INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
1560–8530/2002/04–1–8–11
http://www.ijab.org

Allelic and Genotypic Frequency Distribution at three STR Loci (vWA, D3S1358 and D16S539) in Pakistani Population

NAZNEEN ASLAM†, ZIA-UR-RAHMAN‡ AND SHEIKH RIAZ-UD-DIN
National Institute for Biotechnology & Genetic Engineering (NIBGE), PO Box 577, Jhang Road, Faisalabad–Pakistan
†*Corresponding author: E-mail: nazneenaslam66@hotmail.com*
‡*Centre of Excellence in Molecular Biology (CEMB), Punjab University Lahore–Pakistan*

ABSTRACT

CEMB's forensic DNA typing project is directed towards the introduction of DNA typing technique in Pakistan's criminal justice system. This technique can provide decisive scientific evidence to establish a suspect's guilt and help acquit wrongfully accused persons. Current project is a part of CEMB's population genetic survey of random Pakistani population using DNA typing technique. In this population genetic study, a sample of 100 unrelated individuals was analyzed for three hyper-variable STR loci (vWA D3S1358 and D16S539). The PCR products were fractionated on a denaturing polyacrylamide gel and visualized by silver staining. The genotypic patterns of all the samples were determined and then the data were analyzed for population genetic parameters (allele frequencies, genotypic frequencies, heterozygosity and fit to the Hardy-Weinberg equilibrium). For three STR loci, deviation was not observed from Hardy-Weinberg expectation. No microvarient was found in any of these loci. This allelic and genotypic frequency data may be useful in forensic casework with other loci of CODIS system.

Key Words: Allelic; STR loci; Population; Pakistan

INTRODUCTION

The advent of DNA typing technologies has generated considerable excitement in the forensic community. In fact, the human genome is rich in polymorphic sequences that lie outside of amino acid coding regions (Jeffreys, 1979). The sequence variation between individuals can be exploited ultimately to the point of absolute identification. Since 1985, DNA typing of biological material has become one of the most powerful tools for personal identification in forensic medicine and in criminal investigations (Jefferys *et al.,* 1985; Du Chesne *et al.,* 1993). Forensic DNA typing examines the properties of non-coding loci spread over the entire human genome. Because of non-coding loci not being expressed, DNA typing seldom reveals information about an individual except for his mere identity or his relatedness to other individuals. Forensic DNA analysis is based on the randomly repeated sequences being known. Knowledge about frequency of a certain STR allele in population enables the forensic biologist to calculate how often an allele combination appears in a given population. So a survey of the population must be made to measure the prevalence or frequency of occurrence of a certain allele at a specific locus.

Population data for STR loci has been reported by a number of scientists around the world (Edwards *et al.,* 1992; Puers *et al.,* 1993; Hammond *et al.,* 1994; Bevers & Creacy 1995; Lins *et al.,* 1996). The FBI Laboratory sponsored a community wide forensic science effort to establish the core STR loci for the national DNA index (CODIS System). In 1997, the core loci for the national system were agreed upon by participating laboratories. The 13 core STR loci are CSFIPO, FGA, THO1, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 and D21S11. Population data for these 13 STR loci have been generated and used in several European countries. Three STR loci, vWA, D16S539 and D3S1358 proved to be highly polymorphic and were decided to be amplified and genotyped in Pakistani population too.

MATERIALS AND METHODS

Sample collection, DNA extraction and estimation. Hundred blood samples were collected in EDTA containing tubes at four major hospitals of Lahore. Genomic DNA was extracted from blood according to the modified FBI protocol reported by Singer *et al.* (1988) and Grimberg *et al.* (1989). DNA concentration was estimated by using Hitachi Spectronic 2000 spectrophotometer.

Amplification. DNA samples were amplified for three STR (Short tandem repeats) loci vWA, D3S1358 and D16S539 separately using primers, PCR buffer from Research Genetics, USA (Map Pairs), $MgCl_2$, dNTP'S, from Sigma, Taq DNA polymerase (locally made in CEMB enzyme lab). Amplifications were carried out in thin-walled MicroAmp tubes (0.5 mL) in thermal cycler PTC-100 TM (MJ Research, Inc.). The PCR thermal cycling profile was as follows: Initial denaturation at 95°C for 5 min, denaturation at 94°C for 30 sec, annealing at 59°C (for vWA and D3S1358) and 61°C (for D16S539) for 30 sec, extension at 72°C for 45 sec, for 30 cycles and

followed by 72°C final extension for 5 min. 20-25 ng DNA was used for amplification/reaction (20 µL). Amplicons were confirmed on 2% agarose gel.

