Characterization of 58 STRs and 94 SNPs with the ForenSeq™ DNA Signature Prep Kit in Mexican-Mestizos from the Monterrey City (Northeast, Mexico)

José Alonso Aguilar-Velázquez
Universidad de Guadalajara Centro Universitario de la Cienega
Miguel Ángel Duran-Salazar
Universidad de Guadalajara Centro Universitario de la Cienega
Miranda Fabiola Córdoba-Mercado
Fiscalia General de Justicia del Estado de Nuevo Leon
Carolina Elena Coronado-Avila
Fiscalia general de justicia del estado de Nuevo Leon
Orlando Salas-Salas
Fiscalia General de Justicia del Estado de Nuevo Leon
Gabriela Martinez-Cortés
Universidad de Guadalajara Centro Universitario de la Cienega
Ferrán Casals
Universitat de Barcelona
Francesc Calafell
Pompeu Fabra University: Universitat Pompeu Fabra
Benito Ramos-González
Fiscalia General de Justicia del Estado de Nuevo Leon
Héctor Rangel-Villalobos (hrangel13@hotmail.com)
Universidad de Guadalajara https://orcid.org/0000-0003-2641-8782

Research Article

Keywords: MiSeq FGx, ForenSeq DNA Signature Prep, MPS, Mexican population, Verogen.

Posted Date: November 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-996532/v1

Abstract

Background. STR allele frequency databases from populations are necessary to take full advantage of the increased power of discrimination offered by massive parallel sequencing (MPS) platforms.

Material and methods. For this reason, we sequenced 58 STRs (aSTRs, X-STRs, and Y-STRs) and 94 identity informative SNPs (iiSNPs) on 105 Mestizo (admixed) individuals from Monterrey City (Northeast, Mexico), with the Primer Set-A of the ForenSeq^{TM} DNA Signature Prep Kit.

Results. Most of the STR markers were in Hardy Weinberg equilibrium, with a few exceptions. We found 346 different length-based alleles for these 58 STRs; nevertheless, they became 528 alleles when the sequence was assessed. The combined power of discrimination from autosomal STRs (aSTRs) was –virtually– 100 % in both length and sequence-based alleles, while the power of exclusion were 99.9999999976065 and 99.9999999999494%, respectively. Haplotypes based on X-STRs and Y-STRs showed 100% of discriminatory capacity.

Conclusion. These results provide –for the first time– forensic genomic population data from Mexico necessary for interpretation in kinship and criminal analyses.

Introduction

Analysis of PCR products by massive parallel sequencing (MPS) overcomes many of the limitations attained with capillary electrophoresis (CE) analysis. MPS is getting widespread acceptance in forensic genetics because of its throughput generating information on a much higher number of markers and sequencing of short tandem repeats (STRs), which allows allele discrimination with the same length but different sequences that enhance its potential in human identification (HID). Moreover, MPS platforms allow detecting different markers simultaneously, such as STRs, SNPs, and insertion-deletions (InDels) [1–8]. For these reasons, several commercial PCR-MPS assays for human identification have been introduced and evaluated by forensic genetic labs [2, 3, 7–11]. Among the forensic MPS systems, highlights the ForenSeq DNA Signature Prep Kit (Verogen, San Diego, CA) that contains two separated primer sets: (1) DNA Primer Set-A (DPS-A) targets and amplifies 27 autosomal STRs (aSTRs), 7 X-chromosome STRs (X-STRs), 24 Y-chromosome STRs (Y-STRs) and 94 identity-informative SNPs (iiSNPs), and; (2) DNA Primer Set-B (DPS-B) targets all those markers included in DPS-A plus 56 ancestry-informative SNPs (aiSNPs) and 22 phenotype-informative SNPs (piSNPs) [12].

Regarding the inclusion of MPS in forensic genetics, the DNA Commission of the International Society of Forensic Genetics (ISFG) promotes obtaining allele frequency population databases to take full advantage of the increased power of discrimination offered by MPS generated data [13]. Unfortunately, this task has been scarcely done in Latin America, where –to our best knowledge– only one Peruvian population has been described [6]. In this paper, we report allele and haplotype frequencies and forensic parameters for the markers included in the ForenSeq™ DNA Signature Prep DPS-A in the Mexican Mestizo

(admixed) population from Monterrey City (Northeast, Mexico), which is the second main economic, cultural and political metropolis of this country.

