



ASOCIACIÓN NACIONAL
DE
MÉDICOS FORENSES

Spanish Journal of Legal Medicine

Revista Española de Medicina Legal

www.elsevier.es/mlegal



ORIGINAL ARTICLE

Allele frequencies of 21 autosomal STR markers in a mixed race Peruvian population applied to forensic practice[☆]



Edgardo Delgado^{a,b,*}, Carlos D. Neyra^{c,d}

^a Laboratorio de Biología Molecular y de Genética (LABIMOG), Instituto de Medicina Legal y Ciencias Forenses del Ministerio Público, Lima, Peru

^b Universidad Nacional Mayor de San Marcos, Posgrado de Ciencias Biológicas, Lima, Peru

^c Gerencia de Calidad e Innovación, Registro Nacional de Identificación y Estado Civil, Lima, Peru

^d Facultad de Medicina Humana, Universidad Alas Peruanas, Lima, Peru

Received 22 June 2018; accepted 11 September 2018

Available online 4 September 2019

KEYWORDS

Genetic marker;
Forensic genetics;
Allelic frequencies;
Genotype;
Peruvian population

Abstract

Introduction: Short Tandem Repeat (STR) markers are used in forensic practice due to their great discriminative power and help to identify individuals genetically. At least 20 STRs are currently used in Peru.

Objective: The objective of the present study is to determine the allelic frequencies of 21 genetic markers of type STR (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S391 and D2S1338) in a mixed race Peruvian population.

Materials and methods: The allelic frequencies were obtained from 300 unrelated individuals from different provinces of Peru. A statistical analysis was performed on the data obtained.

Results and conclusions: All loci, except for D10S1248 and D2S1338, were in Hardy-Weinberg equilibrium. The most frequent allele of the D3S1358 marker is 15, with 56.33% in the Peruvian population and with 34.65% in the Hispanic population, while the alleles less frequent in the Peruvian population were 11 and 19, and in the Hispanic one, it was the 9 allele, which the Peruvian population does not have. The vWA marker of the Peruvian population does not have the 11, 12, and 17.3 alleles seen in the Hispanic population. The PD is greater than 0.9999999 and the PE is 0.9999997. When comparing the allelic frequencies of the Peruvian mixed-race population and the Hispanic frequencies of the manual of Globalfiler by Life Technologies For Forensic or Paternity 2013, differences are observed in the allelic frequencies of the markers.

© 2018 Asociación Nacional de Médicos Forenses. Published by Elsevier España, S.L.U. All rights reserved.

DOI of original article: <https://doi.org/10.1016/j.reml.2018.09.002>

[☆] Please cite this article as: Delgado E, Neyra CD. Frecuencias alélicas de 21 marcadores STR autosómicos en una población mestiza peruana aplicado a la práctica forense. Rev Esp Med Legal. 2019;45:92–97.

* Corresponding author.

E-mail address: edelrrams@hotmail.com (E. Delgado).

PALABRAS CLAVE

Marcador genético;
Genética forense;
Frecuencias alélicas;
Genotipo;
Población peruana

Frecuencias alélicas de 21 marcadores STR autosómicos en una población mestiza peruana aplicado a la práctica forense

Resumen

Introducción: En la práctica forense los marcadores del tipo STR (Short Tandem Repeat) son utilizados por su gran poder discriminativo y por ayudar a identificar individuos genéticamente. En Perú actualmente se utilizan como mínimo 20 STR.

Objetivo: El objetivo del presente estudio es determinar las frecuencias alélicas de 21 marcadores genéticos de tipo STR (D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S391 y D2S1338) en una población mestiza peruana.

Materiales y métodos: Las frecuencias alélicas fueron obtenidas a partir de 300 individuos no emparentados provenientes de diferentes provincias del Perú. Con los datos obtenidos se realizó un análisis estadístico.

