values with AFR samples (results between 0.02 and 0.052), followed by Br-CW (0.03 to 0.067), when compared to other regions results (ranging 0.037 to 0.095). An inverse approach shows same regions with higher genetic distance to EUR populations (minimum value for Br-NE and Br-CW is 0.013 and 0.006 with IBS and TSI, respectively), than its regional counterparts (with values as low as 0.0022 for Br-S and IBS pair), with FIN group classified as the most genetically distant of Brazilian populations among Europeans (0.015 for the whole country). When considered other American populations data, Brazilian regional values present markedly resemblance to PUR (from 0.002 to 0.005 for Br-S and Br-NE, in order) and CLM (Fst of 0.002 with Br-N and 0.007 with Br-NE), with PEL samples demonstrating higher genetic distances (from 0.034 for Br-S and Br-SE to 0.041 to Br-NE). Both Asian groups presented intermediate genetic resemblance with Brazilian populations, with SAS samples displaying an overall genetic similarity with investigated regions slightly higher than EAS (average pairwise F_{ST} values to Brazilian regions equivalent to 0.026 and 0.044, respectively).

In order to further examine above results and clarify eventual genetic relationship between 26 worldwide populations and Brazilian regions, a MDS plot based on pairwise F_{ST} values was drawn for obtained data, and is presented in Fig. 2. A clear pattern can be observed, with negative values for Dimension 1 as a characteristic feature for AFR populations. Both Asians groups (EAS and SAS) are grouped together in upper-right quadrant, with apparent distinct cluster separation between both geographical origins. As expected, AMR and EUR have close distribution, occupying the fourth quadrant. However, while EUR populations seem to be grouped relatively close to each other, AMR and BRA have broader dispersion, with a clear directional tendency of Br-NE and Br-CW towards negative values in Dimension 1 (trending to AFR cluster). PEL seems to be an outlier, with extreme negative values for Dimension 2. Wider dispersion displayed by AMR populations apparently reflect their admixed nature, phenomena also observed for other geographical origins (as AFR, for instance, where groups of samples with African ancestry collected outside Africa (ACB and ASW) show a tendency towards axis origins. Although discrete, Br-SE location represents the Brazilian population closer to Asian groups, maybe reflecting the significant historical presence of immigrant from these particular areas in Southeastern Brazil [23,25]. A cladogram was also drawn applying N-J methods, and is presented in Supplementary Figure S4 (equivalent figures were constructed considering Brazilian population as a whole, displaying complete combined dataset, which are presented in Supplementary Figures S5-S6). Resulting phylogenetic tree shows Br-N and Br-S populations positioned closer to EUR branches, followed by other AMR admixed populations. Br-SE and Br-CW occupy intermediate positions between AMR populations and a unique ramification gathering all Asian groups. Furthermore, Br-NE is the single population closer to the tree bottom, where AFR clades were positioned. PCA plots were also executed to help evaluate relevance of genetic distance between populations or individuals. Supplementary Figure S7 shows a PCA analysis for Brazilian and worldwide populations based on allele frequencies of 88 overlapped autosomal SNPs, and top two principal components account for a total of almost 80% of total variation (65.35% and 14.28%). Once again, AFR and EUR are plotted in extreme positions, and Asian populations occupy closed positioned clusters. AMR and BRA have a sparser distribution, with intermediate locations. Supplementary Figure S8 shows a PCA plot based on individual genotypes of all samples included in this study, grouped by its geographical origin. Once again, AMR and BRA individuals were plotted alongside EUR samples. However, first principal component (5.67%) seems to correctly discriminate AFR samples, while second (3.23%) and third (1.97%) principal components might be useful in Asian samples classification. Low values for variation accounted by each principal component in this analysis suggests that most part of genetic differences among samples is located in an individual level, and only a small fraction of total genetic variation is due to differences in samples distinct continental geographical origins. Results above are consistent among all employed statistical methods (PCA, MDS and phylogenetic analyses), and display genetic relationships among all populations that are supported by genetic, historical or ethnographic information, especially for Brazilian case [21–27].

Bayesian inference methods were also used to further characterize differences among populations investigated in this study. STRUCTURE software applies MCMC methods, where admixing levels in different populations can be inferred. Fig. 3 presents STRUCTURE plot of Brazilian samples alongside 26 worldwide populations, with inferred K clusters ranging from 2 to 5. As the number of selected K clusters increases, diverse strata comprising general genetic structure of all populations can be accessed. Although ideal number of clusters was designated as K = 3 using Evanno´s method [43], traditional estimations based on Probability of Data for K = k suggests optimum population numbers as 5. Since this last estimative seems to reflect actual differences in evaluated groups, according to similarities presented by populations sharing a same geographical origin as identified by clustering patterns, it was adopted in this paper. For K = 5, investigated Brazilian populations shows clustering patterns similar to others AMR admixed populations, while AFR, EUR, EAS and SAS populations display a clear predominance of distinct clusters in overall components associated with