Subjects And Methods

Population sample and DNA extraction method

DNA was obtained from peripheral blood in 105 unrelated (42 males and 63 females) residents of the Monterrey City in the Nuevo Leon state (Northeast, Mexico). For this purpose, the AutoMate Express DNA extraction system was used according to the supplier's instructions (Applied Biosystems). Next, DNA was quantified with the Quantifiler® trio DNA quantification kit in a 7500 Applied Biosystems Real-Time PCR system. Volunteers signed an informed consent form before the inclusion in the study, according to the Helsinki Declaration Ethical Guidelines. This work was approved by the Ethical Research Committee at the Institute of Criminalistics and Forensic Services of the Attorney General of Nuevo Leon state (Northeast, Mexico). The anonymity of the participants will be preserved all time.

Massive Parallel Sequencing (MPS) method

Libraries were generated using the DPS-A of the ForenSeq™ DNA Signature Prep Kit (Verogen®, San Diego, CA, USA). Library preparation involved PCR to amplify the DNA targets (STRs and iiSNPs) and to incorporate dual indexed adaptors [12]. The libraries were normalized and then pooled. A total 12-32 samples were pooled on each run. The library pool was diluted, denatured, and then added to the MiSeq FGx™ reagent Standard and Micro Kits for cluster generation on the flow cell according to manufacturer recommendations. Sequencing was conducted following the procedures outlined in the MiSeq FGx™ Instrument Reference Guide [14]. After each sequencing run was completed, a post-run wash was performed. We carried out five sequencing runs, two of them with 10 samples on microflow cell and three runs with 30 samples on Standard Flow cell, including positive and negative amplification controls in each run. Sequencing results were analyzed with the Universal Analysis Software (UAS) provided by the manufacturer and using its default parameters for variant calling and then retrieved and downloaded in an Excel sheet.

Data analysis

STR allele sequences were analyzed through the R script (lFator for autosomal STRs, YIFator for Y-STRs, and XIFator for X-STRs) previously reported by Casals et al. [3] and available in github (https://github.com/fcalafell/). Then, sequences reported herein were compared against a set of reference sequences built with previous sequencing results [3], and updated with our results. The R script provides the following results: (i) repeat sequence-based (RSB) and length-based (LB) allelic frequencies (aSTRs and Y-STRs); (ii) haplotype frequencies of both RSB and LB for Y-STRs; (iii) number of different haplotypes for X- and Y-STRs and; (iv) statistics of forensic-interest such as the power of discrimination and chance of exclusion for autosomal STRs. For practical purposes, we implemented the notation for RSB alleles proposed by Casals et al. [3], always taking into account the suggestions of the DNA

commission of the ISFG [13]. The notation consists of the repeat number as provided by UAS followed by a lowercase letter, which is different for each different sequence (i.e. allele 16a, 16b or 16c). For a more detailed explanation, please see the Material and methods section of [3]. The full list of RSB variants and their notation can be found as Supplementary Material for autosomal, X- and Y-STRs, respectively (SM1–SM3). Moreover, the following analyses were computed with the GenALEx complement of Excel [15]: (i) the Hardy-Weinberg equilibrium (HWE) of the aSTRs, iiSNPs, and X-STRs (only women, n= 62); (ii) allelic frequencies and statistics of forensic interest of the X-STRs and iiSNPs, and; (iii) expected and observed heterozygosity of the aSTRs and iiSNPs.

Results

Fifty-eight STRs and 94 iiSNPs were sequenced for 105 samples from Monterrey City, Mexico, which is in line with the guidelines for the publication of genetic population data generated by MPS published by Gusmão et al. [16], who recommended at least 50 full genotypes. For the three runs, quality metrics (Cluster density, Clusters passing filter, Phasing, and Pre-phasing) were within the boundaries defined by the manufacturer. All negative controls were blank and all positive controls gave full and expected profiles in all runs (2800 M supplied in ForenSeq™ DNA Signature Prep Kit).