Resultados y conclusiones: Todos los loci, excepto el D10S1248 y el D2S1338, estuvieron en equilibrio de Hardy-Weinberg. El alelo más frecuente del marcador D3S1358 es el 15, con un 56,33% en la población peruana y un 34,65% en la población hispana, mientras que los alelos menos frecuentes en la población peruana son el 11 y el 19, y en la hispana es el alelo 9, que no posee la población peruana. El marcador vWA de la población peruana no posee los alelos 11, 12 y 17.3 que posee la población hispana. El PD es mayor de 0,9999999 y el PE es de 0,9999997. Al compararse las frecuencias alélicas de la población mestiza peruana y las frecuencias hispanas del manual de Globalfiler by Life Technologies For Forensic or Paternity 2013 se observan diferencias en las frecuencias alélicas de los marcadores.

© 2018 Asociación Nacional de Médicos Forenses. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Introduction

The search for reliable methods of human identification has always been greatly needed by society to resolve both civil and criminal cases.¹ Over the years, different techniques have been used to characterise and individualise people, including fingerprint, anthropological, and blood group analysis,² as well as molecular analysis of human leukocyte antigen (HLA) polymorphisms. With the discovery of the structure of deoxyribonucleic acid (DNA) in 1953 and its polymorphic regions in 1980, DNA analyses have become an indispensable tool for forensic genetics.

Through the analysis of DNA samples, obtained from traces found at the crime scene – of blood, saliva, semen, hair, nails, tissues, bones, among others – and organising the profiles of these samples in a database, the possibility of a successful criminal investigation is made considerably higher.^{3–5}

In human identification, the use of DNA polymorphism analysis is based on the biological individualisation of each human being, since their genetic profile is exclusive and immutable and registered in all the cells of their organism, providing all cells equality and invariability throughout the individual's life (except for cases involving monozygotic twins). An individual's genetic profile is made up of several markers that are inherited from their parents. These markers correspond to DNA sequences that vary between individuals, called hypervariable or polymorphic regions, and are distributed in the autosomal and sex chromosomes.

Thus, variability among individuals is determined by repeating sequences; depending on the size of the repeating unit, these sequences are termed satellite DNA (repeating unit of between 100 and 10,000 nucleotides), minisatellite DNA (repeating unit of between 10 and 100 nucleotides) and microsatellite DNA (repeating unit of between 2 and 7 nucleotides).^{6–9}

The efforts of the forensic community have resulted in the generation of 41 population data sets that provide a solid basis for the frequency of estimated profiles in several sampled populations, such as African Americans, Caucasians, Hispanics, Asians, Far Eastern populations, and Native Americans.¹⁰ There are significant differences in each population group in certain genetic markers and for certain alleles; hence the need for population frequency bases for each population group or region.

Since January 1, 2017 the FBI's working group has required laboratories to include 7 additional markers (D1S1656, D2S441, D2S1338, D10S1248, D12S391, D19S433 and D22S1045) to the existing 13 (D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, CSF1PO, FGA, TH01, TPOX, vWA), which make a total of 20 autosomal genetic markers in addition to the sex-typing marker amelogenin, which should be in the bases of each country.¹¹ The European Community also recommends standard markers (FGA, TH01, VWA, D1S1656, D2S441, D3S1358, D8S1179, D10S1248, D12S391, D18S51, D21S11, D22S1045) and additional D2S1338, D16S539, D19S433, SE33 markers and amelogenin. The aim is to reduce the accidental coincidence probability, because the databases are increasing considerably in

size and exchange between countries is expanding, which is leading to the identification of missing or accused persons.

