Pattern at the D16S539, PENTA D, PENTA E and TPOX loci by forensic analysis in Peru Southern region

Cecilia Manrique-Sam[1,2,3], Ernesto Aragón-Vela[3], Johany Sánchez-Guillén[1], Alice Zegarra-Adanaque[1], Sandra Manrique-Sam[1], Luis Carlos Lizarraga Vargas[4], María Teresa Meza-Aragon[5], Mary Angela Aragon-Vela[1] , Edgard Patrick Zeballos Gonzales[3], Jani Pacheco-Aranibar[4], Julio Cesar Bernabé-Ortiz[1,3]

[1].Universidad Católica de Santa María. Arequipa-Peru
[2].Facultad de Medicina Universidad Católica de Santa María. Arequipa-Peru
3.Universidad Nacional de San Agustin. Arequipa-Peru
[4].Facultad de Biologia Universidad Nacional de San Agustín. Arequipa-Peru
[5].Facultad de Medicina Universidad Nacional de San Agustín. Arequipa-Peru
[6].Instituto de Biotecnologia del ADN Uchumayo. Arequipa-Peru
Corresponding author: Email: jbernabe@unsa.edu.pe

Abstract

The study of genetic variability through microsatellite markers helps to characterize a population molecularly. Moreover, the genetic parameters obtained are useful in forensic statistical calculations, population genetics, and person identification. The present study aimed to identify the alleles and determine the allelic frequencies of the STR markers D16S539, PENTA D, PENTA E, TPOX present in the Southern Region of the Peru human population. Alleles were identified, allele frequencies of STRs were found, and it was determined if these markers were in Hardy-Weinberg genetic equilibrium. The study analyzed 100 samples from individuals who attended the Molecular Biology laboratory of the Universidad Nacional de San Agustín. DNA extraction was performed and amplified by PCR technique, the amplification products were visualized in polyacrylamide gels, and the computer analysis was done in the GENEPOP program. In the human population of the Southern Region of Peru, 4 alleles were identified for the Tpox marker, 5 alleles for the D16S539 STR, 8 alleles for the Penta D marker and 10 alleles were identified for the PENTA E marker; the STR with the highest discriminatory power was the Penta E STR. STRs markers TPOX, D16S539, PENTA E and Penta D are in Hardy Weinberg equilibrium in the human population of the Southern Region of Peru with a $P<0.05$ for all markers.

Keywords: Microsatellites, allele frequencies, Hardy Weinberg Equilibrium

1. Introduction

In recent years, the human genome study has presented great advances. For example, the use of molecular markers allows to determine genomic variability between individuals of the same species (1). The most commonly used molecular markers in the study of genomic variation in humans are Short Tandem Repeat or STRs due to their high degree of polymorphism. In population genetics, these markers allow us to know the degree of interbreeding of a population, to know the relative distances between different cultures, to determine the ancestral origins of a population and are widely used in the forensic area, in paternity tests and human identification (2).

The human genome is formed by all the DNA found in the nucleus and mitochondria of a cell, and the nuclear DNA is a linear macromolecule formed by two antiparallel chains joined together by hydrogen bridges. Each chain is made up of units called nucleotides (3). Each nucleotide is composed of a nitrogenous base, a pentose and a phosphate group, and the nitrogenous bases are divided into purine bases such as Guanine (G) and Adenine (A) and pyrimidine bases such as Cytosine (C) and Thymine (T) (4)(5).

The size of genomes is measured in base pairs (bp), the size of the human genome is 3.2×10^6 bp. The study of the human genome can be divided according to its repetitiveness and coding character (6), in non-repetitive DNA or single copy, which represents 70% of the entire human genome and the other 30% is formed by repetitive DNA (7).

Repetitive DNA comes in different shapes and sizes, which can be widely interspersed repeats, tandem repeats or nested repeats that can comprise only two or millions of copies and can vary in size from 1-2 bp to form sequences of millions of bases (8).

Tandem repeats (TR) are DNA sequences whose repeat units are short; due to their unique characteristics, which have been widely used as molecular markers in population genetics studies. They constitute 3% of the human genome and are grouped into three groups; microsatellites comprise tandem repeats ranging from 2 to 10 bp in length,

doi.org/10.31838/hiv23.03.24 Received: 05.10.22, Revised: 02.11.22, Accepted: 04.12.22

minisatellites are formed by tandem repeats between 10 and 60 bp and satellites ranging up to 100 bp (8).
The most widely used molecular markers in studying human populations are microsatellites or STR (Short Tandem Repeat). They are codominant and have simple Mendelian inheritance, there are several alleles for the same locus, so they are highly polymorphic. Their genotyping is relatively easy and automated; multiple amplifications can be performed in the same PCR reaction. The sample used does not need to be of good quality, and it is even possible to analyze DNA samples in an advanced state of degradation (9) (10).
The study of the STR profile involves the use of specific primers for regions that flank the microsatellite DNA to be amplified; then amplicons are generated by the PCR technique, and the amplified sequences are visualized in an agarose gel by electrophoresis, sized and assigned a numerical value, The data obtained are easy to interpret and to be compared. Moreover, determining the allelic and genotypic frequencies of these markers in a population helps to form a database that allows the molecular characterization of a population. These data can also be used in forensic statistical calculations and population genetics.
The study aims to identify alleles and determine allele frequencies of STR markers D16S539, Penta D, Penta E, Tpox present in the Southern Region of Peru.

