Genetic variation of informative short tandem repeat (STR) loci in an Iranian population

Reyhaneh Lahmi, Sadeq Vallian*

Department of Biology, Division of Genetics, Faculty of Science, University of Isfahan, Isfahan, I.R. Iran

Abstract
In the present study, genotyping of six short tandem repeat (STR) loci including CSF1PO, D16S539, F13A01, F13B, LPL and HPRTB was performed on genomic DNA from 127 unrelated individuals from the Iranian province of Isfahan. The results indicated that the allele and genotype distributions were in accordance with Hardy-Weinberg expectations. The observed heterozygosity (*Ho*), expected heterozygosity (H_e) as well as forensic and paternity indices including power of discrimination (*PD*) and exclusion (*PE*), polymorphism information content (*PIC*), typical paternity index ($PI_{typical}$) and probability of paternity (*W*) were determined for the examined STR alleles. In addition, genetic diversity index (*GD*) and population parameter (θ) were calculated for the six loci. The combined power of discrimination ($P_{dcombined}$) and combined probability of exclusion ($PE_{typical}$) were 0.9999998 and 0.999856 over the six loci, respectively. Together, the examined STR loci in this study have proven a relatively high genetic variation in Iranian population. The data could be used for construction of a forensic genetic database for Iranian population.
Keywords: Short tandem repeat (STRs); Population data; Paternity testing; Forensic science; Iranian population

INTRODUCTION

Tandem repeat DNA sequences, which are widespread throughout the human genome, show sufficient variability among individuals in a population. They have become important in several fields including mapping studies, disease diagnosis, and human identity testing (Edwards *et al*., 1991; Sutherland and Richards, 1995). Notably, short tandem repeats (STRs), which contain 2-6 base pair (bp) regular repeat units have gained importance in forensic analysis of biological specimens, because they have a high degree of polymorphism in human populations. Moreover, they are amenable to analysis by the polymerase chain reaction (PCR), which allows for the analysis of low concentrated DNA samples (Schneider, 1997). Besides, the intermediate number of alleles (5-15 common alleles per locus) keeps the locus size range small enough for multi-locus PCR amplification (Schumm, 1996). In view of the above advantages, the STR loci considered as popular and useful markers for forensic sciences and paternity analysis (Butler, 2006; Krawczak and Schnidtke, 2001).

Identification of the best markers for these applications is complicated by the fact that the degree of allele frequencies for each STR marker is specific in each population. This is based on the previous studies performed on STRs in the population of interest (Ülker *et al*., 2004). This shows the necessity of analysis of known STR markers in each population (Dong Hoon *et al*., 1998; Edwards *et al*., 1992).

A decade ago, United State Federal Bureau of Investigation (FBI) introduced 13 genetic markers and funded the FBI Laboratory's Combined DNA Index System (CODIS) data base. The CODIS genetic markers are frequently used in most populations for criminal justice and forensic investigations (Butler, 2006).

Previous studies on the application of the STR markers in Iranian population, indicated the presence of several alleles with high polymorphism (Shepard and Herrera, 2006; Vallian and Moeini, 2006). In the present study six different STR markers including

*Correspondence to: **Sadeq Vallian,** Ph.D.
Tel/Fax: +98 311 793 2456
E-mail: svallian@biol.ui.ac.ir

CSF1PO, D16S539, F13A01, F13B, LPL and HPRTB were investigated. The data shown in this study, in combination with the previous data on the STR loci in Iranian population, could facilitate the initiation of construction of the first Iranian STR genetic database.

MATERIALS AND METHODS

DNA samples: Whole blood was collected from a total of 127 unrelated healthy donors from the Iranian province of Isfahan referred to Isfahan medical genetics center, Isfahan. The total genomic DNA was extracted from the leukocytes using standard salting out procedure (Miller *et al*., 1988).

Genotyping: The DNA samples were used to genotype CSF1PO, D16S539, F13A01, F13B, LPL and HPRTB STR loci using PCR technique. The PCR reactions were carried out in a 25 µl total volume containing 0.2 mM of dNTP and *Taq*-DNA polymerase 5 units (Cinagen, IR Iran) in a PCR buffer (10 X) (50 mM Tris-HCl pH 8.8, 15 mM $(NH4)_2SO_4$, 1.5 mM $MgCl_2$, 0.01% (v/v) gelatin). The reactions were performed in a Gradient Thermal Cycler (Eppendorf, Germany). The amplification cycles were as: one cycle at 94°C for 1 min; 20 cycles (94°C for 1 min, annealing 60°C for 1 min, 70°C for 1.5 min), followed by one cycle 70ºC for 5 min. The sequences for each set of primers are shown in Table 1.

