INT. BIOL. BIOTECH., 18 (2): 275-279, 2021.

ESTIMATION OF ALLELE'S FREQUENCY OF FOUR STR MARKERS (D5S818, D7S820, D13S317 & D16S539) IN PASHTUN POPULATION IN BALOCHISTAN, PAKISTAN

Palwasha Khan*, Shakeela Daud, Asma Abro, Nida Ajmal and Saira Jamali

Department of Biotechnology, Faculty of Life Sciences and Informatics, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Balochistan, Pakistan.
*Corresponding author: Tel: +92 (0)312-8320733
E-mail: khanpalwasha224@gmail.com

ABSTRACT

STR markers, also known as microsatellites are dispersed across the whole genome of humans. STR markers are highly polymorphic due to different sort of mutations being carried out in genome. In forensics, genotyping of polymorphic STR markers lead to the solution of different issues such as paternity cases and identification of individuals. The present study aims in identification and determination of allele's frequencies of four STR loci D5S818, D7S820, D13S317 and D16S539 in Pashtun population of Quetta Balochistan, Pakistan. Blood samples from 45 random individuals were collected after consent and were processed for DNA extraction. Amplification of STR markers by PCR followed by genotyping was performed. For marker D5S818, allele 147 bp was most common among all nine observed alleles with allele frequency of 37.77 % and allele 127 bp was least common with allele frequency of 6.66 %. For marker D7S820 most prevalent allele (62.22 %) was 212 bp among 6 observed alleles and alleles 200 bp and 208 bp were present in less individuals (11.11 %). Similarly,152 pb (66.66 %) was common allele among all individuals and allele 180 bp (11.11 %) was least common for marker D13S317. Out of five alleles observed for marker D16S530, allele 152 bp was commonly present (66.66 %) and allele 140 bp was least common allele (13.33%).

Keywords: Marker, Polymorphism, Fingerprinting, Microsatellites, Genotyping

INTRODUCTION

Human genome is composed of two types of DNA sequences, coding and non coding DNA sequences. Non coding DNA sequences that do not code for any proteins are also known as "junk DNA" (Watson, 2003). Junk DNA has repetitive sequences of DNA of different lengths that vary from individual to individual (Lewis, 2005). Thus a technique known as DNA fingerprinting, DNA profiling or DNA typing was developed on the basis of polymorphism in these repetitive sequences (Bains, 2004).

Repetitive sequences present in human genome are of two kinds, minisatellites and microsatellites. Term minisatellite was coined in 1985 by Alec Jefferys (Jefferys, 1885) and coworkers that are tandemly repeated sequences of up to about 100 bp. Microsatellites are made of relatively short DNA sequences usually 1-10 bps. Microsatellites are also known as "Simple Tandem Repeats" or "Short Tandem Repeats" (Edwards *et al*., 1991).

STR analysis is the most common method used in forensics for person identification and paternity testing because STR loci has short repeat units and they are distributed all over human genome.Among these loci there are number of highly polymorphic markers which are identified through the frequency of repeat units (Lewis, 2005). STR analysis requires very less amount of DNA or even degraded DNA can also be used. These characteristics of STR loci make them effective for human identification (Butler, 2001). It has been suggested by different studies that main source of polymorphism in microsatellites is strand slippage mechanism during DNA replication (Tautz and Renz, 1984; Levinson and Gutman, 1987).

Frequencies of DNA profiles obtained after conducting survey of some populations are usually collected into genetic data. Genetic data bases are assembly of DNA profiles which are taken from randomly selected persons of a specific population (Butler, 2001). Examples of present genetic databases of human allele's frequencies are African Jordanian, Rwandan, Spanish and American populations (Yasin *et al*., 2005; Budowle *et al*., 2005; Linacre, 2001). These allele frequencies are further used to calculate genotype frequencies and interpret DNA profiles.The result obtained roughly calculates the likelihood that two persons from a specified population have the similar genotype for every DNA loci analyzed (Lewis, 2005).

Thus, allele frequencies act as a source reference for assessing human genome statistically. They play an important role to give clear results in genome identification when traditional methods are not sufficient enough. The

aim of the present study was to determine allele frequencies in four short tandem repeat loci that are D5S818, D7S820, D13S317 and D16S539 of Pashtoon population in Quetta, Balochistan Pakistan.

