Forensic Science International: Genetics 33 (2018) 129–135

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Genetic analysis of Southern Brazil subjects using the PowerSeq™ AUTO/Y system for short tandem repeat sequencing

Deborah S.B.S. Silva[a,b,*], Fernanda R. Sawitzki[c], Melissa K.R. Scheible[a,b], Sarah F. Bailey[a,b], Clarice S. Alho[c], Seth A. Faith[a,b]

[a] Molecular Biomedical Sciences, North Carolina State University, 1060 William Moore Dr., Raleigh, NC, 27607, USA
[b] Forensic Sciences Institute, North Carolina State University, 1060 William Moore Dr., Raleigh, NC, 27607, USA
[c] Laboratory of Human and Molecular Genetics, Pontificia Universidade Catolica do Rio Grande do Sul,Av. Ipiranga 6681, Porto Alegre, RS, 90619-900, Brazil

ARTICLE INFO

Keywords:
Short tandem repeat
PowerSeq™ AUTO/Y system
South Brazil
Sequence variation

ABSTRACT

With the advent of Next-Generation Sequencing technology, sequencing of short tandem repeats (STRs) allows for a more detailed analysis when compared to size-based fragment methods (capillary electrophoresis-CE). The implementation of high-throughput sequencing can help uncover deeper genetic diversities of different populations. Subjects from the South region of Brazil present a particular and more homogeneous ancestry background when compared to other regions of the country. Both autosomal and Y- STRs have been analyzed in these individuals; however, all analyses published to date encompass data from CE-based fragment analysis. In this study, a genetic analysis of 59 individuals from Southern Brazil was performed on STR sequences. Forensically relevant STRs were PCR-enriched using a prototype of the PowerSeq™ AUTO/Y system (Promega Corp.). Next-generation sequencing was performed on an Illumina MiSeq instrument. The raw data (FASTQ files) were processed using a custom designed sequence processing tool, *Altius*. Isoalleles, which are sequence-based allelic variants that do not differ in length, were observed in nine autosomal and in six Y- STRs from the core global forensic marker set. The number of distinctive alleles based on sequence was higher when compared to those based on length, 37.3% higher in autosomal STRs and 13.8% higher in Y-STRs. The most polymorphic autosomal locus was D12S391, which presented 38 different sequence-based alleles. Among the loci in the Y chromosome, DYS389II presented the highest number of isoalleles. In comparison to CE analysis, Observed and Expected Heterozygosity, Polymorphic Information Content (PIC) and Genetic Diversity also presented higher values when the alleles were analyzed based on their sequence. For autosomal loci, Polymorphic Information Content (PIC) was 2.6% higher for sequence-based data. Diversity was 9.3% and 6.5% higher for autosomal and Y markers, respectively. In the analysis of the repeat structures for the STR loci, a new allele variant was found for allele 18 in the vWA locus. The STR flanking regions were also further investigated and sixteen variations were observed at nine autosomal STR loci and one Y-STR locus. The results obtained in this study demonstrate the importance of genetic analysis based on sequencing and highlight the diversity of the South Brazilian population when characterized by STR sequencing.

1. Introduction

Next-Generation Sequencing (NGS) technology is becoming part of the routine for many molecular biology areas. In the forensic field, even though its implementation is still in the early stages, many studies have been performed in order to better interpret, validate and apply NGS to forensic caseworks [1–3]. Further, this technology is opening doors to better understand genetic diversities in different populations [4,5].

Populations around the world have been typed and analyzed by length-based STR methods, such as capillary electrophoresis (CE). However, with the advances in high-throughput sequencing methods, sequence-based studies of individuals from different populations become imperative, mostly to understand and discover their allelic variance and genetic diversity in a deeper and more detailed analysis.

Brazil is a large country divided in five major regions, and the Brazilian population was formed by many waves of migration from different countries, mainly from the European, African and Asian continents, which contributed to its high genetic admixture. A recent study by Naslavsky et al. [6] analyzed data from 609 elderly Brazilian individuals in order to characterize genetic variants in this group. As a

* Corresponding author at: NC State University, 1060 William Moore Dr., Raleigh, NC, 27607, USA.
E-mail address: silva_deborah@hotmail.com (D.S.B.S. Silva).

https://doi.org/10.1016/j.fsigen.2017.12.008
Received 7 September 2017; Received in revised form 12 December 2017; Accepted 15 December 2017
Available online 16 December 2017
1872-4973/

result of the high degree of genetic diversity among individuals from Brazil, the authors identified 207,621 variants that were absent from major public international databases, including 1000 genomes and dbSNP.

The South region in Brazil, which includes the State of Rio Grande do Sul, is composed of individuals that present a particular and more homogeneous ancestry background when compared to other regions of the country: the majority of subjects with European ancestry (especially Portuguese, Italian, Spanish, and German) and a smaller number of individuals with African and Amerindian ancestry [7]. South Brazil population has been typed and analyzed by different forensic length-based STR systems [8–13]. However, given the growing applications of NGS in the forensic field, and to better understand the genetic variation in forensically relevant STR loci, sequence-based studies should be performed.