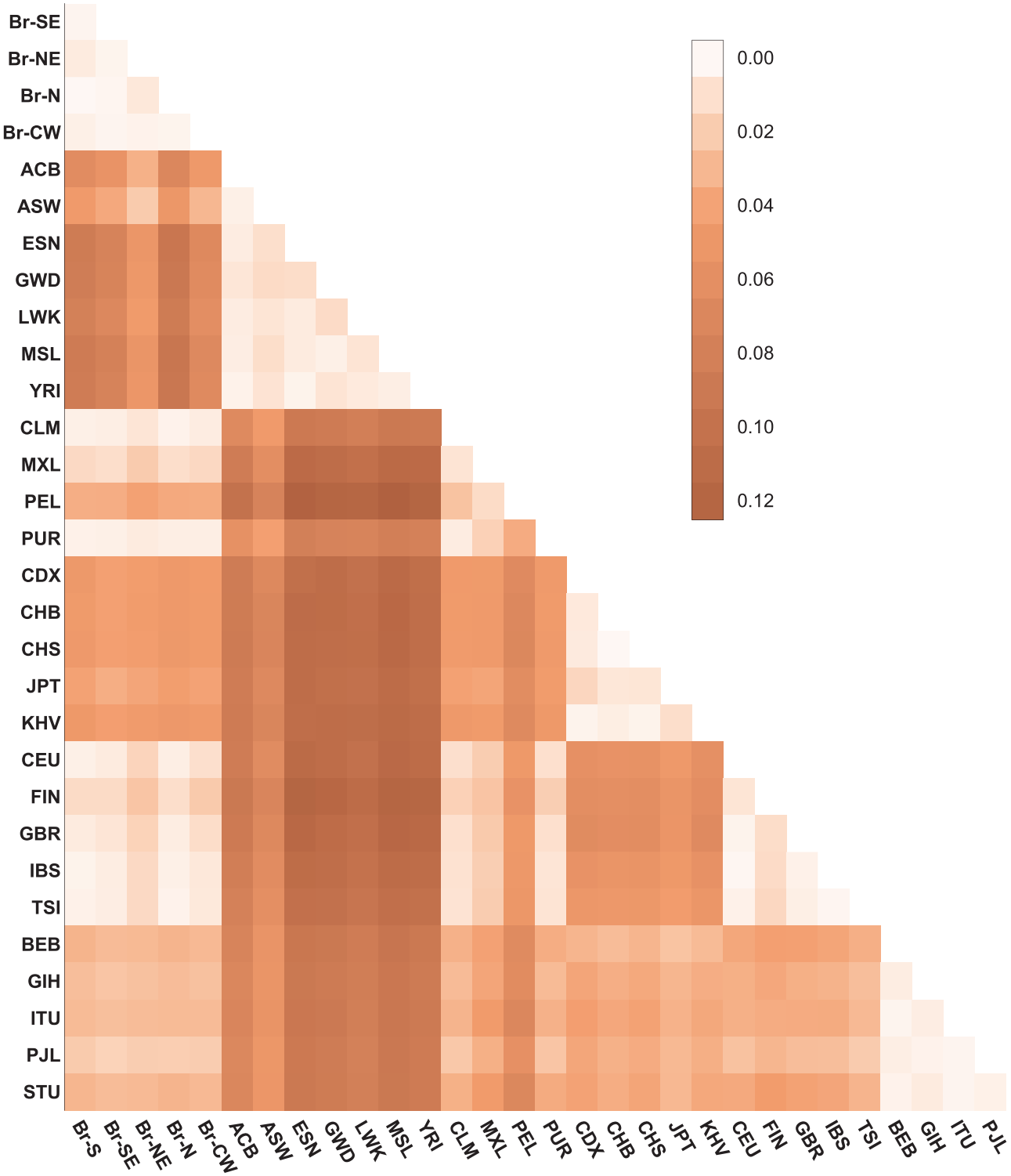


Fig. 1. Heat map of pairwise F_{ST} values for 88 overlapped autosomal SNPs included in HID Ion Ampliseq Identity Panel among 26 worldwide populations and Brazilian regional populations. Population full names described in Supplementary Table S2.

populations with the same geographical origin. Supplementary Table S8 summarizes relative contributions of each cluster for all Brazilians and worldwide populations evaluated, as well as show population geographical origin more strictly linked to each inferred cluster. Noteworthy, with K = 5, components associated with Cluster 2 (represented in red color in Fig. 3) seems to represent residual genetic structure not associated with clusters previously identified with smaller K numbers. Interesting enough, at least part of this residual genetic structure appears to be related to genetic contributions of NA ancestral populations, since admixed AMR populations present higher ratios of this particular cluster association. Brazilian populations show predominant affiliation with EUR and AMR clusters, with roughly 72% to 80% of genetic structure linked to both inferred groups. Above results for genetic Bayesian population inference are consistent with previously mentioned MDS, phylogeny and PCA analyses.

