ARCH. MED. SĄD. KRYM., 2004, LIV, 215-222 PRACE ORYGINALNE

**M. Saqib Shahzad, S. Yassir Abbas Bokhari, Abdul Qayyum Rao, M. Hashim Raza, Obaid Ullah, Zia-ur-Rahman, A. Ali Shahid, Zahoor Ahmad, S. Riazuddin**

# Population Studies for STR Loci (D3S1358, D5S818, D7S820, D18S51 and FGA) in NWFP and Sindhi Populations of Pakistan for Forensic Use

## Badania populacyjne loci STR (D3S1358, D5S818, D7S820, D18S51 i FGA) w wybranych populacjach Pakistanu (NWFP i Sindhi) w praktyce sądowo-lekarskiej

Centre of Excellence in Molecular Biology (CEMB),
University of the Punjab, Lahore, Pakistan.

CEMB's Forensic DNA typing project is directed towards the introduction of DNA typing technique in Pakistan's criminal justice system so as to exonerate an innocent and wrongly accused person and incriminates the culprit. The present study is a part of the project of CEMB to analyze Sindhi and NWFP (North West Frontier Province) populations for five STR (Short Tandem Repeat) loci out of 13 CODIS (Combined DNA Index System) loci. Allelic frequencies and heterozygosity for STR markers D3S1358, D5S818, D7S820, D18S51 and FGA (FIBRA) were determined. Samples from unrelated individuals were amplified by multiplex PCR using the unlabelled primers for these markers followed by denaturing Polyacrylamide Gel Electrophoresis (PAGE). Statistical analysis was performed to determine the allelic frequencies and was evaluated using the Chi Square Test.

Autorzy przedstawiają wyniki populacyjnych badań pięciu loci DNA, zrealizowanych w ramach projektu w Centre of Excellence in Molecular Biology w Pakistanie. Próbki od 209 niespokrewnionych osób zostały oznaczone dla układów D3S1358, D5S818, D7S820, D18S51 i FGA zgodnie z zaleceniami Combined DNA Index System. Wyniki przedstawiono w postaci tabelarycznej. Częstości alleli były analizowane statystycznie z zastosowaniem testu $Chi^2$.

Keywords / słowa kluczowe:
Forensic biology, STR loci, Allelic frequency, NWFP (North West Frontier Province) and Sindhi population.

Abbreviations / skróty:
North West Frontier Province (NWFP)
Centre of Excellence in Molecular Biology (CEMB)
FGA (FIBRA)
Short Tandem Repeat (STR)
Polyacrylamide Gel Electrophoresis (PAGE)
DeoxyRiboNucleic Acid (DNA)
Amplified Fragment Length Polymorphism (AmpFLP)
Variable Number of Tandem Repeats (VNTR)
Combined DNA Index System (CODIS)
Tris EDTA (TE)
Polymerase Chain Reaction (PCR)

## INTRODUCTION

The genetic uniqueness of individuals is a central tenet of human biology. This uniqueness is defined by the combination of genetic markers that an individual inherits from his or her parents. Genetic markers can be analyzed either at the level of protein or DNA variation. DNA technology makes possible the study of human variability at the most basic level, the level of genetic material, DNA (Deoxy Ribonucleic Acid) [17]. Traditionally, the term DNA fingerprinting refers to the patterns, which are highly characteristic for any human individual, of the ridged skin of the distal finger phalanges. The genetic complexity of DNA fingerprints is enormous [11].

Previous methods using blood groups and proteins have analyzed gene products, rather than DNA itself. In addition to providing more direct genetic information, DNA can withstand environmental conditions that destroy proteins, so old, badly degraded samples of body fluids still can provide abundant information as DNA typing has been done of a human body exhumed after 27 years [20].

The advent of DNA typing in the mid-1980s has had an enormous effect on the ability of crime laboratories to identify individuals uniquely by testing a variety of their body fluids. In the past 15 years, technologies for individual identification have moved at a breathtaking pace. The first forensic samples to be typed by DNA procedures required at least 25–100 times more sampled DNA than is currently needed for a result. The incorporation of modern molecular biological techniques in the crime laboratory has resulted, at a remarkably increasing rate, in the identification of criminals and the exoneration of the innocent. The advent of modern DNA technology has resulted in the

increased ability to perform human identity. Individual identification is desirable in a number of situations including the determination of perpetrators of violent crime such as murder and rape, resolving unestablished paternity, and identifying remains of missing persons or victims of mass disasters.