Many research studies provide databases of allele frequencies worldwide, which show that allele frequencies differ among populations or population groups; it has been observed that some alleles are more frequent in populations, while there are alleles that are absent in certain populations. For this reason, the aim is to characterise 21 autosomal STR markers in a mixed-race Peruvian population, which will serve to obtain a genetic frequencies base as an aid for forensic practice. We should bear in mind that the Peruvian population, as in other Latin American countries, is considered mixed-race, a product of the genomic mixture between the European conquerors and the native population.¹²

Materials and methods

The population in this descriptive, cross-sectional study comprised people involved in judicial proceedings in 2013. The population consisted of 1000 individuals, and 300 unrelated individuals were randomly selected. The exclusion criteria were having undergone a bone marrow transplant and/or blood transfusion in the past 6 months, which could generate genetic profiles of the donor. The inclusion criteria were unrelated individuals over the age of 18 who had signed their informed consent. For each sample, signed informed consent and chain of custody were obtained; the identity of the participant was dissociated from the sample collected, to respect confidentiality.

DNA sampling and extraction

A blood sample was taken from the index finger by finger prick with a lancet. The samples were taken on separate FTA (Wathman) cards (including barcode). The FTA cards contain a chemically treated matrix that lyses a great variety of tissues (for example, blood and saliva, among others). After lysing the cells, the released DNA attaches to the card where the matrix protects the nucleic acids from harmful agents that could cause degradation, thus reducing the degradation of the cells.¹³

Obtaining the genetic profile

The genetic profile was determined by direct amplification of the genetic markers of the non-coding DNA with the GlobalFiler™ Express kit, following the indications of the manufacturer (Life Technologies), validated beforehand.

The amplification of STR markers was performed using the multiplex amplification technique. PCR amplification products were detected by the fluorescence detection in capillary electrophoresis technique in the Applied Biosystems™ 3500genetic analyzer.

Following the capillary electrophoresis, the data were imported into GeneMapper® ID-X v1.1 genetic typing software, which was used to obtain and edit the genetic profile. The negative control and the target were checked for peaks in order to discount contamination. The genetic profiles had to be corroborated with a standard reference marker. The

sizes of the observed peaks (above 50 RFU), quality, concentration, presence or absence of interferences and possible null alleles were assessed.

Finally, after successfully passing the quality control tests, and only then, these profiles were exported from the GeneMapper® ID-X v1.1 software in the form of numerical codes (see Table 1 of the supplementary material).

Statistical analysis

In order to characterise the 21 autosomal STR markers of the Peruvian population under study, the following forensic statistical parameters were calculated: power of discrimination (PD), power of combined discrimination, polymorphic information content (PIC), power of exclusion (PE), power of combined exclusion, observed heterozygosity (Hobs) and expected heterozygosity (Hesp), Hardy-Weinberg equilibrium (*p*-value), and probability of coincidence (PM) (Table 1 of the supplementary material). All these parameters helped us to determine the usefulness of these markers in resolving forensic cases using the programmes PowerState.v1.2,¹⁴ Arlequin®.¹⁵ The minimum allele frequency (MAF) was also calculated for the alleles that appear less often.

Comparison of allele frequencies between the Peruvian population under study and the Hispanic population

Allele frequencies were compared between the Peruvian and Hispanic populations to detect significant differences between both populations for each marker analysed. For the purposes of this study, the Peruvian population is understood as that of the three defined regions of Peru (coast, highlands and jungle) and characterised as multiracial, multilingual and pluricultural. The Hispanic population includes countries where there is notable influence of the Spanish language and culture.

Results

A total of 300 individuals were included (113 women and 187 men, over 18 years old). All were mixed race.

Allele frequencies

The allele frequencies of a genetic marker indicate the number of times an allele is observed in the population relative to the total number of alleles of that marker and is represented as a fraction or percentage. Table 1 of the supplementary material shows the allele frequencies obtained from the 21 autosomal STR markers in the Peruvian population under study. The most polymorphic locus (SE33) has 31 alleles, while the least polymorphic locus (TPOX) has 6 alleles. It was demonstrated in the study that the 21 STRs studied are good genetic markers to characterise the Peruvian population as presenting high polymorphism. The three most informative STRs due to their PIC were SE33 (.9095), followed by FGA (.8460), and D18S51 (.8404), while the marker with the lowest index was D22S1045 (.5071). The

markers studied had PIC values higher than .5, which meets the condition of being very informative.¹⁶