In this study, a deeper genetic analysis was performed in 59 individuals from the State of Rio Grande do Sul, in Southern Brazil, on forensic STR sequences using Next-Generation Sequencing.

2. Materials and methods

2.1. DNA samples

Buccal swabs were collected from 59 South Brazilian subjects, 30 females and 29 males. Genomic DNA was extracted using NucleoSpin® Tissue kit (Macherey-Nagel, Bethlehem, PA), following the protocol as described by the manufacturer, and quantified using Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA). This project was approved by the Research Ethics Committee of the Pontifical Catholic University of Rio Grande do Sul (protocol #11-05722, Of.CEP-0295/12 and Of.CEP-1041/12) and by the NC State Internal Review Board (IRB protocol 6569). An informed written consent was completed and signed by all participants.

2.2. Amplification and sequencing

DNA samples were amplified with the PowerSeq™ Auto/Y System Prototype (Promega Corporation, Madison, WI) as described by van der Gaag et al. [14], in a *Veriti*™ Thermal Cycler (Applied Biosystems, Foster City, CA). A total of 0.5 ng was targeted for PCR input, and Component B from Standard Reference Material 2391c (National Institute of Standards and Technology, Gaithersburg, MD) was used as positive control.

Post-PCR clean-up was performed with Agencourt AMPure XP SPRI beads (Beckman Coulter, Brea, CA) using the epMotion 5075tc Liquid Handling Workstation (Eppendorf, Hamburg, Germany). Libraries were prepared with the KAPA Hyper Prep Kit (Kapa Biosystems, Wilmington, Massachusetts), following an automated PCR-free protocol and using custom adapters manufactured by Integrated DNA Technologies (Coralville, Iowa). Libraries were then pooled in equal volumes and the final pooled libraries were quantified using the KAPA Library Quantification kit for Illumina platforms (KAPA Biosystems, Wilmington, MA), following the manufacturer recommended protocol, and the StepOnePlus Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA). Sequencing was performed on an Illumina MiSeq instrument with MiSeq V3 sequencing chemistries (400 cycles) (Illumina, Inc., San Diego, CA), according to manufacturer's recommendations, and a 15% PhiX spike to account for low diversity in the pooled library. Adapter trimming and demultiplexing of sample libraries were performed post sequencing using the Illumina MiSeq software. Raw sequence data in the FASTQ format were generated for each sample library and used in downstream analysis.

2.3. Sample analysis: bioinformatics, quality control and statistics

The raw data (FASTQ files) were processed using the *Altius* Cloud system [15], which is built upon the Amazon Web Services (AWS) platform. The *Altius* system identifies target regions in the flanking regions of the STR sequences and produces an output data following the nomenclature guidance of Parson et al. [16].

For quality control of NGS data, a filtering method based on percentage of reads within a locus was applied as an analytical threshold. This way, for an allele to be called, it had to present at least 15% of the total of the locus. Also, a profile was only approved if it did not exceed a maximum of 3 loci with heterozygous imbalance of 50%. Profiles were independently analyzed by two reviewers and were then compared electronically to check for any discrepancies. The last step of the QC process was the comparison of the NGS data with capillary electrophoresis (CE) data. To obtain the size-based STR genotypes, PCR amplification was performed using commercially available STR kits: AmpFLSTR™ Identifiler™ (Applied Biosystems, Foster City, CA), PowerPlex® Fusion System and PowerPlex® Y23 System (Promega Corporation, Madison, WI). Amplification products were separated on an ABIPRISM® 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA) with POP-4 matrix in a capillary of 36 cm. The analysis was performed using GeneMapper ID software version 4.0 (Applied Biosystems, Foster City, CA) and GeneMarker HID v2.9.0 (SoftGenetics, State College, PA). For the detailed study of STR sequences, statistical parameters, including alleles frequencies, Hardy–Weinberg equilibrium test, observed and expected heterozygosity, polymorphism information content (PIC) and diversities, were assessed for both size- and sequence-based results, using CERVUS version 3.0.7 [17] and GenAlEx 6.5 [18,19]. The two packages were used to double-check all data and confirm if all output were identical. Y-STR haplotypes/haplogroups were estimated using an online tool (http://www.nevgen.org) and the STR flanking sequences were aligned using CLC Genomics Workbench 10 (Qiagen, Hilden, Germany).

3. Results and discussion

Two sequencing runs were performed and a total of 59 libraries were constructed for this study. The first run consisted of 96 pooled indexed libraries and the second one consisted of 67 pooled indexed libraries. In the first run, 54 libraries were part of this current study and the other libraries were part of another study. In the second run, only 5 libraries were part of this study. For the first sequencing run (I), the pooled indexed libraries were at a concentration of 27.52 nM, and for the second run (II) the pooled indexed libraries were at a concentration of 11.93 nM. Libraries concentration were above the requirements of the Illumina sequencing protocol, which allowed for a proper dilution within the correct range of loading concentration. After the sequencing run, high coverage and total reads matching STRs, which means number of sequencing reads that match the STR sequences analyzed, were obtained. Summary of each run is presented in Table 1.