Allele frequencies and forensic parameters for aSTRs

The forensic parameters and allelic frequencies for LB and RSB variation for the 27 autosomal STRs are reported as supplementary material (SM4-SM5). Furthermore, 105 different RSB genotypes were observed for the 27 aSTRs (SM6). A total of 252 different alleles were found when the allele calls were based only on the number of repeats. However, the number of alleles become 367 when the sequence variation was analyzed (46 % greater), involving 19 autosomal STRs (Fig. 1a). The largest increase of allele diversity was observed in D12S391 (13 vs. 34), D21S11 (12 vs. 28), and D2S1338 (10 vs. 26). Conversely, no increase in the allele number was observed in the following aSTRs: CSF1PO, D10S1248, D16S539, D20S482, D20S482, D22S1045, PentaD, TH01, and TPOX. As expected, observed heterozigosity of aSTRs with RSB variation increases importantly (SM4-SM5). The combined power of discrimination was –virtually– 100 % for both LB and RSB variation, whereas the combined power of exclusion for RSB and LB were 99.9999999999494 and 99.9999999976065 %, respectively. As expected, the combined power of exclusion was slightly higher in RSB than in LB alleles, which is the result of the increase in allele number for some STRs when the sequence variation is analyzed (SM4-SM5).

Allele frequencies and forensic parameters for X-STRs

For the seven X-STRs, haplotype frequencies, as well as allele frequencies for LB and RSB variation were reported as supplementary material SM7-SM9, respectively. We found 62 genotypes (females) and 43 haplotypes (males) that showed a discriminatory capacity of 100 % (Supplementary Table S7). When LB variation was analyzed 27 different alleles were reported (Supplementary Table S8), which increased 66 % (from 27 to 56) when RSB variation was taken into account (Supplementary Table S9). The increase in

RSB variation was observed in the following three X-STRs: DXS10074 (10 vs. 11), DXS10103 (6 vs. 10) and DXS10135 (21 vs. 41) (Fig. 1b).

Allele frequencies and forensic parameters for Y-STRs

For 24 Y-STRs, 39 complete haplotypes were sequenced, of which 39 different haplotypes were observed in both LB and RSB variation (SM10-SM11, respectively), both rendering a discriminatory capacity of 100%. Therefore, RSB variation did not imply increase in the number of haplotypes because LB variation was sufficient to differentiate all haplotypes. When the LB diversity was assessed, we obtained 67 different alleles (SM12). However, 105 different alleles were observed when the allelic RSB diversity was analyzed (SM13). RSB variation was detected in the following nine Y-STRs: DYF387S, DYS389II, DYS390, DYS448, DYS481, DYS505, DYS549, DYS570, and DYS635 (Figure 1c).

Allele frequencies and forensic parameters for iiSNPs

Forensic parameters, allele frequencies, and HWE agreement of 94 identity-informative SNPs were reported as supplementary material (SM14). The average of expected heterozygosity was 0.4427, close to the maximum possible value (0.5), which is similar to previous descriptions in European populations [3, 4]. For the 94 iiSNPs the combined power of discrimination was 99.99999999999999999626 (~1), while the power of exclusion was slightly lesser 99.995778%.

Discussion

We observed a significant deviation in HWE for both LB (D19S433, D21S11, D22S1045, and Penta E) and RSB (D19S433, D22S1045, D1S1656, D2S441, and Penta E) variation in aSTRs after Bonferroni correction, as well as in three iiSNPs (rs6444724, rs6955448, and rs1493232). These results may be caused by sampling errors, especially when it comes from a cosmopolitan population such as Monterrey City, the second-most-important metropolis in Mexico. Although a larger population sample size could exhibit more allele diversity, producing more accurate estimates, we were able to observe as far as 367 different aSTR alleles, which is similar or superior to previous studies with the DPS-A of the ForenSeq™ DNA Signature Prep [3–5].

This study follows the ISFG recommendation regarding the generation of STR allele population databases [13], which are required for biostatistical interpretation in kinship analysis and criminal cases where ForenSeq^{TM} DNA Signature Prep kit is used. For instance, in a recent complex paternity case reported in Mexico, the suitability of a forensic genomic platform was invoked [17]. However, the largest social impact of genomic platforms is expected in criminal investigation, given the 40,180 missing persons and 37,000 unidentified bodies in Mexico [18]. Consequently, the genomic population database reported herein will be helpful for this task in Mexico, but also Latin American populations using the ForenSeq^{TM} DNA Signature Prep kit without a proper population database.

To our best knowledge, this is the first study where forensic MPS-based data was reported in a Mexican population and is the second report from Latin America. Although the previous Latin American study in

Peruvian population reported forensic parameters of aSTRs and iiSNPs with the same forensic MPS platform, they did not report X-STR and Y-STRs population data [6]. Thus, for these STRs the reported Mexican population database is the first one available in Latin America for interpretation purposes.