The PD is characterised by the probability that two unrelated and randomly selected individuals can be genetically differentiated by the analysis of a marker or a set of markers. Analyzing the results set out in Table 1 of the supplementary material, it can be seen that the PD index for marker SE33 is .9841, which is the highest, followed by FGA at .9642 and D18S51 at .9627; while the marker with the lowest PD index is D2S441 at .7635. The average PD of the markers is .8836.

The PE parameter allows us to establish the proportion of individuals falsely involved in an expert appraisal. The most informative marker is SE33 (.7818), followed by D18S51 (.6559) and D1S165 (.6494), while the least informative markers are D22S1045 (.2421), D2S441 (.2440) and D3S1358 (.2752).

Analysis of the Hobs and Hesp indices showed that for all the markers the Hesp indices were very close to those observed. The average Hobs value for the total population is .7237 and the Hesp value is .7395; the highest Hobs index was obtained in the marker SE33 (.8933), followed by FGA (.8533) and D18S51 (.8300); in the same way, the same markers with the highest Hesp can be appreciated, which are SE33 (.9169), followed by FGA (.8626) and D18S51 (.8581). As for the lowest indices obtained, in the Hobs index, these are the markers D2S441 (.5585) and D22S1045 (.5567), and in the Hesp index, the markers D2S441 (.5559) and TPOX (.5856). According to Crow and Kimura,¹⁶ heterozygosity is the probability that two alleles of the same locus taken randomly from the population will be different. Hesp is also known as gene diversity.

The Hardy-Weinberg equilibrium test (p -value greater than .05) indicated a deviation of equilibrium in two genetic markers: D10S1248 ($p = .0467$) and D2S1338 ($p = .0074$). However, this deviation was not significant after applying the Bonferroni correction.¹⁷ Therefore, it was established that the 21 markers are in equilibrium.

Finally, the marker with the highest coincidence probability index is D22S1045 (.2416), followed by D2S441 (.2365) and D3S1358 (.2253), while the marker with the lowest index is SE33 (.0159), followed by FGA (.0358) and D18S51 (.0373).

The MAF for the alleles that do not appear regularly is .0083, which was obtained by dividing 5 by $2n$ (" n " is the number of samples).

Comparison of allele frequencies among the Peruvian population under study and the Hispanic population

Comparison between the markers and alleles inherent to the Peruvian and Hispanic populations reveals that there are alleles inherent to the Peruvian population and alleles that are not inherent to the Peruvian population (Table 2 of the supplementary material); likewise, it is noted that some alleles are more and less frequent only in the Peruvian population (Table 3 of the supplementary material). The differences found in some markers of forensic importance are described below.

Alleles 11, 12, and 17.3 are not present in the vWA marker for the Peruvian population, but are reported in the

Hispanic population. The most frequent alleles are 16 and 17 in both populations; while the least frequent allele in the Peruvian population is 13, and the most frequent in the Hispanic population are alleles 11, 13 and 17.3.

In marker D3S1358, allele 11 is only found in the Peruvian population; while alleles 9, 12 and 13 are only found in the Hispanic population. The most frequent allele in both populations is 15, while the least frequent alleles in the Peruvian population are 11 and 19, and allele 9 in the Hispanic population.

Marker D10S1248 of the Peruvian population under study has allele 18, which the Hispanic population does not have, while the Peruvian population does not have alleles 8 and 9 that the Hispanic population does have. The most frequent allele in both populations is allele 14; while the least frequent alleles in the Peruvian population are alleles 10 and 18, and alleles 8, 9 and 10 in the Hispanic population.