Across all STR loci and Amelogenin, the average DoC (Depth of Coverage) was 1798 reads. The average DoC for each locus ranged from 98 to 4438 reads (Fig. 1). Among Y-STRs, DYS448 presented the lowest performance and DYS392 had the highest number of total reads. Among autosomal loci, D2S1338 and D22S1045 consistently showed the lowest and highest number of reads, respectively.

Allele Coverage Ratio (ACR) was calculated for all autosomal loci and DYS385ab, as the remaining Y-STRs only have one allele (Fig. 2).

Table 1
Summary of Sequencing Runs I and II.

Sequencing Run	Loading concentration (pM)	Chemistry version	Number of cycles	Cluster density (K/mm^2)	Total reads passing filter
I	10	3	400	897	19,589,518
II	15	3	400	1140	23,177,362

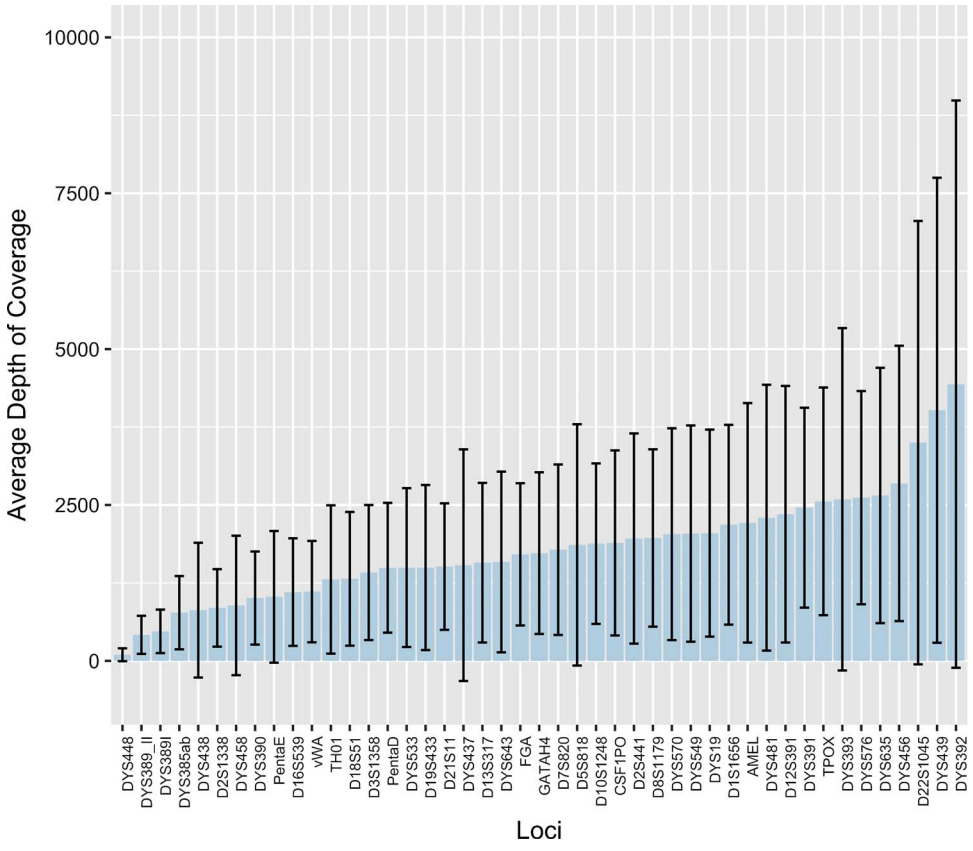

Fig. 1. Average Depth of Coverage (DoC) for Amelogenin and STR loci. Standard deviation is presented as error bar.