The number of different alleles found here (528 different STR alleles) was higher than those reported in Spanish [3], French [4], Asian [5], and Peruvian [6] populations. Also, our RSB results improve importantly the combined PD and CE compared with those previously obtained in Monterrey City but based on 23 aSTRs genotyped by CE [19]. As expected, our sequencing results for the 24 Y-STRs also improve the genetic informativity from Monterrey City regarding that previously reported with 23 Y-STRs based on CE, where a discriminatory capacity of 93.75 % was described [20]. Although our sample size is smaller, with sequencing results a higher discriminatory capacity is expected due to RSB variation.

Conclusion

In brief, we report allele frequencies and forensic statistical parameters of STRs and iiSNPs based on the analysis of ForenSeq™ DNA Signature Prep kit in the Monterrey City population (Northeast, Mexico). Our sequencing results describe the internal repeat variation of different STRs, such as aSTRs, X-STRs, and Y-STRs. The reported genomic population database provides valuable information for statistical interpretation in forensic casework in Mexico and probably some other Latin American populations.

Declarations

Competing Interests and Funding

Authors declare no conflicts of interest.

Compliance with Ethical Standards

All individuals signed a written informed consent agreeing to the ethical guidelines of the Helsinki Declaration. This study was approved by the Local Ethical Research Committee. The anonymity of the participants will be preserved all time.

References

1. Friis SL, Buchard A, Rockenbauer E, Borsting C, Morling N (2016) Introduction of the Python script STRinNGS for analysis of STR regions in FASTQ or BAM files and expansion of the Danish STR sequence database to 11 STRs. Forensic Sci Int Genet 21:68-75. doi:https://doi.org/10.1016/j.fsigen.2015.12.006
2. Novroski NM, King JL, Churchill JD, Seah LH, Budowle B (2016) Characterization of genetic sequence variation of 58 STR loci in four major population groups. Forensic Sci Int Genet 25:214-226. doi:https://doi.org/10.1016/j.fsigen.2016.09.007

3. Casals F, Anglada R, Bonet N, Rasal R, van-der-Gaag KJ, Hoogenboom J, Solé-Morata N, Comas D, Calafell F (2017) Length and repeat-sequence variation in 58 STRs and 94 SNPs nformed consent agreeing to the ethical guidelines of the Helsinki Declaration. This stuin two Spanish populations. Forensic Science International: Genetics 30:66-70. doi:10.1016/j.fsigen.2017.06.006
4. Delest A, Godfrin D, Chantrel Y, Ulus A, Vannier J, Faivre M, Hollard C, Laurent FX (2020) Sequenced-based French population data from 169 unrelated individuals with Verogen´s ForenSeq DNA signature prep kit. Forensic Science International: Genetics 47. doi:10.1016/j.fsigen.2020.102304
5. Peng D, Zhang Y, Ren H, Li H, Li R, Shen X, Wang N, Huang E, Wu R, Sun H (2020) Identification of sequence polymorphisms at 58 STRs and 94 iiSNPs in a Tibetan population using massively parallel sequencing. Sci Rep 10. doi:https://doi.org/10.1038/s41598-020-69137-1
6. Guevara EK, Palo JU, King JL, Buś MM, Guillén S, Budowle B, Sajantila A (2021) Autosomal STR and SNP characterization of populations from the Northeastern Peruvian Andes with the ForenSeqTM; DNA Signature Prep Kit. Forensic Science International: Genetics 52. doi:10.1016/j.fsigen.2021.102487
7. Phillips C, Devesse L, Ballard D, van-Weert L, de-la-Puente M, Melis S, Álvarez-Iglesias V, Freire-Aradas A, Oldroyd N, Holt C, Syndercombe-Court D, Carracedo Á, Lareu MV (2018) Global patterns of STR sequence variation: Sequencing the CEPH human genome diversity panel for 58 forensic STRs using the Illumina ForenSeq DNA Signature Prep Kit. Electrophoresis 39 (21):2708-2724. doi:10.1002/elps.201800117
8. Barrio PA, Martin P, Alonso A, Muller P, Bodner M, Berger B, Parson W, Budowle B, Consortium D (2019) Massively parallel sequence data of 31 autosomal STR loci from 496 Spanish individuals revealed concordance with CE-STR technology and enhanced discrimination power. Forensic Sci Int Genet 42:49-55
9. Just RS, Moreno LI, Smerick JB, Irwin JA (2017) Performance and concordance of the ForenSeq™ system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens. Forensic Sci Int Genet 28:1-9. doi:10.1016/j.fsigen.2017.01.001
10. Xu M, Du Q, Ma G, Chen Z, Liu Q, Fu L, Cong B, Li S (2019) Utility of ForenSeq™ DNA Signature Prep Kit in the research of pairwise 2nd-degree kinship identification. International Journal of Legal Medicine. doi:10.1007/s00414-019-02003-6
11. Wu J, Li JL, Wang ML, Li JP, Zhao ZC, Wang Q, Yang SD, Xiong X, Yang JL, Deng YJ (2019) Evaluation of the MiSeq FGx system for use in forensic casework. Int J Legal Med 133 (3):689-697
12. Verogene (2020) ForenSeq™ DNA Signature Prep Reference Guide. D2018005 Rev. C. San Diego, California 92121 U.S.A.
13. Parson W, Ballard D, Budowle B, Butler JM, Gettings KB, Gill P, Gusmão L, Hares DR, Irwin JA, King JL, Knijff Pd, Morling N, Prinz M, Schneider PM, Neste CV, Willuweit S, Phillips C (2016) Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. Forensic Science International: Genetics 22:54-63. doi:https://doi.org/10.1016/j.fsigen.2016.01.009