The tandemly repeated sequence of 3 to 7 bp is core sequence of STR DNA. These abundant repeats are well distributed throughout the human genome and are rich sources of highly polymorphic markers, which often may be detected using PCR [18]. Polymorphic STR markers are those, which are varying in the length of core repetitive sequence of nucleotides. STR typing is more tolerant to the use of degraded DNA templates than other methods of individual identification because the amplification products are less than 400bp long, much smaller then the material detected with AmpFLP (Amplified fragment Length Polymorphism) [6] or VNTR (Variable Number of Tandem Repeats) [14] analysis. The STR markers were selected according to their high discrimination power (>0.9) and high heterozygosity (>80%) already reported for different population of the world [7]. Thirteen STR loci were selected for use in the combined DNA index system shortly called CODIS [4], which is STR database of genotype of selected STRs.

If a probability calculation on the basis of the population data is needed then the forensic DNA scientist must know genotype of that specific local population for selected STR marker [1]. This indicates the necessity of preparation and maintenance of records of genotypes of a specific STR marker or database of selected STR marker for the local population. All those countries of the world that are exploiting STR DNA for crime investigation have established databases.

Results of the present study will not only help the DNA analyst to generate the DNA profile for five STR loci but also give facts and figures about allelic frequency, genotypic frequency and heterozygosity of the selected STR markers for the Pakistani population. The finding of this study will be useful to undertake statistical analysis in order to prove the matching of the DNA profile between crime scene sample and the suspect to be unique in our population.

## MATERIALS AND METHODS

Blood samples were collected from 209 unrelated informed volunteers from NWFP (North West Frontier Province) and Sindh provinces. From each individual, 3ml of blood was drawn. Phenol chloroform DNA extraction [19] was performed on each of the blood samples. The quantity of the extracted DNA in the TE (Tris EDTA) solution was estimated initially by Spectrophotometery followed by Yield Gel Electrophoresis.

PCR (Polymerase Chain Reaction) conditions for all five STR loci (D3S1358, D5S818, D7S820, D18S51 and FGA) were optimized using MJ Research Inc. PTC-100 and Hybaid Thermal Cycler and then PCR for all these loci was run

for all the collected samples. PCR products were checked on 2% maxi agarose gel. Denaturing Polyacrylamide Gel Electrophoresis (PAGE) was run to check the genotypic ratios and allelic frequencies of the populations of both the provinces for five STR markers. The data were statistically analyzed using the Chi Square test.

## RESULTS

The genotypic data of NWFP (North West Frontier Province) and Sindhi populations was statistically analyzed. This statistical analysis included; observed allelic frequencies, observed and expected genotypic ratios and observed and expected heterozygosities. The Hardy Weinberg Equilibrium was tested for this data by the chi square test. Table I shows the allelic frequencies of loci D3S1358, D5S818, D7S820, D18S51 and FGA.

## DISCUSSIONS

The establishment of a reference genotypic database of Sindhi and NWFP (North West Frontier Province) populations and obtained statistical findings are presented for these populations in order to use for probability calculations during crime investigation and in forensic casework. The Pakistani population is ethnically diverse and consists of four provinces and many sub groups. Population data for STR loci has been reported by a number of scientists around the world [3, 9, 10, 13, 16]. NWFP and Sindhi population have been screened on non-ethnic basis. The DNA was extracted from whole blood samples using an organic extraction method. The DNA extraction by Phenol Chloroform method gives good yield [19] and is suitable for PCR based DNA typing technique [8].

Each sample was amplified for the five STR loci and resolved on 8% denaturing PolyAcrylamide gel followed by silver staining. Bassam et al. [2], Budowle and Allen [5] has used denaturing polyacrylamide gel and found this technique useful for the resolution of STR loci. Amplicons of the developed multiplex were pooled and reprecipitated. Different alleles at each locus were determined using 50bp DNA ladder marker. At each locus most of the reported alleles were found in sampled population.

Allele frequencies of locus D3S1358, D5S818, D7S820, D18S51 and FGA ranged from 0.1650-0.0071, 0.0574-0.1578, 0.1100-0.0550, 0.0669-0.0789, 0.0239-0.0119 for alleles 13-17.1, 8-12, 5-11, 9.1-19.3 and 16-25 respectively (See Tab. I for all the allele frequencies).