MATERIALS AND METHODS

Enrollment of study population

Blood samples were randomly collected form 45 healthy individuals of Pashtoon population in Quetta, Balochistan Pakistan.

Inclusion and exclusion criteria

People belonged to Pashtoon population were randomly selected and included in this study. Individuals from other ethnic groups and those who were diagnosed with diseases were excluded.

DNA extraction and estimation

3 mL blood samples were drawn from the people involved in the study with the proper written informed consent. Inorganic method of DNA extraction was used to extract the DNA from blood.DNA samples were estimated by performing gel electrophoresis on 1% agarose gel.

Primers

Primers used for amplification of STR markers are fluorescently labeled with dyes FAM, VIC JOE and TMR.
D5S818 Forward 5-3 -GATCCCAAGCTCTTCCTCTT **Reverse 5-3**ACGTTTGTGTGTGCATCTGT
D7S820 Forward 5-3 -GGGTGATTTTCCTCTTTGGT **Reverse 5-3** TGATTCCAATCATAGCCACA
D13S317 Forward 5-3 -ACAGAAGTCTGGGATGTGGA **Reverse 5-3** GCCCAAAAAGACAGACAGAA
D16S539 Forward 5-3 -TGTCATAGTTTAGAACGAACTAACG **Reverse 5-3** CTGAGGTATCAAAAACTCAGAGG

Polymerase chain reaction (PCR)

Amplification of all four markers (D5S818, D7S820, D13S317 and D16S539) was performed by PCR.Reactions were performed in 0.2 mL PCR tubes on a BioRad™ thermo-cycler. Reactions were performed with 1ul of template DNA (0.5-2ng) in 20 µl reaction mixture containing 10ul of master mix [BIORON; *Taq*Master Mix (2X)] 7 uL of PCR water, and 1 uL of each of forward and reverse primers.Touchdown thermo cycler program used for amplification that was performed with an initial denaturation step at 95°C for 4 min followed by first 10 cycles as touchdown PCR (with annealing temperature from 60°C to 55°C) and additional 20 cycles with denaturation at 94°C for 45 sec, annealing at 55°C for 1 min and extension at 72°C for 1 min with a final extension at 72°C for 20 minutes. PCR products were confirmed on 2% agarose gel (Goodwin *et al*., 2007).

Genotyping

Genotyping was performed on ABI PRISM 3130 Genetic Analyzer. After genotyping the data was analyzed and allele frequencies were calculated for each STR marker.

RESULTS AND DISCUSSION

Amplification results of all four markers are shown in Fig. 1, 2, 3 and 4. Allele frequencies calculated after performing genotypes are shown in Table 1.

For the first marker D5S818, observed alleles were 127 bp, 131 bp, 135 bp, 139 bp, 143 bp, 147 bp, 151 bp, 155 bp and 159 bp. Among these nine alleles, most common allele was 147 bp observed in 37.77 % individualswith the allele frequency of 18.88. Allele 143 bp was second most common allele present in 33.33 %individuals with the allele frequency of 16.66. other alleles 139 bp, 159 bp, 155 bp, 151 bp, 135 bp and 131 bp were observed in 26 %, 22.22 %, 17.77, 13.33 %, 11.11% and 11.11 individuals withallele frequencies of 13.33, 11.11, 8.88, 6.66, 5.55 and 5.55 respectively. 127 bp was least prevalent present in 6.66 % individuals and its frequency was 3.33.

For second marker D7S820, among six observed alleles, most commonly present allele was 212 bp (62.22 %) and calculated allele frequency was 31.11. Alleles 216 bp, 220 bp and 204 bp were present in 33.33 %, 28.88 and 20% individuals and their frequencies are 16.66, 14.44 and 10 respectively. Alleles 200 bp and 208 bp were least prevalentobserved in 11.11 % and 11.11 % individuals and their frequencies are 5.55 and 5.55.