Allele scoring. The amplified products were typed by electrophoretic separation on 8% denaturing vertical polyacrylamide gel (7 M Urea, 0.5 X TBE, 0.35 mm thick, 500 µL sodium thiosulphate and 25-50 µL TEMED) for all the three STR loci. The gels were allowed to polymerize for about one hour at room temperature. Four µL of amplified DNA sample were mixed with four µL of STR 2X loading solution (10m MNaOH, 95% formamide 0.05% bromophenol blue, 0.05% xylene cyanol FF). The sample was loaded into the gel wells (23 wells comb). Before loading on the gel, the amplified products were denatured at 95ºC for two minutes, chilled on ice and then loaded within 20 min. The gel electrophoresis was carried out for 2 h at 50-60 W, in 0.5 X TBE buffer. The gels were stained by using Promega's DNA Silver Staining System (Anonymous, 1993). Reference allelic ladders were prepared by pooling amplified products, concentrated by using Ultrafree-MC 30, 000 NMWL, running on 8% PAGE with *Hinf 1 ϕx174*. Gel, a counter program, (Schaffer & Sederoff, 1981) was used to determine the size of each ladder band. Then alleles scoring of amplified products were carried out relative to reference allelic ladder by visual comparison.

Statistical analyses. Allelic and genotypic frequencies of the three STR loci were calculated from the genotypic data in the sample set by gene count method (Hayes *et al.,* 1995). Chi square test was applied to check the population for Hardy Weinberg Equilibrium (Hartl, 1988).

RESULTS AND DISCUSSION

CEMB's forensic DNA typing project has been designed to carry out genetic analysis of Pakistani sub-populations. Pakistani population is ethnically diverse and consists of many sub-populations with different genetic structures. Lewontin and Hartle (1991) have advocated the study of population subgroups. Major ethnic groups of Pakistani population such as Araeen, Rajput (Rahman, 2001) and Jaat (Ali, 2000) have been genetically analyzed at CEMB. The findings of their research made the basis of the current study. In the present study, blood samples were collected randomly from the population of Lahore.

Genomic DNA was extracted from whole blood samples using phenol chloroform extraction method. This protocol gives good yield of DNA from small blood samples (Singer *et al.,* 1988; Grimberg *et al.,* 1989). This DNA extraction procedure has been found suitable for both the RFLP based (Aldridge *et al.,* 1984) and PCR based (de Pancorbo *et al.,* 1998) DNA typing techniques.

Three microsatellites vWA, D3S1358 and D16S539 were selected for genotypic analysis. The PCR amplification technology is ideally suited for the analysis of forensic DNA

Table I. Observed allele frequencies distribution at three STR loci in Pakistani population

Allele	STR LOCI		
	VWA(11-21) Frequency (N=100)	D3S1358(12-19) Frequency (N=100)	D16S539(5,8-15) Frequency (N=100)
5			0.125
6			
7			
8			0.185
9			0.215
10			0.210
11	0.000		0.190
12	0.080	0.000	0.075
13	0.100	0.050	0.000
14	0.115	0.345	0.000
15	0.210	0.290	0.000
16	0.240	0.235	
17	0.115	0.080	
18	0.140	0.000	
19	0.000	0.000	
20	0.000		
21	0.000		