2. Material and Methods

Study population: The study was conducted at the Molecular Biology Center of the National University of San Agustin, in the city of Arequipa-Peru. 100 buccal swab samples were analyzed from inhabitants of the southern region of Peru who attended the Molecular Biology Center between 2010 and 2014.
Informed consent was filled out, the study's objective was explained to each individual, the participants were of legal age, and only individuals whose family ancestry belonged to two generations ago born in the southern region of Peru were considered; none of the participants were related by blood.
DNA extraction and PCR
The MasterPure TM Epicentre kit was used for DNA extraction, the primers used are listed below:
FIRST STR: TPOX
Forward 5'-GCACACAGAACAGGCAGGCACTTAGG-3'
Reverse 5'-CGCTCAAACGTGAGGAGGTTG-3'
PRIMERS PENTA D
Forward 5'-
GGGGGGGTCTAAGAGAGCTTGTAAAAAAG-3'
Reverse 5'-
GTTTGTGTGTGTGCATCTCTGTAAGCATGTATATC-3'
FIRST STR: PENTA E
Forward 5'-ATTACCAACACATGAAAGGGGTACCAATA-3'
Reverse 5'-
TGGGGGTTATTAATTGAGAAAACTCCTTACAATTT-3'
FIRST STR: D16S539
Forward 5'-
GGGGGGGTCTAAGAGAGCTTGTAAAAAAG-3'
Reverse5'-
GTTTGTGTGTGTGCATCTCTGTAAGCATGTATATC-3'
For the amplification of STRs markers TPOX, D6S539, Penta D and Penta E, 1µL of DNA, 1µL Primers, 25µL Platinium PCR Super Mix (Invitrogen brand), the prepared samples were taken to the thermal cycler, the programmed activation temperature was 95° C for 5 min, followed by 35 cycles composed of Denaturation Temperature 95° C for 90 sec, Hybridization temperature 60° C for 60 sec. and Extension temperature 72° C for 60 sec. After 35 cycles, a cycle of 72° C for 10 minutes was added. The PCR product was analyzed in polyacrylamide gels to identify the alleles loaded with
5ul of the PCR result, for electrophoresis TAE 1x, 90V was used and the run time was 3 hours. The weight marker was 25 bpb Invitrogen brand, and SybrGreen were used for staining. The bands were observed with the Safe Imagen transilluminator, and photos were taken for analysis; for statistical calculations and forensic parameters, the Genepop program was used.

3. Results

Molecular markers such as microsatellites or STRs can be used in the identification of individuals, in studies of cadaveric remains, in the analysis of biological traces, in paternity tests, and in genetic variability studies, among others (Wyner (11)(12).

Table 1. Alleles identified for markers D16S539, PENTA D, PENTA E, TPOX in the human population of the Southern Region of Peru.

loci	Size of the amplified (bp)	Allele	Number of samples
D16S539	280	9	33
	284	10	48
	288	11	54
	292	12	36
	296	13	29
PENTA D	404	8	3
	409	9	14
	414	10	20
	419	11	71
	424	12	43
	429	13	26
	434	14	14
	439	15	1
PENTA E	404	10	12
	409	11	15
	414	12	31
	419	13	33
	424	14	34
	429	15	35
	434	16	13
	439	17	10
	449	19	7
	454	20	10
TPOX	270	8	120
	274	9	5
	282	11	50
	286	12	25

Source: Own elaboration

The study population consisted of 5 alleles for the

D16S539 marker 5 alleles, for the Penta D marker 8 alleles, 10 alleles identified for the Penta E STR and 4 alleles for the TPOX marker, as shown in Table 1.

Table 2. Allelic frequency of the TPOX, D16S539, Penta D and Penta E markers in the human population of the Southern Region of Peru.