Finally, since occurrence of genetic stratification seems to be clearly

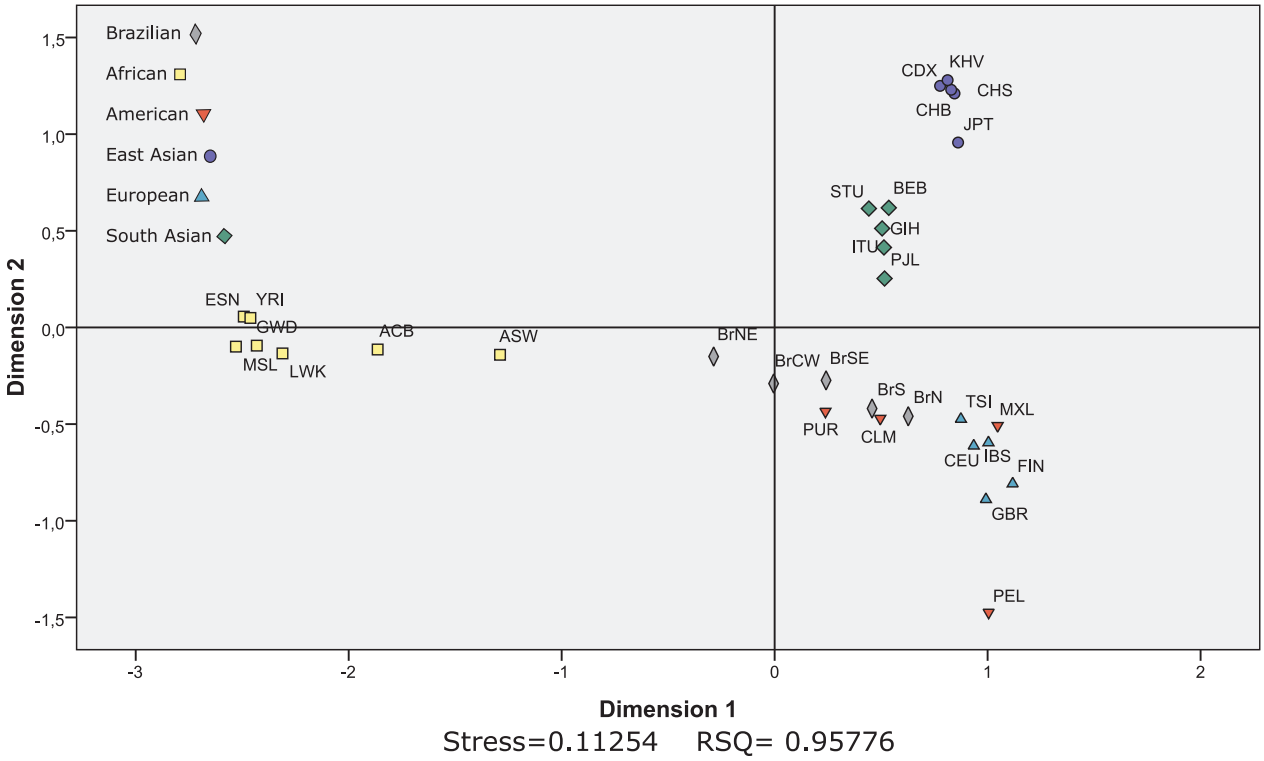


Fig. 2. Genetic distance evaluation for inter population analysis of 26 worldwide populations and Brazilian regions, presented as a MDS plot based on pairwise F_{ST} values for 88 overlapped autosomal SNPs included in HID Ion Ampliseq Identity Panel. Supplementary Figure S4 presents a phylogenetic tree build with Neighbor-joining methods, based on same pairwise F_{ST} values obtained.

defined for worldwide and, to a lesser extent, Brazilian regional populations, statistical tests were performed to evaluate the amount of genetic variance associated exclusively with verified genetic structure. Global AMOVA tests were executed with all five Brazilian regional and 26 worldwide populations, arranged according to geographical origin (Brazilian populations formed a separate group from remaining AMR populations) for 88 autosomal SNPs including in this panel, and results are detailed in Table 3. Within individuals variation accounts for an estimate of around 94% of total genetic variability, with a little over 5% of total genetic differences explained by among populations divergence. These results support convenience of employing HID Ion Ampliseq Identity Panel in forensic identification studies, since population genetic structure associated with autosomal SNP markers included in the kit seems to have a limited impact on affecting genetic diversity at

Table 3
Fixation indexes and global AMOVA results for Brazilian and 26 worldwide populations, based on individual genotypes of 88 autosomal SNPs included in HID Ion Ampliseq Identity Panel.

Source of Variation	Relative Variation (%)	Fixation Indexes
Among groups	5.11	F_{CT}: 0.05111
Among populations within groups	0.61	F_{SC}: 0.00641
Among individuals within populations	0.16	F_{IS}: 0.00170
Within individuals	94.12	F_{IT}: 0.05878
(Among populations)*	**94.28**	**F_{ST}: 0.05719**

* Among populations subsumes both categories accessing individual level variation.

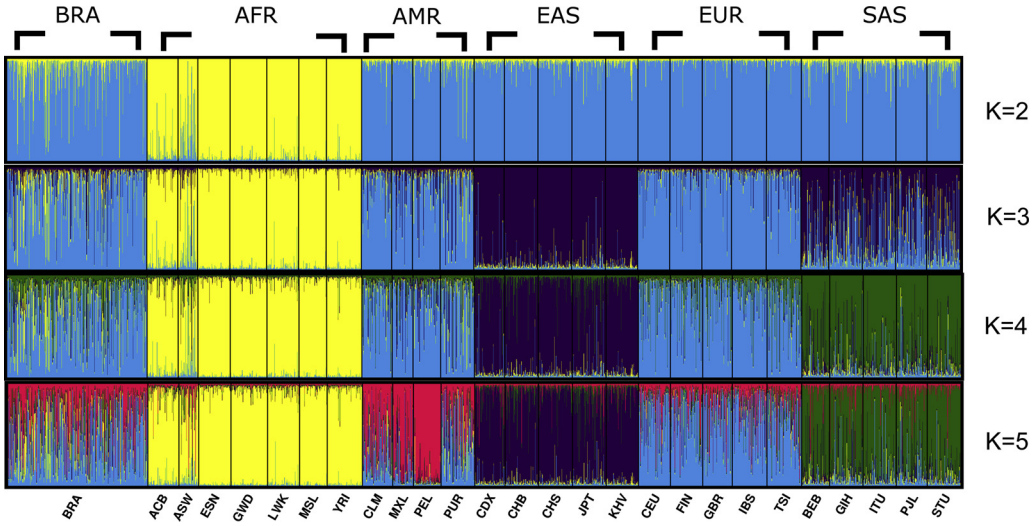


Fig. 3. Population structure of investigated Brazilian along with 26 worldwide populations based on 88 autosomal SNP markers included in HID Ion Ampliseq Identity Panel. Structure plots are presented from K cluster number ranging from 2 to 5 (top to bottom). Each vertical line stands for an individual, with each color representing relative proportion of association with each inferred cluster, assigned by distinct colors. Populations and geographical regions abbreviations are listed in Supplementary Table S2.

individual levels.