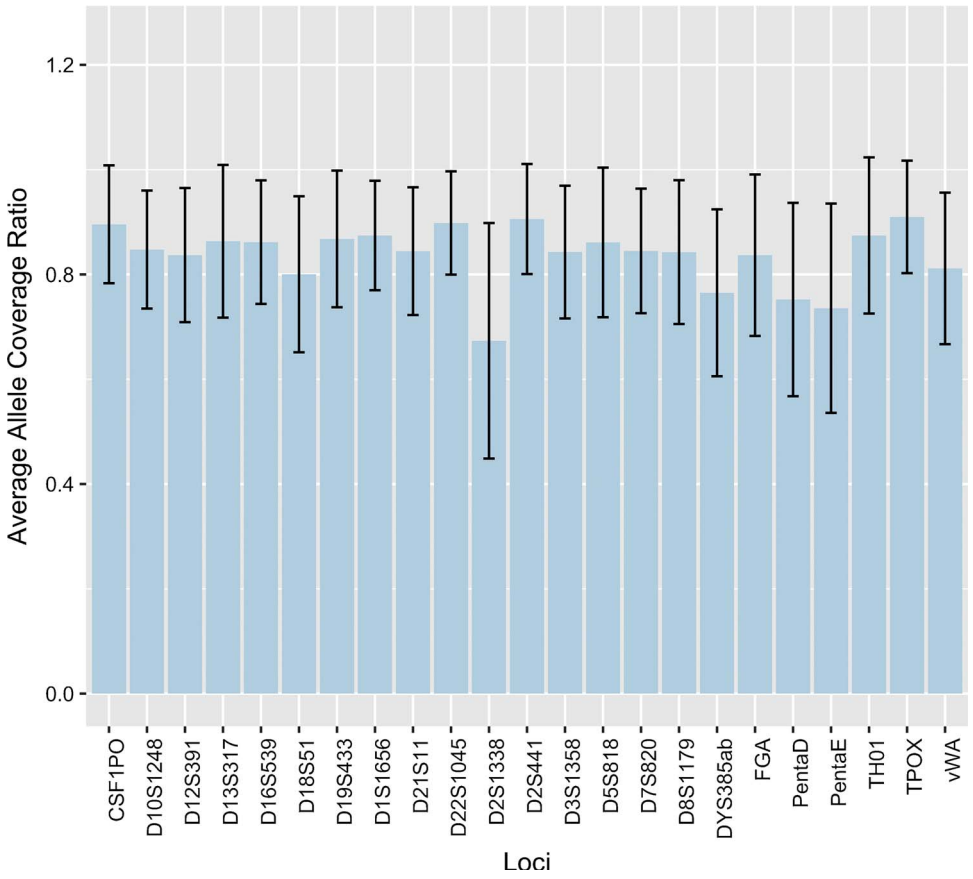

Fig. 2. Average Allele Coverage Ratio (ACR) for Amelogenin, all autosomal loci and DYS385ab. Standard deviation is presented as error bar.

To calculate ACRs, the number of reads for the lower coverage allele was divided by the number of reads for the highest coverage allele at each locus. Overall, the average ACR for all heterozygotes was 0.84. D2S1338 had the lowest ACR with an average of 0.67 and presented the most allele imbalance observed in the dataset due to a lower coverage for the longer alleles.

3.1. STR analysis by length

Autosomal- and Y- STR loci were analyzed based on their length. Allele frequencies for the NGS data were calculated and are presented in Supplementary Tables S1 and S2; for locus DYS385ab genotype frequencies are shown. NGS data were compared to CE data and full concordance was observed after analysis with both GeneMarker HID v2.9.0 (SoftGenetics, State College, PA) and GeneMapper ID v4.0 (Applied Biosystems, Foster City, CA).

All data based on length was also compared against previously published CE data on individuals from South Brazil and the results showed similarity and concordance to the published data [9–13]. For the dataset in this study, tests for Hardy-Weinberg Equilibrium (HWE) in autosomal loci detected that one locus deviated from expectations: D18S51 (Supplementary Table S3). The deviation was also observed after Bonferroni correction. When analyzing the distribution of the genotypes for D18S51 there is a homozygote deficiency in the population tested. The presence of a higher than expected number of heterozygotes, analyzed by sequence-based methods, can be explained by variation in individual admixture in this tested population. This is likely to happen by chance, given the small population size. So, although some HWE deviation was observed for the D18S51 locus, we would expect that in a larger population sampling the HWE would not deviate from expected. Deviations were also observed in previous publications [11–13] but authors stated that no significance difference was found after Bonferroni correction. Also, for autosomal markers, observed and expected heterozygosity and polymorphism information content (PIC) were calculated (Supplementary Table S4). Results obtained in this study are in concordance with previously published size-based data on South Brazil populations [11–13]. For Y markers, diversities and haplotypes/haplogroups were estimated and are presented in Supplementary Tables S5 and S6. Results obtained for the Y loci are also shown to be in concordance with previously published size-based data [[9–13],10].

3.2. STR analysis by sequence

Next-generation sequencing allows a much deeper sample analysis and the generation of detailed sequence information. Because previously published analyses with individuals from South Brazil were performed with CE-based fragment data [8–13], sequence variations of Autosomal and Y- STRs were investigated. Sequence-based allelic variants were observed in nine autosomal and in six Y loci (Fig. 3). These variants, also known as isoalleles, present difference in their sequences but no change in length (size). With the detection of variant alleles of the same size, the total number of alleles in the samples analyzed increased 37.3% and 13.8% for Autosomal and Y- STRs, respectively. New allele frequencies were calculated and are presented in Supplementary Tables S7 and S8.

Observed and expected heterozygosity, Shannon's Information Index and PICs were calculated for sequence-based alleles for the autosomal STRs and results were compared with those obtained for length-based alleles (Table 2). All parameters assessed presented higher values for sequencing data, showing an increased observed allelic richness and portraying how a more detailed genetic analysis enables allelic diversity and genotypic variation. Locus D12S391 presented the highest number of isoalleles, increasing the total number of alleles by more than double of the number of length-based alleles (Table 2). Locus FGA presented only one isoallele and therefore the parameters assessed in this study did not show any significant change.