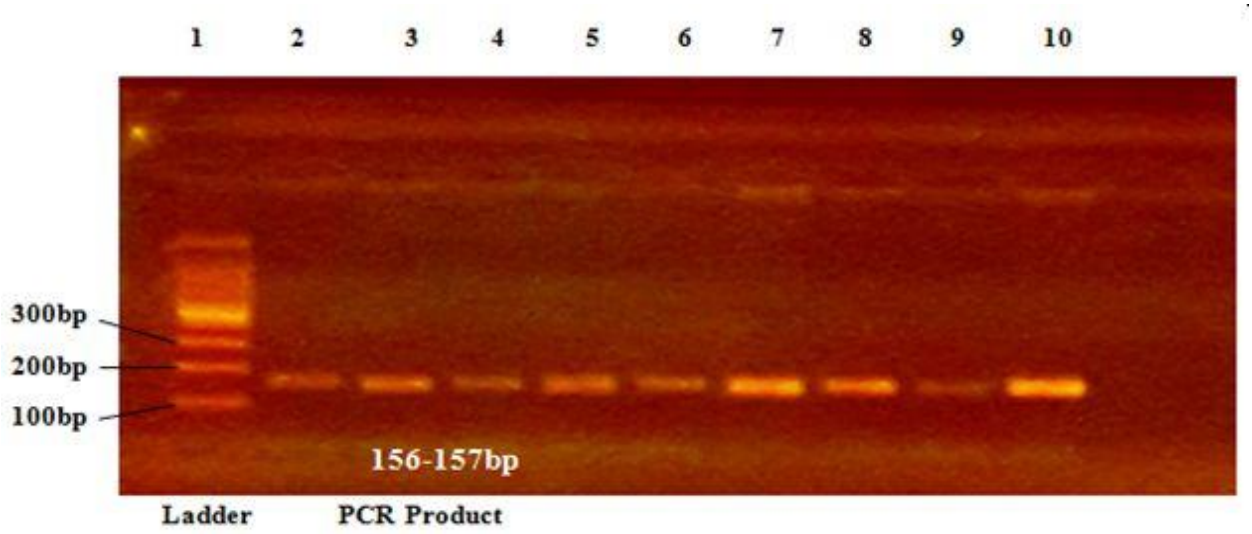
Among six observed alleles for marker D13S317, most common allele was 188 bp (53 %) and its frequency was 26.66. Other alleles 184 bp, 172 bp, 192 bp, and 176 bp were present in 44.44 %, 26.66 %, 20 % and 15.55 %

individuals and their frequencies are 22.22, 13.33, 10 and 7.77 respectively. Least prevelant allele was 180 bp (11.11 %) and its frequency is 5.55.

For marker D16S539, only five alleles were observed among which 152 bp was most common (66.66 %) with the allele frequency of 33.33. Other observed alleles were 144 bp, 148 bp and 156 bp that were present in 31.11 %, 28.88 % and 26.66 % and their frequencies are 15.55, 14.44 and 13.33 respectively. The least common allele was 140 bp. It was present in 13.33 % individuals and its frequency is 6.66.

Table 1. Under study STR markers, their observed alleles, number of observations, total sample size and calculated frequency.

Marker	Allele	No. of observations	Total sample size	Frequency
D5S8181	127	3	45	3.33
	131	5	45	5.55
	135	5	45	5.55
	139	12	45	13.33
	143	15	45	16.66
	147	17	45	18.88
	151	6	45	6.66
	155	8	45	8.88
	159	10	45	11.11
D7S820	200	5	45	5.55
	204	9	45	10
	208	5	45	5.55
	212	28	45	31.11
	216	15	45	16.66
	220	13	45	14.44
D13S317	172	12	45	13.33
	176	7	45	7.77
	180	5	45	5.55
	184	20	45	22.22
	188	24	45	26.66
	192	9	45	10
D16S539	140	6	45	6.66
	144	14	45	15.55
	148	13	45	14.44
	152	30	45	33.33
	156	12	45	13.33



Fig, 1. Amplification result of marker D5S818.