14. Verogen (2018) MiSeq FGx™ Instrument Reference Guide. San Diego, CA 92121, U.S.A.
15. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537-2539
16. Gusmão L, Butler JM, Linacre A, Parson W, Roewer L, Schneider PM, Carracedo A (2017) Revised guidelines for the publication of genetic population data. Forensic Sci Int Genet 30:160-163. doi:https://doi.org/10.1016/j.fsigen.2017.06.007
17. González-Herrera LJ, García-Aceves ME, Domínguez-Cruz MD, López-González PN, Sosa-Escalante JE, Rangel-Villalobos H (2020) A four-step mutation at D22S1045 in one complex paternity case when the brother of the alleged father hypothesis is evaluated. Int J Legal Med 134: 1653 doi: 10.1007/s00414-020-02335-8.
18. Guevara-Bermúdez JA, Chávez-Vargas LG (2018) La impunidad en el contexto de la desaparición forzada en México. Eunomía. Rev en Cult de la Legalidad 14: 162–174. doi: https://doi.org/10.20318/eunomia.2018.4161
19. Ramos-González B, Aguilar-Velázquez JA, Chávez-Briones MdL, Delgado-Chavarría JR, Alfaro-Lopez E, Rangel-Villalobos H (2016) Population data of 24 STRs in Mexican-Mestizo population from Monterrey, Nuevo Leon (Northeast, Mexico) based on Powerplex Fusion and GlobalFiler kits. Forensic Science International: Genetics 21:e15-e17. doi:10.1016/j.fsigen.2015.12.004
20. Ramos-González B, Aguilar-Velázquez JA, Chávez-Briones MdL, del Rocío Escareño-Hernández M, Alfaro-Lopez E, Rangel-Villalobos H (2017) Genetic population data of three Y-STR genetic systems in Mexican-Mestizos from Monterrey, Nuevo León (Northeast, Mexico). Forensic Science International: Genetics 29:e21-e22. doi:10.1016/j.fsigen.2017.04.016