Marker SE33 of the Peruvian population under study provides alleles 17.2, 19.2, 23 and 35.2, which the Hispanic population does not have, while the Hispanic population does not have alleles 12, 12.2, 13, 14.2, 15.2, 16.2, 16.3, 20.2, 23.2, 26, 33, 34, 34.2, or 37. The most frequent allele is allele 17 in the Peruvian population, and allele 18 in the Hispanic population. On the other hand, the least frequent alleles are 11.2, 13.2, 23 and 24 in the Peruvian population, and alleles 11.2, 13.2, 24, 31 in the Hispanic population.

Marker D2S1338 of the Peruvian population under study does not have alleles 26, 27 and 28 that the Hispanic population does have, the most frequent allele in the Peruvian population is 19, and in the Hispanic population it is 17; while the least frequent allele in the Peruvian population is 16 and in the Hispanic population, alleles 27 and 28.

Discussion

Autosomal STR markers are increasingly applied forensically, as genetic profiles in databases increase considerably in size. At present, laboratories working according to the CODIS system use a minimum of 20 genetic markers; the European Community also uses 15 of these markers as standard and adds SE33. The aim is to reduce the probability of accidental matches in the identification of missing or accused persons. The exchange of genetic profiles is expected to expand in future between countries to genetically identify individuals who, due to globalisation, are migrating rapidly in order to commit crimes or disappear.

A study was carried out in 2009 on the population of San Andrés de Colombia¹⁸ using 15 autosomal STR loci, and it was determined that the STR markers contributing most in identification tests, by their PIC, PD, PE and number of effective alleles, were D2S1338, D18S51 and FGA. As in the Colombian population, the Peruvian population studied has the same markers contributing most for identification tests (D2S1338, D18S51, FGA) in addition to markers D19S433 and SE33. In 2009¹⁹ a study was carried out in the Argentine population with 15 autosomal STR markers, and the markers that exhibited the greatest heterozygosity were markers FGA and D2S1338, while TPOX had the lowest value.

In the Peruvian population studied, the same behaviour was found in the same markers. In 2011 Manríquez et al.²⁰ conducted a study in a Hispanic-American population of Valparaíso, Chile, in which they used 6 European markers, and found that marker SE33 was the most polymorphic marker, resulting as the most informative. The behaviour regarding the most polymorphic marker SE33 is similar between the Peruvian population and the Chilean population studied. Rangel et al.²¹ in 2013 carried out a study in the Amerindian population of northern and eastern Mexico, in which they used 15 STR markers, and detected D18S51, D19S433, FGA and D21S11 as the most informative markers; all the markers resulted in Hardy-Weinberg equilibrium, after applying Bonferroni correction to markers D13S317 and FGA ($p < .0033$). The behaviour of the Peruvian population under study is similar to that of the Mexican population in terms of the most informative markers; furthermore, markers SE33, D16S539, D8S1179, D1S1656, D12S391 would have to be added.

Regarding the variability of the markers of the Peruvian population analysed, it was observed that markers SE33, D21S11, FGA, D12S391, D1S1656 and D18S51 presented greater variability with respect to the number of alleles; it was detected that marker SE33 presents 32 different alleles, FGA and D21S11 present 15 alleles, and markers D18S51, D12S391 and D1S1656 present 14 different alleles each. On the other hand, those with the lowest allele variability are TH01 with 6 alleles, and markers D3S1358, D13S317, TPOX and D16S539, each with 7 different alleles.

Making a comparison between the frequencies found in the Hispanic population and those found through analysis of the Peruvian population under study, the average difference in frequencies is 2.9357% (on a scale from 1 to 100, the difference is low), representing a 97.0643% similarity, so the difference would be given by some alleles among the Peruvian population that were not reported in the Hispanic population, and vice versa.

Financing

Molecular Biology and Genetics Laboratory (LABIMOG) of the *Instituto de Medicina Legal y Ciencias Forenses del Ministerio Público del Perú* (Institute of Legal Medicine and Forensic Sciences of the Public Ministry of Peru).