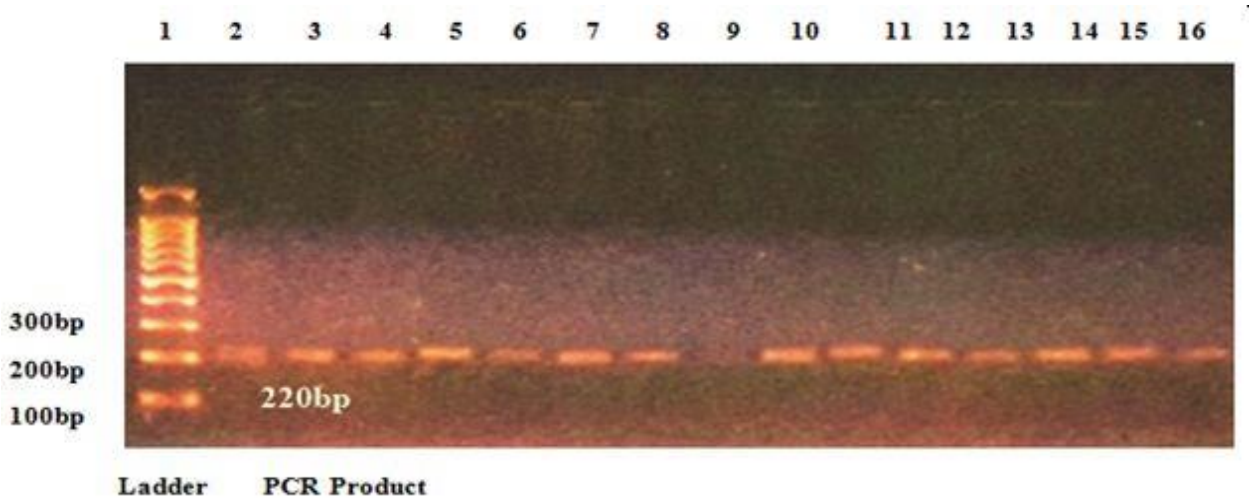


Fig. 2. Amplification result of marker D7S820.

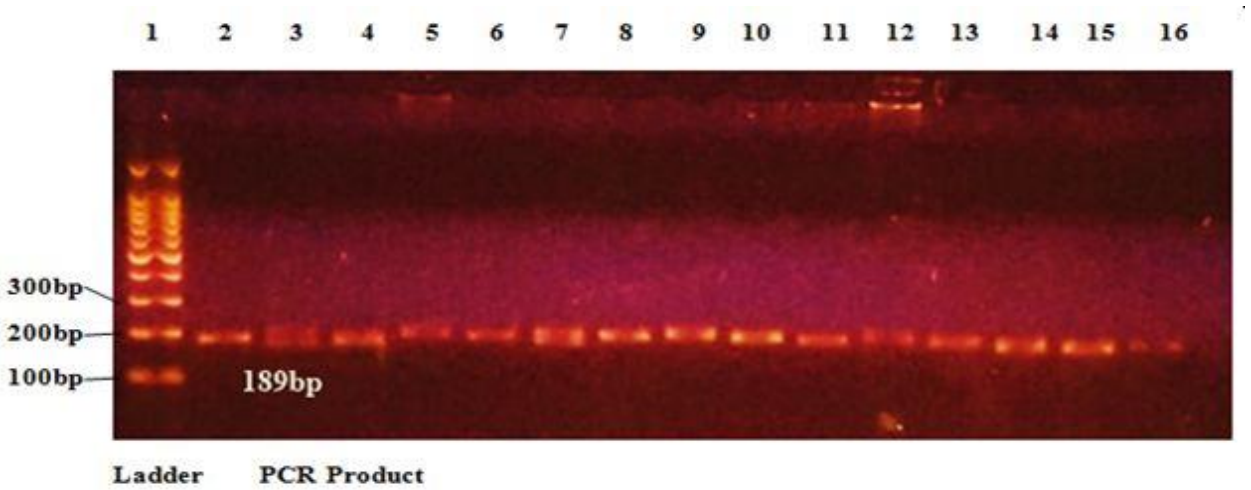


Fig. 3. Amplification result of marker D13S317.