3.4. MPS sequencing performance

All samples in this study were genotyped through Ion PGM Sequencer, employing native data interpretation solutions provided by the manufacturer. HID SNP Genotyper plugin was the main pipeline for data calling. Additionally, Coverage Analysis plugin was used to access sequencing performance for all genotyping procedures. Since sample size largely exceed suggested maximum number of samples to be included in a single chip, all 432 samples were analyzed in six independent runs. Data regarding each individual experiment is presented in Supplementary Table S9. Maximum number of samples recommended by manufacturer to be included in a single run using 318v2 model chips (77 samples, calculated to ensure mean depth of 600x per locus) was slightly surpassed in three experiments, where 80 samples each per run were included. Despite that, mean depth per locus per sample range from 276x to 740x, with an average value of 476x, which can be considered satisfactory for reliable genotyping of single-source reference samples [13,14,44], where proposed coverage of as low as 20 reads might be considered sufficient for genotype determination [13]. Supplementary Figure S9 represents a heat map graphic representation of complete sample set (482 individuals) coverage for each evaluated locus. As expected, overall coverage for haploid Y-SNPs is consistently lower than diploid autosomal SNPs. Samples with no coverage for all Y-SNPs represents female individuals (44 total). Differences in coverage numbers can be noted for both individuals and markers, with low performance SNP markers identified by vertical lines presenting lighter color shades.

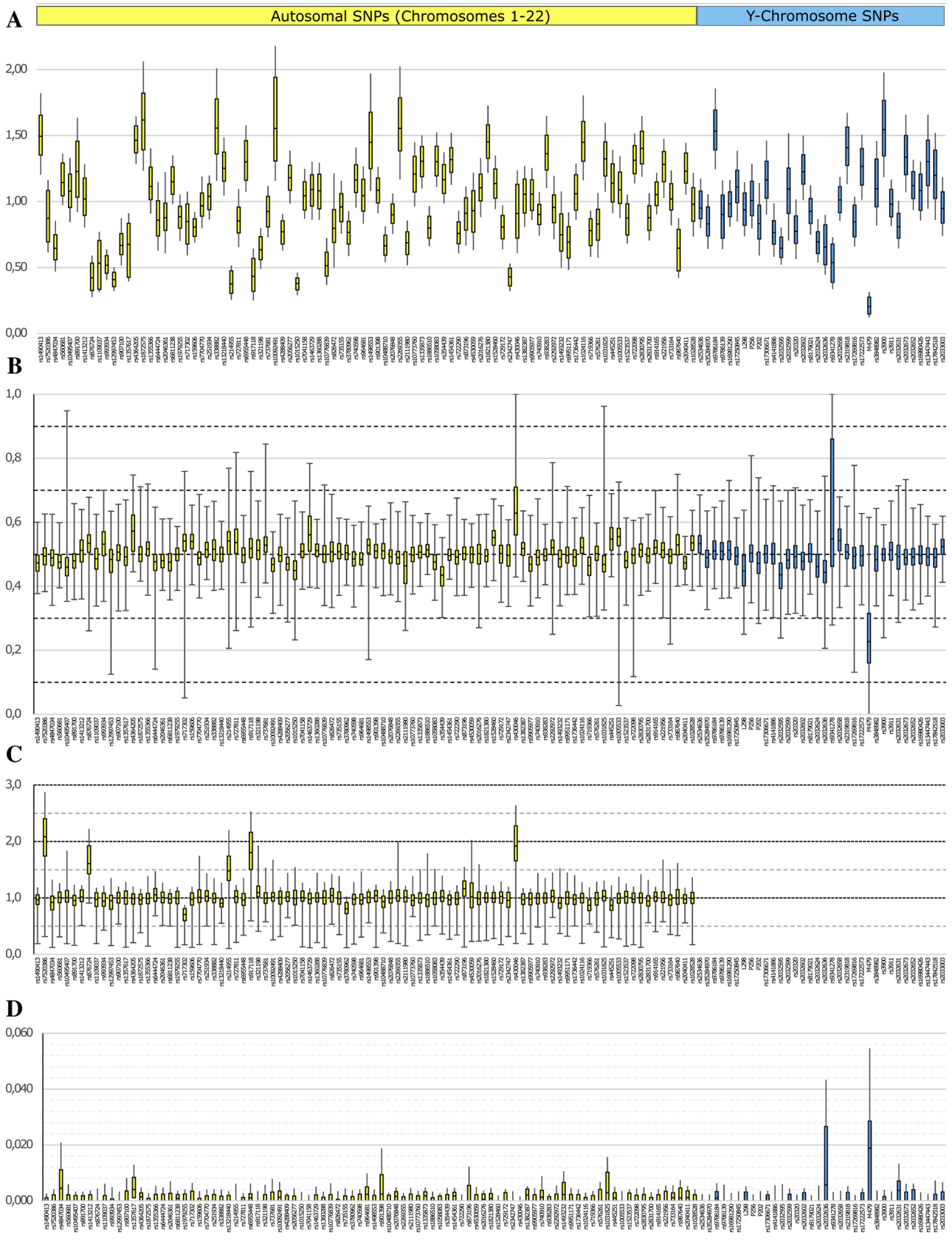
A series of metrics were applied to evaluate genotyping efficiency using HID Ion Ampliseq Panel, and results are summarized in Fig. 4. Average coverage across all 90 autosomal loci (527 ± 316 reads) is nearly twice as high as Y-SNPs (257 ± 172 reads). Therefore, LB calculations were performed independently for autosomal and Y-chromosome SNPs. As shown in Fig. 4A, LB median values for autosomal SNPs ranged from 0.38 at rs214955 and rs1015250 to 1.62 at rs1872575, while Y-SNP median LBs varies from 0.2 at M479 to 1.55 at rs3900. Overall average LB values for all 124 loci is 1.00 (Supplementary Table S10 presents details for individual markers). Six autosomal (rs876724, rs12997453, rs214955, rs917118, rs1015250 and rs2342747) and one Y (M479) markers provided poor LB performance, with values under 0.50 threshold, meaning coverage for this six loci are below half the average coverage. Overall inter locus coverage imbalance has been described for this panel, phenomena observed for different products versions [11–13,15–20,30]. This persistent issue might present obstacles with real forensic casework samples, especially concerning challenging samples, and efforts for achieving optimization of primers efficiency balance and global panel design must be addressed by manufacturer. LSB results ranged from 0.23 at M479 to 0.63 at rs430046, with an average value across all 124 loci equivalent to 0.50 ± 0.03, consistent with expected values (Fig. 4B and Supplementary Table S10). When manufacturer´s proposed thresholds are considered (0.50 ± 0.20), only outlier SNP among all loci is M479. As shown in Fig. 4C and Supplementary Table S10, average HB of all median values for 90 autosomal SNPs is 1.03 ± 0.19, with minimum and maximum values of 0.71 at rs717302 and 2.08 at rs7520386, respectively. Previous validation studies for this panel [13,14] propose HB acceptable thresholds in a 60:40 heterozygote ratio (HB values in 0.66 to 1.5 range), and obtained results show four SNPs (rs7520386, rs876724, rs917118 and rs430046) outside such conditions. Once again, impact of markers presenting low performance for HB metrics in forensic cases may be significant, especially in low-source DNA samples or mixture analysis, and data regarding this particular SNPs should be carefully scrutinized. Base calling efficiency, assessed by NL measures, where generally good, with most median values near zero (Fig. 4D and Supplementary Table S10) across all 124 loci. In a general manner, Y SNPs performed poorly than autosomal markers, including two SNPs (rs2032636 and M479) with 3rd quartile above 0.02, and overall larger median values, standard deviations and maximum observed values. Average median values were different than 0.00 for 22 markers (18 autosomal and 4 Y-SNPs), and only rs1357617 displays 1 st quartile value above zero. Highest NL median and relative value (25%) observed were both seen in M479 Y-SNP. Results found show that sequencing performance of HID Ion Ampliseq Identity Panel was generally good, with some problems observed for specific markers. Even though revised versions of this product have been released by manufacturer, all of them seem to present issues regarding locus and heterozygous balance [11,16–20]. Such questions must be addressed by manufacturer in order to deliver a more polished product, suitable for efficient use in routine forensic casework, especially considering critical sample analysis.