ALELO	D16S539	PENTA D	PENTA E	TPOX
		0,0150		0,6000
	0,1650	0,0700		0,0250
	0,2400	0,1400	0.0600	
	0,2700	0,3550	0.0750	0,2500
	0,1800	0,2150	0.1550	0,1250
	0,1450	0,1300	0.1650	
		0,0700	0.1700	
		0,0050	0.1750	
			0.0650	
			0.0500	
			0.0350	
			0.0500	

Source: Own elaboration

Table 2 shows that the alleles identified for the STR marker D16S539 were 9, 10, 11, 12 and 13 being the most frequent allele 11, alleles 9,10,11,12,13,14,15 were identified for STR Penta D, the most frequent allele was allele 11, allele 15 was identified most frequently in the STR Penta E marker, the alleles identified for this marker were 10,11,12,13,14,14,15,16,17,19,20 and the STR Tpox marker identified 4 alleles 8,9,11 and 12, the most frequent being allele 8.

Table 3 shows the parameters of genetic and forensic interest. The highest number of alleles identified was for the STR marker Penta E, which presented the highest polymorphic content index (PIC) of 0.874. All STR markers obtained a $P < 0.05$, which indicates that our population is in Hardy Weinberg equilibrium.

The highest number of alleles and genotypes was presented by the STR marker Penta E and the lowest number by the Tpox marker. Of the 4 alleles studied, the Tpox marker was the least polymorphic and presented the lowest heterozygosity.

Table 3. Parameters of genetic and forensic interest for TPOX, D16S539, Penta D and Penta E markers in the human population of the Southern Region of Peru.

MARKER	D16S539	PENTA D	TPOX	PENTA E
Population Size	100	100	100	100
No of alleles	5	8	4	10
No of genotypes	14	18	7	37
Homozygotes observed	29	28	53	22
Expected homozygotes	20,7186	21,4874	43.5930	12.6080
Heterozygotes observed	71	72	52	78
Expected heterozygotes	79,2814	78,5126	56.4070	87.3920
Polymorphic Content Index	0,755	0,752	0.502	0.874
Hardy-Weinberg Equilibrium	P<0.05	P<0.05	P = 0.0335	P = 0.0001
Source: Own elaboration.				

4. Discussion

The genetic analysis of STR markers has been studied in different population groups to identify alleles present, allele frequencies, and genotypic frequencies and identify whether the study population is in Hardy Weinberg equilibrium, indicating a possible disruption in the genetic variability of a population due to a non-constant phenomenon such as a migration process (Gomes, 1999)(2).

The TPOX marker or locus is one of the most studied STRs in the study of populations. Its repeat unit is formed by four nucleotides that repeat in tandem (13). The alleles reported in the standard reference database NIST (National Institute of Standards and Technology USA) for the STR marker TPOX are 13. Of the 13 alleles reported in NIST, 8, 9, 11 and 12 were identified in our study population. In Peru, the STR marker TPOX was studied by Cruz Calvo in 2006 in the province of Andahuaylas. In this study, the STR markers TPOX and TH01 were analyzed, and the presence of 5 alleles (8, 9, 10, 11 and 12) was reported for the TPOX marker (14). In the study conducted by Abello in 2001 in a Colombian population, where the study population was 50 samples, 7 alleles were reported (6, 7, 8,9,10,11,12), with allele 8 being the most frequent with a frequency of 43.8% (15). And in the study of Arias in La Paz, Bolivia, in a mestizo population, the frequency for allele 8 of the STR TPOX marker was only 37%. It is worth noting that in this study population, 8 of the 13 alleles registered in the NIST were present, which would explain the lower percentage of presentation for allele 8 of STR TPOX (16). Based on these studies, allele 8 of STR TPOX predominates in Latin American countries.

Table 1 also shows that the STR marker D16S539 identified 5 alleles (9,10,11,12,13): allele 11 was identified with greater frequency in the human population of the Southern Region of Peru, the STR markers Penta D and Penta E identified 8 and 10 alleles respectively, the most frequent alleles were 11 for the STR Penta D and 15 for the Penta E marker. In the study, the most polymorphic STR was Penta E, which had the highest number of alleles and was the STR with the greatest discriminatory power. In the study carried out by Rey in a Colombian population, STRs D2S1338, D19S433, PENTA D, PENTA E and SE-33 were studied, STR Penta D identified 10 alleles and STR Penta E identified 15 alleles, and STR Penta E was the most polymorphic marker, as in our study, the one with the greatest discriminatory power (17).

Table 3 shows the observed and expected genotypes for the 4 markers studied. To verify that the population has not changed from generation to generation and is stable in Hardy Weinberg equilibrium. The Chi-square test was applied and obtained for these 4 markers a $P< 0.05$ value that indicates that the population is in genetic equilibrium for these markers. The allelic diversity of molecular markers is known as polymorphisms; the study of

these polymorphisms has multiple applications in the medical area and the study of populations. For example, the polymorphisms present in the different STR markers are responsible for the genetic variability between individuals of a population (same species), making each member of the same species unique and unrepeatable; polymorphisms are also responsible for the evolution of species (18). Likewise, their study can help identify the susceptibility to present certain diseases by incorporating these markers into molecular epidemiology (19).