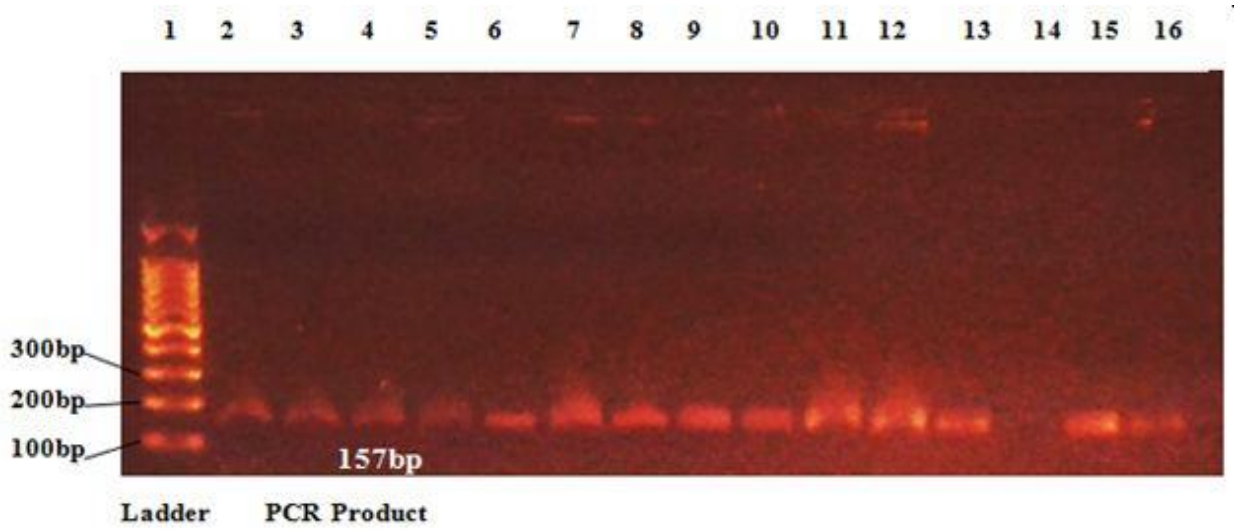


Fig. 4. Amplification result of marker D16S539.

Acknowledgements

Firstly we acknowledge our teachers, family and friends, who provided every kind of help during this study.

Funding Acknowledgements

We acknowledge the Institutional Review Board (IRB) BUITEMS for protocol approval and Organization of Research Innovation and Commercialization (ORIC), BUITEMS Quetta for research funding.

Conflict of interest

The authors declare that there is no conflict of interest.

REFERENCES

Watson, J. (2003). *DNA The secret of life*. Random House Inc. New York.

Lewis, R. (2005). *Human Genetics: Concepts and Applications*. 6th edition. McGraw-Hill.

Bains, W. (2004). *Biotechnology from A to Z*. 3rd edition. New York: Oxford University Press Inc.

Jefferys, A.J., V. Wilson and S.L. Thien (1985). Hypervariable 'minisatellite' regions in human DNA**.** *Nature.,* 314:67–74.

Edwards, A., A. Civitello, H.A. Hammond and C.T. Caskey (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet.*, 49: 746 -756.

Butler, J. (2001). *Forensic DNA Typing: Biology and Technology behind STR Markers*. Academic Press.

Tautz, D. and M. Renz (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucl. Acids Res.,* 12: 4127–38.

Levinson, G. and G.A. Gutman (1987). Slipped-strand impairing: A major mechanism for DNA sequence evolution. *Mol Biol Evol.,* 4(3): 203-21.

Yasin, S.R., M.M. Hamad, A.Z. Elkarmi and A.S. Jaran (2005). African Jordanian Genetic Database on Fifteen Short Tandem Repeat Genetic Loci. *Croatian Med. J.,* 46: 587-592.

Linacre, A.M.T., P. Phillips, A. Wally and W. Goodwin (2001). STR data for the SGM PlusTM loci from three Indonesian populations. *Forensic Sci. Int.*, 123: 232-234.

Budowle, B., M. Adamowicz, X.G. Aranda, C. Barna, R. Chakraborty, D. Cheswick, B. Dafoe, A. Eisenberg, R. Frappier, A.M. Gross, C. Ladd, H-S. Lee, S.C. Milne, C. Meyers, M. Prinz, M.L. Richard, G. Sladahna, A.A. Tierney, L. Viculis and B.E. Krenke (2005). Twelve short tandem repeat loci Y chromosome haplotypes: Genetic analysis on populations residing in North America. *Forensic Sci.,* 150: 1-15.

Goodwin, W., A. Linacre and S. Hadi (2007). *An Introduction to Forensics Genetics*. John Wiley & Sons, Ltd.

(*Accepted for publication March 2021*)