The PCR products (1.5 µl) were checked in a vertical 15 cm × 25 cm × 0.75 mm polyacrylamide gel electrophoresis (PAGE). The separated alleles were silver stained and visualized using a gel documentation system (Biometra, Germany). The size of the alleles was determined by comparison with the M3 DNA size marker (Elchrom Scientifics AG, Switzerland) using Gene-tools software.

Statistical analysis: The allele frequencies were calculated based on the number of the detected alleles for specific locus. For data analysis, the Arlequin software package version 1.3 was used to (i) assess Hardy-Weinberg equilibrium (HWE) using Fisher's exact test, and (ii) to determine population parameter *(θ)* and gene diversity indices (*GD*) (Excoffier, 2005). Several forensic parameters were examined including power of discrimination (*PD*), polymorphic information content (*PIC*) and matching probability (*pM*) using PowerStats program version 1.2 (Tereba, 1999). Also, paternity indices such as power of exclusion (*PE*), typical paternity index ($PI_{typical}$) and probability of paternity (*W*) were calculated using PowerStats program version 1.2 (Tereba, 1999).

RESULTS

Genotyping and allele frequency of the six previously mentioned short tandem repeat (STR) loci including CSF1PO, D16S539, F13A01, F13B, LPL and HPRTB were investigated. In Table 1, the characteristics of the

Table 1. Characteristics of the STR loci and primers used in this study.

STR Locus	Chromosomal location	Allele number	Product size (bp)	Primer sequence (forward and reverse) (5´-3´)
LPL	8p22	5	113-129	5'-CTGACCAAGGATAGTGGGATATAG-3' 5'-GGTAACTGAGCGAGACTGTGTCT-3'
F13B	1q31-q32.1	6	169-189	5'-TGAGGTGGTGTACTACCATA-3' 5'-GATCATGCCATTGCACTCTA-3'
HPRTB	Xq26	7	263-287	5'-ATGCCACAGATAATACACATCCCC-3' 5'-CTCTCCAGAATAGTTAGATGTAGG-3
D16S539	16q22-24	6	280-300	5'-GATCCCAAGCTCTTCCTCTT-3' 5'-ACGTTTGTGTGTGCATCTGT-3
CSF1PO	5q33.3-34	7	299-323	5'-AACCTGAGTCTGCCAAGGACTAGC-3' 5'-TTCCACACACCACTGGCCATCTTC-3'
F13A01	6p24-p25	5	315-331	5'-GAGGTTGCACTCCAGCCTTT-3' 5'-ATGCCATGCAGATTAGAAA-3'

[1]Alleles determined in the present study. Abbreviations are as follows: LPL, Lipoprotein lipase gene; F13B, coagulation factor XIII b subunit gene; HPRTB, hypoxanthine phosphoribosyltransferase (HPRT) gene; CSF1PO, CSF-1 receptor (FMS) gene; F13A01, coagulation factor XIII a subunit gene.

Table 2. Allele frequencies of six STR loci in the population of Isfahan.

Allele	LPL		F13B		HPRTB		D16S539		CSF1PO		F13A01	
	Frq(%)	length(bp) n= 245	Frq(%)	length(bp) n= 245	Frq(%)	length(bp) n= 245	Frq(%)	length(bp) n= 245	Frq(%)	length(bp) n= 245	Frq(%)	length(bp) n= 245
6			15.8	(169)								
7			14.9	(173)	6.6	(263)						
8			17.1	(177)	6.1	(267)			5.3	(299)		
9	20.2	(113)	21.5	(181)	23.2	(271)	16.2	(280)	6.1	(303)		
10	20.6	(117)	25.9	(185)	8.8	(275)	10.1	(284)	18,0	(307)		
11	19.3	(121)	4.8	(187)	27.6	(279)	24.6	(288)	25.4	(311)		
12	19.3	(125)			23.7	(283)	31.1	(292)	28.1	(315)	25.9	(315)
13	20.6	(129)			3.9	(287)	12.7	(296)	12.3	(319)	14.9	(319)
14							5.3	(300)	4.8	(323)	24.1	(323)
15											19.3	(327)
16											15.8	(331)
H_0	0.78		0.78		0.77		0.74		0.76		0.72	
H_e	0.81		0.84		0.79		0.79		0.80		0.79	
P	0.74		0.91		0.30		0.30		0.56		0.44	
GD	0.82		0.83		0.78		0.78		0.81		0.76	
θ	2.69		2.77		2.61		2.61		2.57		2.52	

H_o: observed heterozygosity; H_e: expected heterozygosity; P: HWE, Fisher's exact test P values; GD: gene diversity; θ: population parameter.

examined STR loci were illustrated. As presented in Table 2, in a sample of 127 unrelated Iranian individuals, genotyping of the loci indicated the presence of 7, 6, 5, 6, 5 and 7 different alleles for CSF1PO, D16S539, F13A01, F13B, LPL and HPRTB loci, respectively. Based on the position of the primers on the genomic sequence of the loci, the number of repeats and the size of the alleles, the core four-nucleotide repeat was calculated with the aid of Gene-tools software using the M3 DNA size marker as standard (see Material and Methods). For each locus the alleles found were evenly spaced and resided between 315-331 base pair for F13A01, 299-323 base pair for CSF1PO, 280-300 base pair for D16S539, 263-287 base pair for HPRTB, 169-189 base pair for F13B and 113-129 base pair for LPL (see Table 2). Figure 1 represents a typical genotyping analysis for F13A01 STR using a 12% gel. The observed allele frequencies of the six STR loci are shown in Table 2.