Shannon's Information Index and diversity were calculated for sequence-based alleles for the Y-STRs. Results were compared with those obtained for length-based and, similar to results obtained for autosomal STRs, the sequence-based data presented highest values for all parameters assessed (Table 3). This shows that even a small change in allele sequence variants can have an impact in the genetic diversity of these South Brazil individuals when STR sequences are investigated. One locus, however, showed an increase of three times in the number of total different alleles: DYS389II (Table 3). Such diversity in a Y-STR can be helpful in forensic and civil cases, such as paternity disputes, in this population. Haplotypes were first constructed with NGS data based on length and only unique haplotypes were found (Supplementary Table S6). Thus, no extra haplotype or haplogroups were estimated using NGS data based on sequence.

The sequences structures, including all allelic sequence variants, of the Autosomal and Y-STRs are presented in Supplementary Table S9. These results are in concordance with those reported by other sequence-based studies [2,4,16,20,21], except for one variant in the vWA locus: allele 18 – [TAGA]13[CAGA]4**[TAGT]**. To our knowledge, this sequence has not been published elsewhere. The authenticity of this new variant was confirmed by two separate sequencing runs. The discovery of a new isoallele in this sample of individuals emphasizes how sequence-based genetic studies are important in populations from all regions of Brazil and other countries.

A small number of male samples presented a sequence variant for the DYS393 locus that has a cytosine as the first nucleotide in the repeat sequence instead of an adenine (Supplementary Table S9). However, in this study, the sequences were not considered as isoalleles since the first nucleotide in the repeat structure is a degenerate base for the forward primer that extends into the beginning of the sequence. Therefore, some samples presented the degenerate base as the first nucleotide in the sequence repeat. Since no second sequencing method was used in this study, the sequences observed here cannot be considered as real variants.

3.3. STR flanking regions

STR flanking regions comprise the sequence between the primer binding sites and the core repeat. The importance of analyzing these regions is to provide additional sequence variation and increase the genetic diversity to further individual differentiation. In this study, STR flanking regions were analyzed for sequence variations previously reported by Parson et al. [16]. Sixteen variations were observed at nine autosomal STR loci and one Y-STR locus (Table 4). Among the Y STRs, DYS437 was the only locus that presented sequence variation in the 5′ flanking region. Among autosomal STRs, the highest number of flanking region variations was observed at locus D8S1179. Two SNP positions were observed in D8S1179 flanks, rs116377919 and rs191524697. The former was a complex form of tri-allelic alleles and a variation (G/T/C; deletion), while the later was bi-allelic (G/T).

Interestingly, four of nine loci with flanking region variations (D1S1656, D2S1338, D8S1179 and vWA) also presented sequence variations within the core repeat. The combination of multiple genetic polymorphisms can help improve discrimination by increasing the number of genotypes available for analysis; combined genotypes (flankings + core repeat) can now be generated creating even more diversity within a population.

4. Conclusion

Individuals from the State of Rio Grande do Sul, in South Brazil, have been part of genetic and population studies based on STR analysis with size-based capillary electrophoresis. However, there is a lack of data on sequencing analysis of forensically relevant loci in the Southern Brazil population. With the Next-Generation Sequencing technology,

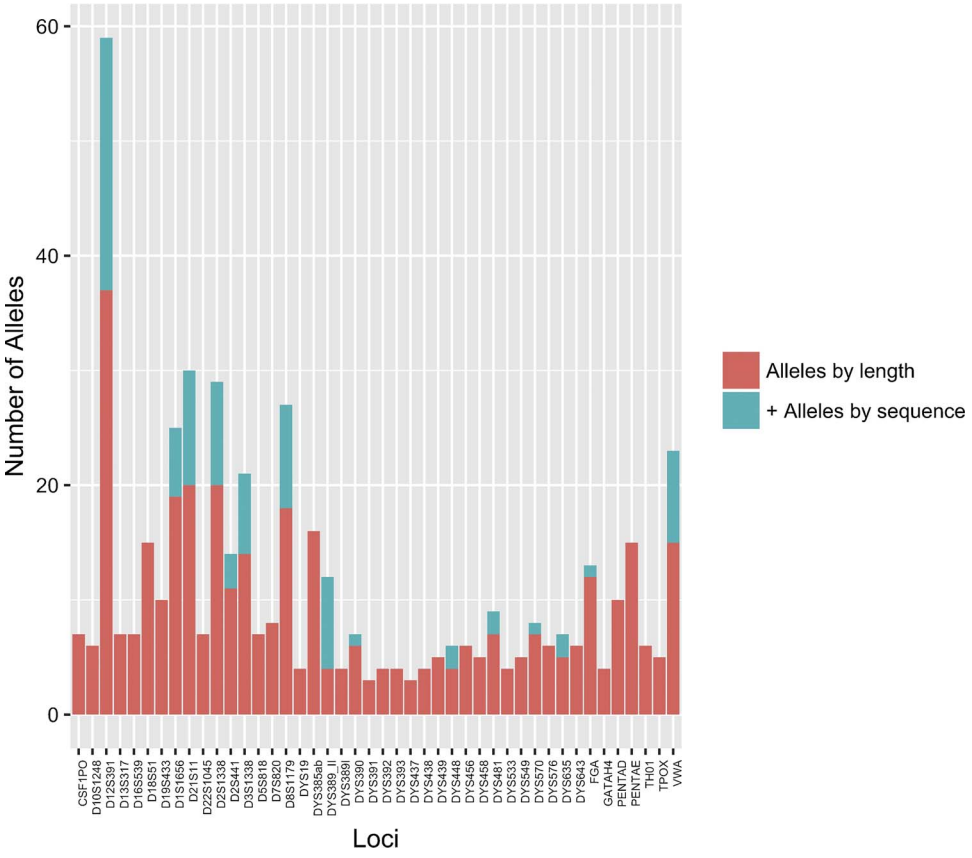

Fig. 3. Number of alleles obtained by CE (size-based) + NGS (sequence-based).