4. Discussion

Present study aims to evaluate utility of HID Ion Ampliseq Identity Panel in Brazilian regional populations, in order to develop a frequency database that allows use of this commercial solution in real forensic casework. 432 samples collected from all five Brazilian geopolitical regions were evaluated. Forensic statistical parameters obtained for all national populations show high polymorphism levels, suitable for human identification purposes.

Investigation of genetic stratification among Brazilian regions shows evidence of some degree of genetic structure nationwide, considering all 90 autosomal SNP markers included in the panel. However, the impact of observed stratification levels in database planning and construction is yet to be determined. A comprehensive analysis of available data for regional STR frequencies demonstrated that unified frequency databases can be used for Brazilian population as a whole, although some corrections might be necessary to adjust for observed levels of genetic structure existent for these markers [45]. Although no such statement can yet be made for polymorphisms investigated in this study, due to lack of sufficient information available for autosomal SNPs discussed here, present frequency database can be the foundation of a representative, accurate Brazilian SNP frequency database that allows reliable use of this set of markers in human identification applications and parentage testing.

Y-haplotypes frequency of major clades was also investigated for Brazilian population, through analysis of 34 Y-SNPs included in this panel. Findings are in accordance with previous studies for Brazilian populations, that demonstrate differential proportions in genetic contributions of immigrants from distinct geographical origins for all five Brazilian regions, according to historical colonization processes [21–27]. Furthermore, gender bias concerning male and female parental lineages distribution in Brazilian modern population, identified as predominance of Y-chromosome haplotypes with EUR origin or AFR and NA mtDNA-associated lineages relative overrepresentation, with diverse proportions in different Brazilian regions [26,27], was also verified within evaluated sample set. Obtained data for Y-haplotype frequencies supports above conclusions, with considerable preponderance of male lineages with EUR origin. In this context, it is important to question the relevance and utility of Y-chromosome haplotyping in forensic identification procedures focusing in Brazilian naturals. High levels of admixture are presented by Brazilians, and studies performed with ancestry informative markers (AIMs) suggest that most of identified Brazilian genetic variance is observed between individuals, rather than among population groups [46,47]. Same studies show geographical region, self-declared ethnicity or phenotypical features are bad ancestry predictors for Brazilian populations. Under such assumptions, the significant predominance of Y-chromosome EUR haplogroups (with recurrent R1b haplogroup representing over half national Y-haplogroup diversity) renders Y chromosome examination counterproductive, since no considerable amount of information on



(caption on next page)