Moreover, the expected (H_e) and observed (H_o) heterozygosity, *GD*, θ and *P* values were calculated and presented in Table 2. There were no deviations from Hardy-Weinberg expectations detected in any of the six loci analyzed in this population ($P>0.05$). The heterozygosity of six STR loci screened in this study ranged from 0.72 to 0.78.

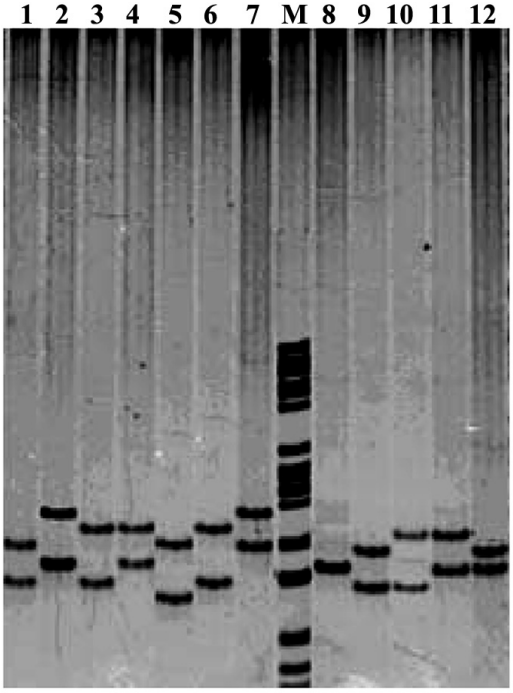

Figure 1. Genotyping of F13A01 STR marker. A typical polyacrylamide gel electrophoresis of F13A01 STR marker, which was genotyped for 12 individuals, is presented. Genotyping was performed on a sample of total genomic DNA from each individual using PCR with primers for F13A01 genomic sequence. The PCR products were analyzed on a 12% polyacrylamide gel and visualized by silver staining. M represents the M3 DNA size ladder. See Materials and Methods for details.

Table 3. Analysis of forensic and paternity values for CSF1PO, D16S539, F13A01, F13B, LPL and HPRTB in Isfahan.

	F13A01	CSF1PO	D16S539	HPRTB	F13B	LPL
Forensic						
Matching probability (*pM*)	0.078	0.081	0.087	0.078	0.074	0.076
Power of discrimination (*PD*)	0.922	0.919	0.913	0.922	0.926	0.924
Polymorphic information content (*PIC*)	0.76	0.77	0.76	0.77	0.78	0.77
Paternity						
Power of exclusion (*PE*)	0.473	0.533	0.502	0.548	0.580	0.564
Typical paternity index (PI_I)	1.84	2.11	1.97	2.19	2.38	2.28
Probability of paternity (*W*)	0.622	0.706	0.664	0.736	0.766	0.746

Forensic parameters including matching probability and power of discrimination as well as polymorphic information content (*PIC*) for the STR loci CSF1PO, D16S539, F13A01, F13B, LPL and HPRTB were shown in Table 3 all the six STR loci showed a high degree of *PIC* values (above 0.5). The paternity indices such as power of exclusion, typical paternity index and probability of paternity were also determined using the same software (see Table 3). Moreover, the data showed that all six STR loci had essentially similar power of exclusion. However, F13B and LPL showed higher typical paternity index and probability of paternity compared to CSF1PO, D16S539, F13A01 and HPRTB.

DISCUSSION

The high level of heterozygosity (ranging from 0.72 to 0.78) for six STR loci which were genotyped in this study, indicated that these loci could be used in determination of individual identification. Data obtained for all the markers examined indicated that F13B locus was potentially very promising marker for identity testing. Moreover, the fairly even distribution of alleles at the LPL locus within the population examined makes it particularly informative for forensic and paternity testing purposes.

Analysis of forensic parameters including matching probability and power of discrimination as well as polymorphic information content (*PIC*) for the STR loci indicated that all the STRs showed relatively similar values for forensic and individual identification purposes (see Table 3). The high degree of *PIC* values (above 0.5), which was shown by all the six STR loci could suggest that the loci were very informative. Moreover, the data showed that all six STR loci had essentially similar power of exclusion. However, F13B and LPL showed higher typical paternity index and probability of paternity compared to CSF1PO, D16S539, F13A01 and HPRTB (see Table 3).