Table 2
Analysis by locus of 22 autosomal STRs in individuals from South Brazil: alleles based on length *vs* alleles based on sequence.

Locus	N	Alleles based on length					Alleles based on sequence				
		Na	I	HObs	HExp	PIC	Na	I	HObs	HExp	PIC
CSF1PO	118	7	1.454	0.627	0.739	0.685	7	1.454	0.627	0.739	0.685
D10S1248	118	6	1.544	0.864	0.763	0.718	6	1.544	0.864	0.763	0.718
D12S391	118	15	2.333	0.915	0.889	0.871	**37**	**3.144**	**0.932**	**0.938**	**0.927**
D13S317	118	7	1.688	0.797	0.784	0.746	7	1.688	0.797	0.784	0.746
D16S539	118	7	1.547	0.763	0.749	0.704	7	1.547	0.763	0.749	0.704
D18S51	118	15	2.265	0.949	0.886	0.866	15	2.265	0.949	0.886	0.866
D19S433	118	10	1.819	0.661	0.793	0.758	10	1.819	0.661	0.793	0.758
D1S1656	118	13	2.327	0.932	0.894	0.876	**19**	**2.566**	**0.949**	**0.910**	**0.895**
D21S11	118	10	1.987	0.915	0.844	0.817	**20**	**2.545**	**0.966**	**0.909**	**0.893**
D22S1045	118	7	1.356	0.746	0.690	0.629	7	1.356	0.746	0.690	0.629
D2S1338	118	11	2.174	0.932	0.875	0.854	**20**	**2.473**	**0.949**	**0.894**	**0.877**
D2S441	118	8	1.515	0.729	0.744	0.694	**11**	**1.732**	**0.763**	**0.776**	**0.737**
D3S1338	118	7	1.639	0.831	0.797	0.758	**14**	**2.167**	**0.864**	**0.867**	**0.845**
D5S818	118	7	1.310	0.729	0.689	0.625	7	1.310	0.729	0.689	0.625
D7S820	118	8	1.744	0.864	0.811	0.776	8	1.744	0.864	0.811	0.776
D8S1179	118	9	1.871	0.847	0.831	0.801	**18**	**2.433**	**0.898**	**0.901**	**0.884**
FGA	118	11	2.079	0.864	0.868	0.845	**12**	**2.105**	**0.864**	**0.869**	**0.846**
PENTAD	118	10	1.941	0.797	0.844	0.816	10	1.941	0.797	0.844	0.816
PENTAE	118	15	2.447	0.831	0.905	0.889	15	2.447	0.831	0.905	0.889
TH01	118	6	1.626	0.780	0.796	0.757	6	1.626	0.780	0.796	0.757
TPOX	118	5	1.072	0.627	0.596	0.516	5	1.072	0.627	0.596	0.516
VWA	118	7	1.674	0.847	0.796	0.758	**15**	**2.072**	**0.881**	**0.835**	**0.808**
Summary	Mean	9.136	1.791	0.811	0.799	0.762	12.545	1.957	0.823	0.816	0.782
	SE	0.656	0.079	0.020	0.017	0.095	1.566	0.109	0.022	0.019	0.106

Data shown in **BOLD** represent calculations where sequence-based alleles yield differences compared to length-based alleles.
N = Number of Chromosomes; Na = Number of Different Alleles; I = Shannon's Information Index = −1* Sum (pi * Ln (pi)); HObs = Observed Heterozygosity = No. of Hets/N; HExp = Expected Heterozygosity = (2N/(2N-1)) * (1 − Sum pi^2); PIC = polymorphic information content; pi is the frequency of the ith allele for the population & Sum pi^2 is the sum of the squared population allele frequencies.

Table 3
Analysis by locus of 23 Y-STRs in individuals from South Brazil: alleles based on length vs alleles based on sequence.