Fig. 4. Sequencing performance of HID Ion Ampliseq Identity Panel. Data is presented as box-plots, with 1st and 3rd quartiles separated by median value in black. Open whiskers represent values for average plus standard deviation, while capped bars indicate maximum or minimum values. Autosomal SNPs presented in yellow, Y-chromosome in blue. (A) Locus balance for all SNPs, measured as depth of each locus divided by mean depth of all locus per sample. LB was calculated independently for autosomal and Y SNPs, and ideal value is equivalent to 1. (B) Locus strand balance for all SNPs, estimated as forward strand coverage divided by locus total coverage. 0.5 values are optimal, while 0.3 and 0.7 ratios (represented by dotted line) were considered lower and upper thresholds, respectively [13]. (C) Heterozygous balance for autosomal SNPs only. HB was calculated as coverage ratio of one allele to the other, in alphabetical order, for heterozygous genotypes only. Once again, optimal value equals 1. (D) Percentage of noise for all SNPs, measured as coverage of non-alleles divided by total locus coverage. Ideal value equals 0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

genetic diversity is added to results. Information on Y-haplogrouping in a forensic context might be useful for highly structured populations as North Americans [48]. In such cases, specific ethnic groups inside the country display different allele frequencies for forensic relevant markers, leading to construction of independent databases for each group [49]. No evidence of such degree of genetic stratification was found in present work, nor described so far for populations from Brazil. Therefore, from a Brazilian perspective, substitution of Y-haplogroup SNPs included in HID Ion Ampliseq Identity Panel for other highly-informative, forensic relevant markers might be advantageous, in order to improve discrimination efficiency without reducing overall genotyping capacity (measured as expected depth for each locus in a multi-sample run) and subsequent sequencing performance. Evaluation of possible genetic structure occurrence among different Y-haplogroups was considered, but executed analyses (not shown) were not able to find proof of such divergences, if present. This situation is affected by low diversity in Y-haplogroups observed in Brazilian populations, and low number of non-AFR and non-EUR samples identified. Database expansion is underway, and new samples addition will be helpful in testing this hypothesis.

Inter population analysis of allele frequencies among Brazilian regions and 26 worldwide populations also shed light on interesting issues about investigated sample set. Despite being designed to be neutral and have no significant variations among geographical regions [39,40], autosomal SNPs included in this panel have some capacity to differentiate intercontinental populations through inter populations analysis. This phenomenon has been previously described [19], but application to Brazilian regional context revealed that different colonization processes distinct regions went through had implications on genetic structure presented by people in these areas, at least to some extent. As verified, sample selection bias might have masked such effect for Brazilian Northern region. However, performed tests were able to evaluate deeper levels of genetic structure not discussed by Liu et al. [19], especially concerning higher numbers of inferred populations. Results of interpopulation structure analysis seem to suggest that, at K = 5, some of the genetic structure observed might be derived from NA ancestral genetic contribution to modern AMR populations. Cluster presented in red color in Fig. 3 is distributed across all continental origins, and can be seen as the residual genetic diversity not detected with K = 4. Despite some relevance in European populations (around 20% of relative contribution for EUR group), this particular cluster comprises most of the genetic variation in AMR populations, with values of especially in 34.5%, 52.7%, 67.5% and 90.8% for PUR, CLM, MXL and PEL populations, respectively. For Brazilians, this cluster encompasses 30.3% of genetic structure identified. These results are in general agreement with a meta-analysis of comparative ancestry in American populations, based on AIM data obtained from several studies [50]. NA relative admixture proportions for above-mentioned populations are, in order: 14%, 44%, 62% and 92%, while overall estimate for Brazil is 17%. Despite discrepancies for BRA and PUR populations (where excess of this particular cluster contribution, when compared to AIM-based NA ancestry, might have origins in EUR ancestry component of both groups, due to fraction of this cluster contribution to EUR observed genetic structure), overall correlation between NA ancestry and this cluster contribution for AMR populations is acceptable.

In conclusion, results described in present paper suggest that HID Ion Ampliseq Identity Panel might be a useful tool to be incorporated to Brazilian forensic analytical repertoire. However, some issues still need to be addressed, including not only technical optimizations in panel design (especially aiming improvements regarding inter and intralocus balance). Adequacy of this reagent use in Brazilian real forensic casework must include attention to national genetic particularities, including but not limited to regional idiosyncrasies, which can significantly influence questions like database conception, implementation cost, and other relevant questions.

Conflict of interest

Authors declare they have no conflict of interest.

Financial support

Present work was funded by grants provided by CAPES-BR (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Edital 25/2014 – Pró-Forenses), CNPq-BR (Conselho Nacional de Pesquisa Científica - Chamada n° 16/2014 – INCT) and FAPERGS-BR (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul).

CRediT authorship contribution statement

Eduardo Avila: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization, Project administration. **Aline Brugnera Felkl:** Formal analysis, Writing - review & editing. **Pietra Graebin:** Investigation, Resources, Data curation. **Cláudia Paiva Nunes:** Investigation, Project administration. **Clarice Sampaio Alho:** Conceptualization, Resources, Data curation, Writing - review & editing, Project administration, Supervision, Funding acquisition.

CRediT authorship contribution statement

Eduardo Avila: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization, Project administration. **Aline Brugnera Felkl:** Formal analysis, Writing - review & editing. **Pietra Graebin:** Investigation, Resources, Data curation. **Cláudia Paiva Nunes:** Investigation, Project administration. **Clarice Sampaio Alho:** Conceptualization, Resources, Data curation, Writing - review & editing, Project administration, Supervision, Funding acquisition.

Acknowledgments

Authors would like to thank collaborators CP Cavalheiro, DB Silva, F Sawitski and CEI Gonçalves for assistance with sample processing and data acquisition, L Ribeiro for map design and MG Bicalho for providing samples used in this work.

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.2019.02.012.