These data were also compared with those reported in a different study in Iranian population and other populations as well (Shepard and Herrera, 2006). Our obtained *PD*, *PIC* and *PE* values for CSF1PO and D16S539 were very similar to those reported in Iranian population (Shepard and Herrera, 2006). In addition, these markers showed a high level of heterozygosity in both studies. Furthermore, comparison of H_O, *PIC*, *PD* and $PI_{typical}$ values which were calculated for CSF1PO and D16S539 in this study, with two Indian populations indicated that the loci were similar in sense of polymorphism (Eaaswarkhanth *et al*., 2007; Hima Bindu *et al*., 2007). However, *pM* seems to be higher in Indian subpopulations (0.2) than that of Iranians (0.07). It shows that the number of individuals with the same STR pattern is higher in those subpopulations. This makes them improper for forensic applications by reducing their validity. Our calculated indices for CSF1PO and D16S539 were also compared with those of two Chinese subpopulations (Zhu *et al*., 2007 and 2008). The data showed that H_O, *PD* and *PIC* were almost similar, although *PE* was a little higher in those subpopulations (0.7). Finally, we examined our data for five loci (all except HPRTB) with those of Argentine population (Martinez *et al*., 2003). D16S539 and F13A01 seemed to have the highest level of heterozygosity in Argentine population. *PE* and *pM* values were found to be generally similar with those of Iranians. However, *PIC* was slightly lower

with the range from 0.5 to 0.7 in Argentine population. Therefore, our obtained values for forensic and paternity indices appear to be more similar with those of Asian populations which are useful to enrich Iranian STR database.

In conclusion, these loci can accurately distinguish between two unrelated people since the discriminating power of them was very high ($P_{dcombined}$ =0.9999998 and *combined PE*=0.999856). These results strongly support the application of this set of genetic markers for personal identity testing in Iranian population.

References

Butler JM (2006). Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci*. 51: 253-276.

Dong Hoon L, Han J, Lee W, Rho HM (1998). Quadruplex amplification of polymorphic STR loci in a Korean population. *Int J Legal Med*. 111: 320-322.

Eaaswarkhanth M, Roy S, Haque I (2007). Allele frequency distribution for 15 autosomal STR loci in two Muslim population of Tamilnadu, India. *Int J Legal Med*. 9: 332-335.

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

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*. 14: 241-253.

Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis, *Evol Bioinform*. 1: 47-50.
(Available: http://cmpg.unibe.ch/software/arlequin3.1)

Hima Bindu G, Trivedi R, Kashyap VK (2007). Allele frequency distribution based on 17 STR marker in three major Dravidin linguistic populations of Andhra Pradesh, India. *Forensic Sci Int*. 170: 76-85.

Krawczak M, Schnidtke J (2001). DNA typing to establish relationships. *Electrophoresis* 17: 626-639.

Martinez G, Vazquez E, Schaller C, Quevedo (2003). Genetic data on 11 STRs in an Argentine northeast population. *Forensic Sci Int*. 133: 254-255.

Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 16: 12-15.

Schneider PM (1997). Basic issues in DNA typing. *Forensic Sci Int*. 88: 17-32.

Schumm W (1996). New Approaches to DNA Fingerprint Analysis. *Gene Print*. 58: 12-18.

Shepard E, Herrera R (2006). Iranian STR variation at the fringes of biogeographical demarcation. *Forensic Sci Int*. 158: 140-148.

Sutherland G, Richards R (1995). Simple tandem DNA repeats and human genetic disease. *Hum Mol Genet*. 4: 523-537.

Tereba A (1999). Tools for analysis of population statistics. *Profiles in DNA*. 2: 14-16. (Available: http://www.promega.com/geneticidtools/ powerstats)

Ülker U, Ülker M, Cuneyte M, Kesici T, Meneves S (2004). Short tandem repeat (STR) polymorphisms in Turkish population. *J Genet*. 83: 6-9.

Vallian S, Moeini H (2006). Genotyping of Five Polymorphic STR Loci in Iranian Province of Isfahan. *J Sci IR*. 17: 113-117.

Zhu J, Shen C, Ma Y, He Y, Zhao J, Li X, Liu Y (2007). Genetic polymorphisms of 15 STR in Chinese salar ethnic minority group. *Forensic Sci Int*. 173: 210-213.

Zhu B, Yan J, Shen C, Li Y, Li T, Yu X, Xiong X, Mu H, Huang Y, Deng Y (2008). Population genetic analysis of 15 STR loci of Chinese Tu ethnic minority group. *Forensic Sci Int*. 174: 255-258.