Table I Allelic frequencies of locus D3S1358, D5S818, D7S820, D18S51, FGA.
Tab. I. Częstości alleli badanych loci.

| Allele No. | D3S1358 | D5S818 | D7S820 | D18S51 | FGA |
|---|---|---|---|---|---|
| 5 | – | – | 0.1100 | – | – |
| 5.3 | – | – | 0.1889 | – | – |
| 7 | – | – | 0.1196 | – | – |
| 8 | – | 0.0574 | 0.1483 | – | – |
| 8.2 | – | 0.1076 | – | – | – |
| 8.3 | – | – | 0.2105 | – | – |
| 9.1 | – | – | – | 0.0669 | – |
| 9.2 | – | 0.0574 | – | – | – |
| 9.3 | – | – | 0.1674 | – | – |
| 10 | – | – | – | 0.0574 | – |
| 10.1 | – | 0.2775 | – | – | – |
| 11 | – | – | 0.0550 | 0.0669 | – |
| 11.1 | – | 0.1818 | – | – | – |
| 12 | – | 0.1578 | – | 0.1028 | – |
| 13 | 0.1650 | – | – | – | – |
| 13.1 | – | – | – | 0.1267 | – |
| 14 | 0.2153 | – | – | 0.1196 | – |
| 15 | 0.2775 | – | – | – | – |
| 15.1 | – | – | – | 0.1339 | – |
| 15.2 | 0.0095 | – | – | – | – |
| 16 | 0.2129 | – | – | – | 0.0239 |
| 16.2 | 0.0119 | – | – | 0.1052 | – |
| 17 | 0.0114 | – | – | – | 0.0598 |
| 17.1 | 0.0071 | – | – | – | – |
| 17.2 | – | – | – | 0.0909 | – |
| 18 | – | – | – | – | 0.1578 |
| 18.2 | – | – | – | – | 0.0574 |
| 18.3 | – | – | – | 0.0502 | – |
| 19 | – | – | – | – | 0.0095 |
| 19.1 | – | – | – | – | 0.0909 |
| 19.3 | – | – | – | 0.0789 | – |
| 20 | – | – | – | – | 0.1889 |
| 20.2 | – | – | – | – | 0.0909 |
| 21 | – | – | – | – | 0.0454 |
| 21.2 | – | – | – | – | 0.0717 |
| 22.2 | – | – | – | – | 0.0837 |
| 23 | – | – | – | – | 0.0861 |
| 24 | – | – | – | – | 0.0047 |
| 24.2 | – | – | – | – | 0.0167 |
| 25 | – | – | – | – | 0.0119 |

Nazneen et al. [15] have reported the allele frequencies for loci vWA, D3S1358 and D16S539 in the Lahori population at random. According to their findings, the allele frequencies for locus D3S1358 ranged from 0.050 for allele number 13 to 0.080 for allele number 17. Their data varied from our findings which were for the NWFP (North West Frontier Province) and Sindhi populations of Pakistan, where frequencies ranged from 0.1650-0.0071 for alleles 13-17. There is no published data for all the other five loci. Shaista Khanum has reported [12] the allele frequencies for the locus D18S51 in Punjabi population of Pakistan in her thesis, which ranged from 0.0142-0.0020 for alleles 9.2-25. This also varied from our data, which showed a range from 0.0669-0.0789 for alleles numbered 9.1-19.3. This comparison indicates that the overall Pakistani population is diverse and allele frequencies vary from place to place.

The Chi square test was applied to check the population for Hardy Weinberg Equilibrium. Chi Square values are 0.5754 for D3S1358, 3.5402 for D5S818, 1.4840 for D7S820, 0.8937 for D18S51 and 0.04834 for FGA indicating that the results are non significant and obey the Hardy Weinberg Equilibrium so these five STR makers alone are not sufficient enough to generate a database. The discrimination power of these STR markers is insufficient but the high heterozygosity values [Tab. II] suggest that in combination with further markers, these STR markers can be used to generate a database with full confidence and accuracy.

Table II Heterozygosity at three STR loci (D5S818, D7S820 AND D18S51) in Pakistani population

Tab. II. Heterozygotyczność w zakresie trzech badanych loci STR

| Locus | Expected<br>Oczekiwana | Observed<br>Obserwowana | Chi square |
|---|---|---|---|
| D3S1358 | 79.3663 | 86.124 | 0.5754 |
| D5S818 | 80.2717 | 97.1292 | 3.5402 |
| D7S820 | 84.046 | 95.215 | 1.4840 |
| D18S51 | 90.0712 | 99.0431 | 0.8937 |
| FGA | 89.3192 | 91.388 | 0.04834 |

These values are non-significant which indicate that our population obeys the Hardy – Weinberg equilibrium.