Conflict of interests

The authors have no conflict of interests to declare.

Acknowledgements

To the management of the Molecular Biology and Genetics Laboratory (LABIMOG) of the *Instituto de Medicina Legal y Ciencias Forenses del Ministerio Público del Perú* (Institute of Legal Medicine and Forensic Sciences of the Public Ministry of Peru)

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.remle.2018.09.001](https://doi.org/10.1016/j.remle.2018.09.001).

References

1. França GV. *Medicina legal*. Rio de Janeiro: Guanabara-Koogan; 2001. p. 32–63.
2. Hirszfeld I, Hirszfeld H. Serological differences between the blood of different races. *Lancet*. 1919;194:675–9.
3. Oz C, Levi JA, Novoselski Y, Volkov N, Motro U. Forensic identification of a rapist using unusual evidence. *J Forensic Sci*. 1999;44:860–2.
4. Mora Sánchez JM. Delitos contra la libertad sexual y análisis de ADN. *Rev Latinoam Der Méd Leg*. 2001;6:7–13.
5. Sweet D, Hildebrand D. Saliva from cheese bite yields DNA profile of burglar: a case report. *Int J Legal Med*. 1999;112:201–3.
6. Butler JM. *Forensic DNA typing. Biology, technology and genetics of STR markers*. 2nd ed. New York: Elsevier Academic Press; 2005.
7. Butler JM, Hill CR. Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. *Forensic Sci Rev*. 2012;24:15–26.
8. Hamada H, Petrino MG, Kakunaga TA. Novel repeated element with DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc Natl Acad Sci USA*. 1982;79:6465–9.
9. Tuatz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res*. 1984;12:4127–38.
10. Budowle B, Shea B, Niezgoda S, Chakraborty C. CODIS STR loci data from 41 sample population. *J Forens Sci*. 2001;46:29–65.
11. Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizona R, Buckleton JS, et al. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. *Forensic Sci Int Genet*. 2016;25:175–81.
12. Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, et al. Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet*. 2008;4:1–9.
13. GE Healthcare Life Sciences. Whatman FTA Brochure; Your forensic samples, our experience. 2011. Available from: <https://www.thermofisher.co.nz/Uploads/file/Supplier-Partners/GE-Whatman-FTA.pdf> [accessed 9.11.15].
14. Tereba A. Tools for analysis of population statistics. *Profiles DNA*. 1999;2:14–6.
15. Schneider S, Roesli D, Excoffier L. Arlequin ver. 2.000: a software for population genetics data analysis. Switzerland: Genetic and Biometry Laboratory, University of Geneva; 2000.
16. Crow JF, Kimura M. *An introduction to population genetics theory*. New York, Evanston and London: Harper and Row Publishers; 1970. p. 83–98.
17. Bonferroni CE. *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze. 1936;8:3–62.
18. Lamprea N [Tesis de maestría] Caracterización genética de la población humana de San Andrés y providencia a partir de los marcadores microsatélites (STRs) empleados por el combined dna index system (CODIS). Bogotá: Universidad Nacional de Colombia; 2009.

19. Abovich M, Arellano A, Szocs A, Alcázar D, Cabeller S, Rodriguez MB. Allele frequencies of 15 STRs loci in an Argentine population simple. *Forensic Sci Int Genet Suppl Ser.* 2009;2:369–70.
20. Manríquez J, Rojas S, Yáñez MO, Molina G. Evaluation of PowerPlex ESI 17 amplification kit in an admixed Hispano-Amerindian population simple of Valparaíso Chile. *Forensic Sci Int Genet Suppl Ser.* 2011;3:113–4.
21. Rangel-Villalobos H, Martinez VM, Salazar J, Martínez G, Muñoz JF, Galaviz C, et al. Forensic parameters for 15 STRs in eight Amerindian populations from the north and west of Mexico. *Forensic Sci Int Genet.* 2013;7:62–5.