References

[1] Y. Tan, P. Bai, L. Wang, H. Wang, H. Tian, H. Jian, R. Zhang, Y. Liu, W. Liang, L. Zhang, Two-person DNA mixture interpretation based on a novel set of SNP-STR markers, Forensic Sci. Int. Genet. 37 (2018) 37–45, https://doi.org/10.1016/j.fsigen.2018.07.021.

[2] E. Murphy, Forensic DNA typing, Annu. Rev. Criminol. 1 (2018) 497–515, https://doi.org/10.1146/annurev-criminol-032317-092127.

[3] P.M. Schneider, Beyond STRs: the role of diallelic markers in forensic genetics, Transfus. Med. Hemotherapy (39) (2012) 176–180, https://doi.org/10.1159/000339139.

[4] T.J. Parsons, R. Huel, J. Davoren, C. Katzmarzyk, A. Milos, A. Selmanović, L. Smajlović, M.D. Coble, A. Rizvić, Application of novel "mini-amplicon" STR multiplexes to high volume casework on degraded skeletal remains, Forensic Sci. Int. Genet. 1 (2007) 175–179, https://doi.org/10.1016/j.fsigen.2007.02.003.

[5] John M. Butler, Coble M.D. JM, P.M. Vallone, STRs vs. SNPs: thoughts on the future of forensic DNA testing, Forensic Sci. Med. Pathol. 3 (3) (2007) 200–205, https://doi.org/10.1007/s12024-007-0018-1.

[6] B. Budowle, A. van Daal, Forensically relevant SNP classes, Biotechniques 5 (44) (2018) 603–610, https://doi.org/10.2144/000112806.

[7] A. Amorim, L. Pereira, Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs, Forensic Sci. Int. 1 (150) (2005) 17–21, https://doi.org/10.1016/j.forsciint.2004.06.018.

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

[9] K.M. Canturk, R. Emre, K. Kınoglu, B. Başpınar, F. Sahin, M. Ozen, Current status of the use of Single-Nucleotide Polymorphisms in forensic practices, Genet. Test. Mol. Biomarkers 7 (18) (2014) 455–460, https://doi.org/10.1089/gtmb.2013.0466.

[10] M.L. Kampmann, A. Buchard, C. Børsting, N. Morling, High-throughput sequencing of forensic genetic samples using punches of FTA cards with buccal swabs, Biotechniques 3 (61) (2016) 149–151, https://doi.org/10.2144/000114453.

[11] E. Salata, A. Agostino, I. Ciuna, S. Wooton, L. Ripani, A. Berti, Revealing the challenges of low template DNA analysis with the prototype Ion AmpliSeq™ Identity panel v2.3 on the PGM™ Sequencer, Forensic Sci. Int. Genet. 22 (2016) 25–36, https://doi.org/10.1016/j.fsigen.2015.07.011.

[12] C. Børsting, S.L. Fordyce, J. Olofsson, H.S. Mogensen, N. Morling, Evaluation of the Ion Torrent™ HID SNP 169-plex: a SNP typing assay developed for human identification by second generation sequencing, Forensic Sci. Int. Genet. (2014), https://doi.org/10.1016/j.fsigen.2014.06.004.

[13] M. Eduardoff, C. Santos, M. de la Puente, T.E. Gross, M. Fondevila, C. Strobl, B. Sobrino, D. Ballard, P.M. Schneider, A. Carracedo, M.V. Lareu, W. Parson, C. Phillips, Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the Ion PGM, Forensic Sci. Int. Genet. 12 (2014) 144–154, https://doi.org/10.1016/j.fsigen.2015.04.007.

[14] A. Buchard, M.L. Kampmann, L. Poulsen, C. Børsting, N. Morling, ISO 17025 validation of a next-generation sequencing assay for relationship testing, Electrophoresis 21 (37) (2016) 2822–2831, https://doi.org/10.1002/elps.201600269.

[15] K.A. Meiklejohn, J.M. Robertson, Evaluation of the precision ID identity panel for the ion torrent™ PGM™ sequencer, Forensic Sci. Int. Genet. 48 (31) (2017) 56, https://doi.org/10.1016/j.fsigen.2017.08.009.

[16] F. Guo, Y. Zhou, H. Song, J. Zhao, H. Shen, B. Zhao, F. Liu, X. Jiang, Next generation sequencing of SNPs using the HID-Ion AmpliSeq™ identity panel on the ion torrent PGM™ platform, Forensic Sci. Int. Genet. 25 (2016) 73–84, https://doi.org/10.1016/j.fsigen.2016.07.021.

[17] O. García, A. Soto, I. Yurrebaso, Allele frequencies and other forensic parameters of the HID-Ion AmpliSeq™ Identity Panel markers in Basques using the Ion Torrent PGM™ platform, Forensic Sci. Int. Genet. (28) (2017) e8–e10, https://doi.org/10.1016/j.fsigen.2017.03.010.

[18] R. Li, C. Zhang, H. Li, R. Wu, H. Li, Z. Tang, C. Zhen, J. Ge, D. Peng, Y. Wang, H. Chen, H. Sun, SNP typing using the HID-Ion AmpliSeq™ Identity Panel in a southern Chinese population, Int. J. Legal Med. 4 (132) (2018) 997–1006, https://doi.org/10.1007/s00414-017-1706-3.

[19] J. Liu, Z. Wang, G. He, X. Zhao, M. Wang, T. Luo, C. Li, Y. Hou, Massively parallel sequencing of 124 SNPs included in the precision ID identity panel in three East Asian minority ethnicities, Forensic Sci. Int. Genet. (35) (2018) 141–148, https://doi.org/10.1016/j.fsigen.2018.05.002.