Locus	N	Alleles based on length			Alleles based on sequence		
		Na	I	Diversity	Na	I	Diversity
DYS19	29	4	1.254	0.700	4	1.254	0.700
DYS389_II	29	4	1.362	0.764	**12**	**2.366**	**0.929**
DYS389I	29	4	1.096	0.631	4	1.096	0.631
DYS390	29	6	1.606	0.805	**7**	**1.699**	**0.818**
DYS391	29	3	0.817	0.549	3	0.817	0.549
DYS392	29	4	0.943	0.584	4	0.943	0.584
DYS393	29	4	0.949	0.507	4	0.949	0.507
DYS437	29	3	1.041	0.650	3	1.041	0.650
DYS438	29	4	1.172	0.670	4	1.172	0.670
DYS439	29	5	1.231	0.663	5	1.231	0.663
DYS448	29	4	1.146	0.663	**6**	**1.405**	**0.722**
DYS456	29	6	1.500	0.768	6	1.500	0.768
DYS458	29	5	1.359	0.732	5	1.359	0.732
DYS481	29	7	1.560	0.751	**9**	**1.729**	**0.783**
DYS533	29	4	0.792	0.446	4	0.792	0.446
DYS549	29	5	1.057	0.606	5	1.057	0.606
DYS570	29	7	1.711	0.815	**8**	**1.789**	**0.823**
DYS576	29	6	1.567	0.798	6	1.567	0.798
DYS635	29	5	1.331	0.707	**7**	**1.549**	**0.751**
DYS643	29	6	1.466	0.756	6	1.466	0.756
GATAH4	29	4	1.060	0.611	4	1.060	0.611
DYS385ab (genotype)	29	16	2.243	0.833	16	2.243	0.833
Summary	Mean	5.273	1.285	0.682	6.000	1.368	0.697
	SD	0.567	0.072	0.022	0.658	0.090	0.025

Data shown in **BOLD** represent calculations where sequence-based alleles yield differences compared to length-based alleles.
N = Number of Chromosomes; Na = Number of Different Alleles; I = Shannon's Information Index = −1* Sum (pi * Ln (pi)); HObs = Observed Heterozygosity = No. of Hets/N; HExp = Expected Heterozygosity = (2N/(2N-1)) * (1 − Sum pi^2); PIC = polymorphic information content; pi is the frequency of the ith allele for the population & Sum pi^2 is the sum of the squared population allele frequencies.

Table 4
Sequence variations in flaking regions of Autosomal- and Y-STRs.

Locus		SNP	Position (GRCh38)	Ancestral Allele	Variant	Frequency
Autosomal						n = 118
	D1S1656	rs4847015	1:230769689	G	T	0.330
	D2S1338	rs6736691	2:218014824	C	A	0.330
	D5S818	rs182073376	5:123775520	A	C	0.008
	D5S818	rs25768	5:123775612	G	A	0.127
	D7S820	rs7789995	7:84160204	A	T	0.093
	D7S820	rs7789995	7:84160204	A	-/deletion	0.059
	D7S820	rs16887642	7:84160286	G	A	0.059
	D8S1179	rs116377919	8:124894824	G	T	0.059
	D8S1179	rs116377919	8:124894824	G	C	0.025
	D8S1179	rs116377919	8:124894824	G	-/deletion	0.186
	D8S1179	rs191524697	8:124894835	G	T	0.042
	D13S317	rs9546005	13:82148069	A	T	0.398
	D16S539	rs11642858	16:86352761	A	C	0.983
	D22S1045	rs190864081	22:37140272	G	T	0.008
	vWA	rs75219269	12:5983970	A	G	0.084
Y						n = 29
	DYS437	rs9786886	Y:12346264	C	T	0.034

DNA samples from 59 South Brazilian subjects were analyzed with higher resolution using sequencing data of short tandem repeats (STRs) markers. Additional sequence-based alleles were found for both autosomal and Y markers. An increase in heterozygosity and in overall genetic diversity was observed and a new sequence variant for the vWA locus was discovered.

Even though this study is not a comprehensive population study due to the number of samples, it is an important work that highlights the diversity of the South Brazilian population. The results obtained demonstrate the importance of genetic analysis based on STR sequencing and how individuals from the State of Rio Grande do Sul present a much higher sequence-based genetic diversity when compared to STRs based on fragment analysis.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This project was supported by Award No. 2015-DN-BX-K062, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this website are those of the authors(s) and do not necessarily reflect those of the Department of Justice. We thank Promega Corporation for providing PowerSeq™ reagents.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2017.12.008.

References

[1] X. Zeng, J. King, S. Hermanson, J. Patel, D.R. Storts, B. Budowle, An evaluation of the PowerSeq™ Auto System: a multiplex short tandem repeat marker kit compatible with massively parallel sequencing, Forensic Sci. Int. Genet. 19 (2015) (2015) 172–179.

[2] K.B. Gettings, K.M. Kiesler, S.A. Faith, E. Montano, C.H. Baker, B.A. Young, R.A. Guerrieri, P.M. Vallone, Sequence variation of 22 autosomal STR loci detected by next generation sequencing, Forensic Sci. Int. Genet. 21 (2016) 15–21.

[3] R.S. Just, L.I. Moreno, J.B. Smerick, J.A. Irwin, Performance and concordance of the ForenSeq™ system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens, Forensic Sci. Int. Genet. 28 (2017) 1–9 (F.R.).

[4] J.D. Wendt, N.M.M. Churchill, J.L. Novroski, J. King, R.F. Ng, K.L. Oldt, J.A. McCulloh, D.G. Weise, S. Smith, B. Budowle Kanthaswamy, Genetic analysis of the yavapai native americans from West-Central Arizona using the illumina MiSeq FGx™ forensic genomics system, Forensic Sci. Int. Genet. 24 (2016) 18–23.