Figures

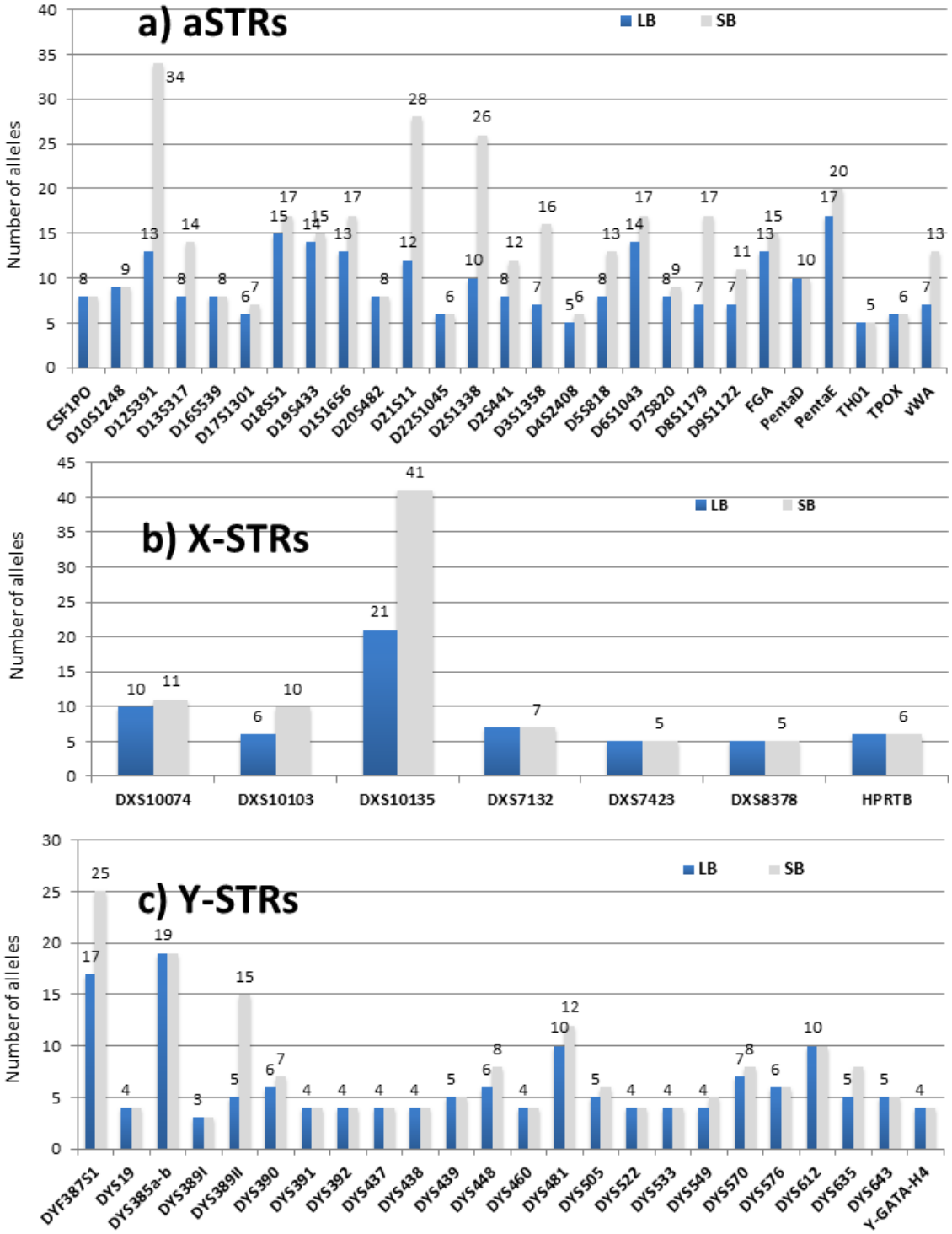

Fig. 1. Differences in the number of length-based (LB) and sequence-based (SB) alleles observed in the studied Mexican population sample for aSTRs (a), X-STRs (b), and Y-STRs (c).

Figure 1

Differences in the number of length-based (LB) and sequence-based (SB) alleles observed in the studied Mexican population sample for aSTRs (a), X-STRs (b), and Y-STRs (c).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SM1.NomenclatureRSBaSTR.xlsx
- SM10.HaplotypeFreqSBYSTRs.xlsx
- SM11.HaplotypeFreqLBYSTRs.xlsx
- SM12.AlleleFreqLBYSTRs.xlsx
- SM13.AlleleFreqSBYSTRs.xlsx
- SM14.AllelicFreqSNP.xlsx
- SM2.NomenclatureRSBXSTR.xlsx
- SM3.NomenclatureRSBYSTR.xlsx
- SM4.ForensicparametersandallelefreqLBaSTR.xlsx
- SM5.ForensicparametersandallelefreqRSBaSTR.xlsx
- SM6.GenotypesSRBaSTR.xlsx
- SM7.HaplotypeFrequencyXSTR.xlsx
- SM8.ForensicparametersandallelefreqLBXSTR.xlsx
- SM9.AlleleFreqRSBXSTRs.xlsx