## REFERENCES

1.Balding D.J., Nichols R.A. DNA match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci. Int. 1994, 64, 125-140. – 2. Bassam B. J., Caetano-Anolles G., Gresshoff P. M. Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal. Biochem. 1991, 196, 80. – 3. Bever R.A., Creacy S. Validation and commercialization of available STR multiplexes for parentage analysis. Proceedings for the 5th International Symposium on Human Identification, 1995, Pp.61-63. – 4. Budowle B. Examples of STR population database for CODIS and for case work Presented at the 9th Symposium on Human Identification. Orlando, FL, 1998. – 5. Budowle B., Allen R. C. Discontinuous polyacrylamide gel electrophoresis of DNA fragments. In: Mathew C. (ed.). Methods in Molecular Biology– Molecular Biology in Medicine. Vol. 7. Humana, London, 1990. – 6. Budowle B., Baechtel F. S., Comey C., Guisti A., Klevan L. Simple protocols for Typing Forensic Biological Evidence: Chemiluminescent Detection for Human DNA Quantitation and Restriction Fragment Length Polymorphism (RELP) Analysis and Manual Typing of polymerase chain reaction (PCR) Amplified polymorphisms. Electrophoresis 1995, 16, 1559-1567. – 7. Budowle B., Moretti T. R., Baumstark A. L., Defenbaugh D. A., Keys K. M. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J. Forensic Sci. 1999, 44 (6), 1277-1286. – 8. de Pancorbo M. M., Castro A., Fernandez I., Garcia – Orad A. Population genetics and forensic application using multiplex PCR (CSFIPO, TPOX, and TH01) loci in the Basque Country. J. Forensic Sci. 1998, 43, 1181-1187. – 9. Edwards A., Hammond H.A., Chakraborty R. Genetic variation at five trimeric and tetra meric random repeats loci in four human population groups. Genomics 1992, 12, 241-253. – 10. Hammond H. A., Jin L., Zhong Y., Caskey C. T., Chakraborty R. Evaluation of 13 short Tandem Repeat loci for use in personal Identification Applications. Am. J. Hum. Genet. 1994, 55, 175-189.

11. Jefferys A. J., Wilson V., Thein L. S., Weaterall D. J., Ponder, B. A. J. Forensic DNA data banking by state crime laboratories. Am. J. Hum. Genet. 1986, 39, 11. – 12. Khanum S. Allelic frequency distribution of STR loci (D8S1179, D18S51 and D21S11) in punjabi population of Pakistan. Thesis, National Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan, 2002. – 13. Lins A.M., Sprecher C.J., Puers C., Schumm J.W. Multiplex sets for the amplification of polymorphic STR loci, silver stain and fluorescent detection. Biotechniques 1996, 20, 882-889. – 14. Nakamura Y., Leppert M., O'Connel P., Wolfe R., Holm T., Culver M., Martin C., Fujimoto E., Hoff M., Kumlin E., White R. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 1987, 235, 1616-1622. – 15. Nazneen A., Rahman Z., Riazuddin S. Allelic and genotypic frequency distribution at three STR loci (vWA, D3S1358 and D16S539) in Pakistani population. Int. J. Agricult. Biol. 2002, 1560-8530 /

04-1-8-11. – 16. Puers C., Hammond H. A., Jin L., Caskey C. T., Schumm J. W. Identification of repeat sequences heterogeneity at the polymorphic short tandem repeat locus HUMTH01 [AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am. J. Hum. Genet. 1993, 53, 953-958. – 17. Rebecca R., Sensabaugh G. Analysis of Genetic Markers in Forensic DNA samples using the Polymerase Chain Reaction. Reprinted from Analytical Chemistry 1991, 63. – 18. Sambrook J., Fritsch E.F., Maniatis T. Chapter 14: In vitro amplification of DNA by the polymerase chain reaction. In: Molecular cloning: A laboratory Manual, Second edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989. – 19. Signer E., Kuenzle C. C., Thomann P. E., Hubscher U. DNA fingerprinting: Improved DNA extraction from small blood samples. Nucleic Acids Res. 1988, 16, 7738. – 20. Tahir M.A., Balraj E., Luke L., Gilbert T., Hamby J.E., Amjad M. DNA typing of samples for polymarker, DQA1, and nine STR loci from a human body exhumed after 27 years, J. Forensic Sci. 2000, 45 (4), 902-907.

Corresponding Author / Adres pierwszego autora:
M. Saqib Shahzad
House No. C/114
Ward No 12
Muzaffar Garh, Punjab
Pakistan 34200
E-mail: saqib1330@yahoo.com.