[5] E.H. Kim, H.Y. Lee, S.Y. Kwon, E.Y. Lee, W.I. Yang, K.J. Shin, Sequence-based diversity of 23 autosomal STR loci in Koreans investigated using an in-house massively parallel sequencing panel, Forensic Sci. Int. Genet. 30 (2017) 134–140.

[6] M.S. Naslavsky, G.L. Yamamoto, T.F. Almeida, S.A.M. Ezquina, D.Y. Sunaga, N. Pho, D. Bozoklian, T.O.M. Sandberg, L.A. Brito, M. Lazar, D.V. Bernardo, E. Amaro Jr., Y.A.O. Duarte, M.L. Lebrao, M.R. Passos-Bueno, M. Zatz, Exomic variants of an elderly cohort of Brazilians in the ABraOM database, Hum. Mutat. 00 (2017) 1–13.

[7] F.C. Parra, R.C. Amado, J.R. Lambertucci, J. Rocha, C.M. Antunes, S.D. Pena, Color and genomic ancestry in Brazilians, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 177–182.

[8] V.R.C. Aguiar, E.V. Wolfgramm, F.S.V. Malta, A.G. Bosque, A.C. Mafia, V.C.O. Almeida, F.A. Caxito, V.C. Pardini, A.C.S. 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. 6 (2012) 504–509.

[9] S.P. Schwengber, T. Kommers, C.H.F. Matte, P.E. Raimann, B.A. Carvalho, F.P.N. Leite, M.A. Medeiros, L.F. Souza, C.S. Castro, F.G.C. Chassot, Sandro L. Bonatto, Population data of 17 Y-STR loci from Rio Grande do Sul state (South Brazil), Forensic Sci. Int. Genet. 4 (2009) e31–e33.

[10] N.N. Da Fré, R. Rodenbusch, A.Z. Gastaldo, E. Hanson, J. Ballantyne, C.S. Alho, Genetic data and de novo mutation rates in father-son pairs of 23 Y-STR loci in Southern Brazil population, Int. J. Legal Med. 129 (2015) 1221–1223.

[11] F.G.L. Chula, R. Rodenbusch, S. Schumacher, T. Grandi, C.T. Michelon, A.Z. Gastaldo, C. Costi, B. Carvalho, C.M.D. da Silva, 15 STR loci frequencies with mutation rates in the population from Rio Grande do Sul Southern Brazil, Forensic Sci. Int. Genet. 3 (2009) e35–e38.

[12] A.Z. Gastaldo, R. Rodenbusch, N.N. Da Fre, A.C. Mardini, C.S. Alho, Allele frequencies of 20 autosomal STR in the population from Rio Grande do Sul Southern Brazil, Forensic Sci. Int. Genet. 19 (2015) 15–17.

[13] C.B. Moyses, W.M. Tsutsumida, P.E. Raimann, C.H.A.S. Motta, T.L.S. Nogueira, O.C.L. Santos, B.B.P. Figueiredo, T.F. Mishima, I.M. Cândido, N.M.O. Godinho, L.S. Beltrami, R.K. Lopes, A.F. Guidolin, A. Mantovani, S.M. Santos, C.A. Souza, L. Gusmão, Population data of the 21 autosomal STRs included in the GlobalFiler® kits in population samples from five Brazilian regions, Forensic Sci. Int. Genet. 26 (2017) e28–e30.

[14] K.J. van der Gaag, R.H. de Leeuw, J. Hoogenboom, J. Patel, D.R. Storts, J.F. Laros, et al., Massively parallel sequencing of short tandem repeats-population data and mixture analysis results for the PowerSeq system, Forensic Sci. Int. Genet. 24 (2016) 86–96.

[15] S. Bailey, M.K. Scheible, C.L. Williams, D.S.B.S. Silva, C. Eichman, S.A. Faith, Secure and robust cloud computing for high-throughput forensic microsatellite sequence analysis and databasing, Forensic Sci. Int. Genet. 31 (2017) 40–47.

[16] W. Parson, D. Ballard, B. Budowle, J.M. Butler, K.B. Gettings, P. Gill, et al., Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements, Forensic Sci. Int. Genet. 22 (2016) 54–63.

[17] S.T. Kalinowski, M.L. Taper, T.C. Marshall, Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment, Mol. Ecol. 16 (2007) 1099–1106.

[18] R. Peakall, P.E. Smouse, GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research, Mol. Ecol. Note 6 (2006) 288–295.

[19] R. Peakall, P.E. Smouse, GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update, Bioinformatics 28 (2012) 2537–2539.

[20] D.H. Warshauer, J.D. Churchill, N. Novroski, J.L. King, B. Budowle, Novel Y-chromosome short tandem repeat variants detected through the use of massively parallel sequencing, Genomics Proteomics Bioinf. 13 (2015) 250–257.

[21] M. Geppert, J. Edelmann, R. Lessig, The Y-chromosomal STRs DYS481, DYS570 DYS576 and DYS643, Leg. Med. 11 (2009) S109–S110.