5. Conclusion

Four alleles were identified for the Tpox marker, 5 alleles for the STR D16S539, 8 alleles for the Penta D marker and 10 alleles were identified for the PENTA E marker in the population of the Southern Region of Peru. Therefore, the STR with the highest polymorphic and discriminatory power was STR Penta E in the Southern Region of Peru population. Therefore, STRs markers TPOX, D16S539, PENTA E and Penta D are in Hardy Weinberg equilibrium in the human population of the Southern Region of Peru with a $P<0.05$ for all markers.

References

Schulman, A.H. Molecular markers to assess genetic diversity. *Euphytica* 158, 313–321 (2007). https://doi.org/10.1007/s10681-006-9282-5

Gomes, I., Collins, A., Lonjou, C., Thomas, N., Wilkinson, J., Watson, M., & Morton, N. (1999). Hardy–Weinberg quality control. Annals of Human Genetics, 63(6), 535-538. doi:10.1046/j.1469-1809.1999.6360535.x

Silva Zolezzi, I. (2011). Genómica y medicina. *Educación química*, *22*(1), 15-27. doi:10.1016/s0187-893x(18)30109-5.

WATSON, J., CRICK, F. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature171, 737-738(1953). https://doi.org/10.1038/171737a0.

Martínez-Frías, M. L. (2010). Estructura y función del ADN y de los genes. I Tipos de alteraciones de la función del gen por mutaciones. *Semergen*, *36*(5), 273-277.

Reguero Reza, Ma. Teresa. (2013). 10 años de la secuenciación del genoma humano: Encuentro entre el imaginario y la realidad. Revista Colombiana de Biotecnología, 15(1), 5-7

Corella, D., & Ordovas, J. M. (2017). Conceptos básicos en biología molecular relacionados con la genética y la epigenética. Revista Española de Cardiología, 70(9), 744– 753.

Treangen, T. J. y Salzberg, S. L. (2011). ADN repetitivo y secuenciación de próxima generación: desafíos y soluciones computacionales. Reseñas de la naturaleza. Genética, 13(1), 36-46. https://doi.org/10.1038/nrg3117

González, E. G. (2003). Microsatélites: sus aplicaciones en la conservación de la biodiversidad. Graellsia, 59(2-3), 377–388.

Barletta C., Builes J.J., Bravo M.L., Aguirre D., Descailleaux J., Gusmao L.(2013) Genetic data of 10 X-STRs in a population sample from Lima, Perú Forensic Science International: Genetics Supplement Series, 4 (1) , pp. e168-e169

Wyner N, Barash M and McNevin D (2020) Forensic Autosomal Short Tandem Repeats and Their Potential Association With Phenotype. Front. Genet. 11:884. doi: 10.3389/fgene.2020.00884

Andleeb Zahra, Bilal Hussain, Amer Jamil, Z. Ahmed, Shahid Mahboob. (2018) Forensic STR profiling based smart barcode, a highly efficient and cost effective human identification system, Saudi Journal of Biological Sciences, Volume 25, Issue 8, Pages 1720-1723.

Lukka, M., Tasa, G., Ellonen, P., Moilanen, K., Vassiljev, V., & Ulmanen, I. (2006). Triallelic patterns in STR loci used for paternity analysis: Evidence for a duplication in chromosome 2 containing the TPOX STR locus. Forensic Science International, 164(1), 3–9. doi:10.1016/j.forsciint.2005.11.006 10.1016/j.forsciint.2005.11.006

Cruz Calvo Augusto Fernando, (2006) Análisis de la Variabilidad Genética en una muestra Poblacional de la Provincia de Andahuaylas – Apurímac, Perú - con los marcadores Microsatélites TPOX y TH01.

Abello, A., Carris, M. (2001). Análisis Genético Poblacional de 50 muestras Provenientes de Barranquilla utilizando tres Loci Microsatélites CTT. Acta Biol. Colomb., Volumen 6, Número 2, p. 14. ISSN electrónico 1900-1649. ISSN impreso 0120-548X

Arias Loza, Anahí y Colb., Análisis de 9 Loci STR en la Población Mestiza de la Paz para la Identificación de Individuos. Cuadernos del Hospital de Clínicas. Volumen 47 Número 2.

Rey Mauricio, Usaquen William (2009). Datos Poblacionales de los Microsatélites de ADN Humano D2S1338, D19S433, PENTA D, PENTA E Y SE-33 de la Región Central Colombiana. Acta Biol. Colomb. Vol. 14, N. 2, 161 –168.

Torrades Sandra. (2002). Diversidad del genoma humano: Los polimorfismos Offarm. Vol.21. Núm 5 Mayo 2002

Iniesta, Raquel, Guinó, Elisabet, & Moreno, Víctor. (2005). Análisis estadístico de polimorfismos genéticos en estudios epidemiológicos. Gaceta Sanitaria, 19(4), 333-341.