Table II. Genotypic Frequency Data for STR Locus vWA

Genotype	Frequency				
	Allele 1	Allele 2	Al 1* A 2	Possible Genotypes	Chi-square value
11	0.000	0.000	0.000	0	0.000
12, 12	0.080	0.080	0.006	1	29.703*
12, 13	0.080	0.100	0.008	2	1.600
12, 14	0.080	0.115	0.009	2	0.014
12, 15	0.080	0.210	0.017	2	0.550
12, 16	0.080	0.240	0.019	2	3.840
12, 17	0.080	0.115	0.009	2	0.383
12, 18	0.080	0.140	0.011	2	0.686
13, 13	0.100	0.100	0.010	1	9.000
13, 14	0.100	0.115	0.012	2	0.735
13, 15	0.100	0.210	0.021	2	2.438
13, 16	0.100	0.240	0.024	2	0.008
13, 17	0.100	0.115	0.012	2	0.039
13, 18	0.100	0.140	0.014	2	0.014
14, 14	0.115	0.080	0.009	1	0.007
14, 15	0.115	0.210	0.024	2	0.143
14, 16	0.115	0.240	0.028	2	7.607*
14, 17	0.115	0.115	0.013	2	0.157
14, 18	0.115	0.140	0.016	2	3.220
15, 15	0.210	0.210	0.044	1	0.038
15, 16	0.210	0.240	0.050	2	1.524
15, 17	0.210	0.115	0.024	2	0.975
15, 18	0.210	0.140	0.029	2	0.132
16, 16	0.240	0.240	0.058	1	3.934
16, 17	0.240	0.115	0.028	2	0.419
16, 18	0.240	0.140	0.034	2	2.726
17, 17	0.115	0.115	0.013	1	0.079
17, 18	0.115	0.140	0.016	2	0.189
18, 18	0.140	0.140	0.020	1	0.001
19	0.000	0.000	0.000	0	0.000
20	0.000	0.000	0.000	0	0.000
21	0.000	0.000	0.000	0	0.000

* Significant and all others are non-significant

Table III. Genotypic frequency Data for STR Locus D3S1358

	Frequency			Possible genotype	Chi-Square Value
Genotype	Allele 1	Allele 2	A1 * A 2		
12	0.000	0.000	0.000	0	0.000
13, 13	0.050	0.050	0.003	1	0.250
13, 14	0.050	0.345	0.017	2	1.885
13, 15	0.050	0.290	0.015	2	0.003
13, 16	0.050	0.235	0.012	2	2.350
13, 17	0.050	0.080	0.004	2	0.050
14, 14	0.345	0.345	0.119	1	0.068
14, 15	0.345	0.290	0.100	2	0.447
14, 16	0.345	0.235	0.081	2	1.096
14, 17	0.345	0.080	0.028	2	0.042
15, 15	0.290	0.290	0.084	1	0.020
15, 16	0.290	0.235	0.068	2	0.029
15, 17	0.290	0.080	0.023	2	0.580
16, 16	0.235	0.235	0.055	1	2.190
16, 17	0.235	0.080	0.019	2	0.015
17, 17	0.080	0.080	0.006	1	0.203
18	0.000	0.000	0.000	0	0.000
19	0.000	0.000	0.000	0	0.000

All values are non-significant

samples in that it is sensitive, rapid and not limited by the quality of DNA as needed by RFLP method.

The standardized masters mix components and optimized PCR conditions have been used. Mg^{2+} concentrations range from 1.5 mM to 3.0 mM for these loci. The PCR profile differs in annealing temperature for these loci i.e. 59ºC for vWA and D3S1358 while 61ºC for D16S539. The confirmation of amplification reaction and estimation of PCR product concentration was carried out by 2% agarose gel electrophoresis.

Table IV. Genotypic frequency Data for STR Locus D16S539

	Frequency				
Genotype	**Allele 1**	**Allele 2**	**A1*A2**	**Possible genotype**	**Chi-square Value**
5,5	0.125	0.125	0.016	1	0.203
5,8	0.125	0.185	0.023	2	0.571
5,9	0.125	0.215	0.027	2	0.491
5,10	0.125	0.210	0.026	2	0.298
5,11	0.125	0.190	0.024	2	0.329
5,12	0.125	0.075	0.009	2	0.675
8,8	0.185	0.185	0.034	1	0.052
8,9	0.185	0.215	0.040	2	0.115
8,10	0.185	0.210	0.039	2	0.640
8,11	0.185	0.190	0.035	2	0.000
8,12	0.185	0.075	0.014	2	0.541
9,9	0.185	0.215	0.046	1	1.223
9,10	0.185	0.210	0.045	2	1.799
9,11	0.185	0.190	0.041	2	0.084
9,12	0.185	0.075	0.016	2	1.535
10,10	0.210	0.210	0.044	1	0.038
10,11	0.210	0.190	0.040	2	0.511
10,12	0.210	0.075	0.016	2	0.007
11,11	0.190	0.190	0.036	1	0.718
11,12	0.190	0.075	0.014	2	0.254
12,12	0.075	0.075	0.006	1	0.340
13	0.000	0.000	0.000	0	0.000
14	0.000	0.000	0.000	0	0.000
15	0.000	0.000	0.000	0	0.000