[20] S. van der Heijden, S.J. Oliveira, M.L. Kampmann, C. Børsting, N. Morling, Comparison of manual and automated AmpliSeq™ 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.

[21] 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.

[22] 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.

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

[24] 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), https://doi.org/10.1371/journal.pone.0017063 e17063.

[25] IBGE, Atlas do Censo Demográico 2010. Rio de Janeiro, Instituto Brasileiro de Geografia e Estatística., 2013 216 p.

[26] 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.

[27] 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), https://doi.org/10.1371/journal.pone.0152573 e0152573.

[28] 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.

[29] J. McEntyre, J. Ostell (Eds.), The NCBI Handbook, 2nd edition, National Center for Biotechnology Information, Bethesda (MD), 2013201 p..

[30] E. Ochiai, K. Minaguchi, P. Nambiar, Y. Kakimoto, F. Satoh, M. Nakatome, K. Miyashita, M. Osawa, Evaluation of Y chromosomal SNP haplogrouping in the HID-Ion AmpliSeq™ identity panel, Legal Med. (Tokyo) 22 (2016) 58–61, https://doi.org/10.1016/j.legalmed.2016.08.001.

[31] A. Gouy, M. Zieger, STRAF - A convenient online tool for STR data evaluation in forensic genetics, Forensic Sci. Int. Genet. 30 (2017) 148–151, https://doi.org/10.1016/j.fsigen.2017.07.007.

[32] 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. 10 (3) (2010) 564–567, https://doi.org/10.1111/j.1755-0998.2010.02847.x.

[33] IBM, IBM SPSS Statistics for Windows, Version 17.0. IBM Corp., Armonk, NY, 2008.

[34] 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.

[35] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics. 2 (155) (2000) 945-959. 155:945–959.

[36] N.M. Kopelman, J. Mayzel, M. Jakobsson, N.A. Rosenberg, I. Mayrose, CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K, Mol. Ecol. Resour. 5 (15) (2015) 1179–1191, https://doi.org/10.1111/1755-0998.12387.

[37] 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.

[38] H.J. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies, Mol. Biol. Evol. 1 (16) (1999) 37–48, https://doi.org/10.1093/oxfordjournals.molbev.a026036.

[39] K.K. Kidd, A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L. Kajuna, N.J. Karoma, S. Kungulilo, J.J. Kim, R.B. Lu, A. Odunsi, F. Okonofua, J. Parnas, L.O. Schulz, O.V. Zhukova, J.R. Kidd, Developing a SNP panel for forensic identification of individuals, Forensic Sci. Int. 1 (164) (2006) 20–32, https://doi.org/10.1016/j.forsciint.2005.11.017.

[40] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 9 (27) (2006) 1713–1724, https://doi.org/10.1002/elps.200500671.

[41] V.R. Aguiar, E.V. Wolfgramm, F.S. Malta, A.G. Bosque, A.C. Mafia, V.C. Almeida, F.A. Caxito, V.C. Pardini, A.C. Ferreira, I.D. Louro, Updated Brazilian STR allele frequency data using over 100,000 individuals: an analysis of CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, Penta d, Penta e, TH01, TPOX and vWA loci, Forensic Sci. Int. Genet. 4 (6) (2012) 504–509, https://doi.org/10.1016/j.fsigen.2011.07.005.

[42] T.M.R. Machado, Migrantes sulistas: Caminhadas e aprendizados na região Acreana, Edufac, Rio Branco, 2016.

[43] G. Evanno, S. Regnaut, J. Goudet, Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study, Mol. Ecol. 8 (14) (2005) 2611–2620, https://doi.org/10.1111/j.1365-294X.2005.02553.x.

[44] R. Daniel, C. Santos, C. Phillips, M. Fondevila, R.A. van Oorschot, A. Carracedo, M.V. Lareu, D. McNevin, A SNaPshot of next generation sequencing for forensic SNP analysis, Forensic Sci. Int. Genet. (14) (2015) 50–60, https://doi.org/10.1016/j.fsigen.2014.08.013.

[45] T. Hessab, R.S. Aranha, R.S. Moura-Neto, D.J. Balding, C.G. Schrago, Evaluating DNA evidence in a genetically complex population, Forensic Sci. Int. Genet. (36) (2018) 141–147, https://doi.org/10.1016/j.fsigen.2018.06.019.

[46] F.S.N. Manta, R. Pereira, R. Vianna, R.B. Araújo, D.L.G. Gitaí, D.A. Silva, E.V. Wolfgramm, I.M. Pontes, I.J. Aguiar, O.M. Moraes, E.F. Carvalho, L. Gusmão, Revisiting the genetic ancestry of Brazilians using autosomal AIM-Indels, PLoS One 9 (8) (2013), https://doi.org/10.1371/journal.pone.0075145 e75145.

[47] 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.

[48] M.F. Hammer, V.F. Chamberlain, V.F. Kearney, D. Stover, G. Zhang, T. Karafet, B. Walsh, A.J. Redd, Population structure of Y chromosome SNP haplogroups in the United States and forensic implications for constructing Y chromosome STR databases, Forensic Sci. Int. 1 (164) (2006) 45–55, https://doi.org/10.1016/j.forsciint.2005.11.013.

[49] C.D. Steele, D.J. Balding, Choice of population database for forensic DNA profile analysis, Sci. Justice 6 (54) (2014) 487–493, https://doi.org/10.1016/j.scijus.2014.10.004.

[50] R.R. Moura, A.V. Coelho, V.Q. Balbino, S. Crovella, L.A. Brandão, Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries, Am. J. Hum. Genet. 5 (27) (2015) 674–680, https://doi.org/10.1002/ajhb.22714.