All values are non-significant

Table V. Heterozygosity at three STR Loci in Pakistani Population.

Locus	Expected Heterozygosity	Observed Heterozygosity	Chi-square Value
VWA	83.99	82	0.04714
D3S1358	73.28	71	0.07093
D16S539	81.80	82	0.00048

All values are non-significant

Three STR Loci were amplified separately for each locus rather than multiplex system. Multiplex system has now been used for DNA typing frequently (Hammond & Caskey, 1992; Lins *et al.***,** 1996). But in the present study, multiplex amplification was not possible due to different PCR conditions and close proximity of the alleles. Each locus was amplified separately and resolved on 8% denaturing polyacrylamide gel followed by silver staining. de Pancorbo *et al.* (1998) have used denaturing polyacrylamide gel and found the technique useful for the resolution of STR loci because the resolution capacity of 4% polyacrylamide gel is one base pair.

Band stuttering is the common problem with STR loci. Stutter bands (sometime shadow bands) were amplified with the three STR primer pairs. The amplification of stutter bands was also observed at vWA STR locus (Weber & May, 1989; Sprecher *et al.,* 1996). Band stuttering is common in dinucleotide repeats and is produced due to slippage mechanism of the polymerase during amplification (Luty *et al.,* 1990).

All the three loci were polymorphic, without any microvarient. At each locus most of the reported alleles were found in sampled population, neither the variant nor alleles other than already known were observed. For locus vWA reported alleles ranging from 11 to 22 repeats (126-170 bp) (Anonymous, 1993). They also reported a variant allele 15.2 (144 bp) however no such allele was observed in sampled population, whereas alleles with 12 to 18 repeats (130-154 bp) were found. For STR locus D3S1358, alleles with 12 to 19 repeats (114–142 bp) were reported and alleles with 13-17 repeats (118–134 bp) were found in our sampled population. At locus D16S539, alleles with 5, 8-15 repeats (141 bp-173 pb) have been reported out of these alleles with 5 and 8-12 repeats (141-161 bp) were observed in our sampled population.

For the three loci, statistical analysis was also performed (Hartl, 1988). Observed and expected heterozygosity values were calculated (Table V). Chi square test was applied to check the population for Hardy Weinberg Equilibrium. Chi square values were insignificant indicating that all three STR loci met Hardy Weinberg Equilibrium. These values clearly indicated that in our population inbreeding effect has decreased at these three STR loci. The distribution of the allelic frequencies and

genotypic frequencies for the loci is given in Tables I, II, III and IV). Statistically, no departure from Hardy Weinberg Equilibrium was observed. The trends in selection at these loci could not be determined because this data has been reported first time in Pakistan. However, some of the genotypic frequencies showed significant behavior reflecting that selection was not favoring these genotypes (Table II).

This study will lay the foundation for further studies in population genetics to check the trends of selection in our population. Moreover, this study is one part of national DNA database for 13 CODIS loci. A police crime lab to calculate the likelihood of a suspect being the actual criminal can use the present data along with other CODIS loci.

ACKNOWLEDGEMENT

The authors are thankful to all the doctors of KEMC, Jinnah hospital, Ganga Ram Hospital, General Hospital and volunteers who helped to complete this manuscript.

REFERENCES

Aldridge, J., L. Kunkel, G. Bruns, U. Tantaravahi, M. Lalande, T. Brewster, E. Moreau, M. Wilson, W. Bromley, T. Roderick and S.A. Latt, 1984. A strategy to reveal high frequency RFLPs along the human X- chromosome. *American J. Hum. Genet.,* 36: 546–64.

Ali, S., 2000. Population genetic analysis of Jaat Sub population of Lahore employing DNA typing technique. *Thesis* National Centre of Excellence in Molecular Biology, Lahore.

Anonymous, 1993. DNA silver staining system. *Technical Manual,* Promega Corporation.

Bever, R.A. and S. Creacy, 1995. Validation and utilization of commercially available STR multiplexes for parentage analysis. Proc. 5th Intl. Symp. Human Identification, pp: 61–3.

de Pancorbo, M..M., A. Castro, I. Fernandez and A. Garcia-Orad, 1998. Population genetics and forensic applications using multiplex PCR (CSF1PO, TPOX, and TH01) loci in the Basque Country. *J. Forensic Sci.,* 43: 1181–7

Du Chesne, A., S. Rand and B. Brinkmann, 1993. Stain examination with DNA technology-a retrospecific analysis. *Archiv Fur Kriminologie*, 192: 87.

Edwards, A., H.A. Hammond, L. Jin, C.T. Caskey and R. Chakraborty, 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics,* 12: 241–53.

Grimberg, J., S. Nawoschik, L. Belluscio, R. McKee, A. Turck and A. Eisenberg, 1989. A simple and efficient non-organic procedure for the isolation of genomic DNA from blood. *Nucl. Acids Res.,* 17: 8390.

Hammond, H. and C.T. Caskey, 1992. Personal identification via short tandem repeats. Proc. 3rd Intl. Symp. Human Identification, Promega Corporation.

Hammond, H.A., L. Jin, Y. Zhong, C.T. Caskey and R. Chakraborty, 1994. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *American J. Hum. Genet.,* 55: 175–89.

Hartl, D.L., 1988. *A Primer of Population Genetics,* 2nd Ed., pp: 23–8. Sinauer Associates, Inc. Publishers.

Hayes, J. M., B. Budowle and M. Freund, 1995. Arabian population data on the PCR-based loci: HLA-DQA1, LDLR, GYPA, HBGG, D7S8, Gc, and D1S80. *J. Forensic Sci*., 40: 888–91.

Jeffreys, A. J. 1979. DNA sequence variation G-gammar. A. Gamma, delta and Beta-globin gene of Mann. *Cell,* 18: 1–10.

Jeffreys,A.J., V. Wilson and S.L. Thein, 1985. Individual specific fingerprint of Human DNA. *Nature,* 316: 76–9.

Lewontin, R.C. and D.L. Hartl, 1991. Population genetics in forensic DNA typing. *Science,* 254: 1745–50.

Lins, A.M., C.J. Sprecher, C. Puers and J.W. Schumm, 1996. Multiplex sets for the amplification of polymorphic short tandem repeat loci silver stain and fluorescence detection. *BioTechniques,* 20: 882–9.

Luty, J.A., Z. Guo, H.F. Willard, D.H. Ledbetter, S. Ledbetter and M. Litt, 1990. Five polymorphic microsatellites VNTRs on the human X chromosome. *American J. Hum. Genet.,* 46: 776–83.

Puers, C., H.A. Hammond, L. Jin, C.T. Caskey and J.W. Schumm, 1993. Identification of repeat sequence 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. *American J. Hum. Genet.,* 53: 953–8.

Rahman. Z., T. Afroze and B.S. Weir, 2001. DNA typing results from two urban populations of Pakistan. *J. Forensic Sci.,* 46: 111–5.

Schaffer H.E. and R.R. Sederoff, 1981. Improved Estimation of DNA Fragment Lengths from Agarose gels. *Anal. Biochem.,* 115: 113 version 16/11/85.

Singer, E., C.C. Kuenzle, P.E. Thomann and U. Hubscher, 1988. DNA fingerprinting: Improved DNA extraction from small blood samples. *Nucl. Acids Res.,* 16: 7738.

Sprecher, C.J., C. Puers, A.M. Lins and J.W. Schumm, 1996. General approach to analysis of polymorphic short tandem repeat loci. *BioTechniques,* 20: 266–76.

Weber, J.L. and P.E. May, 1989. Abundant class of human polymorphism which can be typed using the polmerase chain reaction. *American J. Hum. Genet.,* 44: 388–96.

(Received 04 November 2001; Accepted 13 December 2001)