

STR polymorphisms in Sri Lanka: evaluation of forensic utility in identification of individuals and parentage testing

Aresha Manamperi¹, Chanditha Hapuarachchi^{1,2}, Nilmini Silva Gunawardene¹, Anura Bandara¹, Damsiri Dayanath¹ and Wimaladharma Abeyewickreme^{1,2}

(Index words: Population genetics, short tandem repeats, alleles, DNA fingerprinting, paternity testing, genetic database)

Abstract

Objectives This preliminary study was carried out to determine the allele frequencies and forensic efficiency parameters for the short tandem repeat loci CSF1PO, TPOX, THO1, D16S539, D7S820, D13S317, vWA, FESFPS and F13B in a test sample population of Sri Lankans.

Design Test samples were obtained from 305 non-related individuals originating from all 9 provinces of Sri Lanka. DNA was extracted from whole blood using chelex-100 and amplified by PCR using the GenePrint STR kit and silver stained. Final DNA profiles were analysed for forensic efficiency parameters and paternity indices using PowerStats version 12. Possible divergence from Hardy-Weinberg Equilibrium was tested using the chi-square test and exact test.

Results All common alleles in the allelic ladders were found in the test sample studied. PIC values >0.5 for all 9 STR loci indicate this STR system to be informative and useful for identification purposes. The D13S317, vWA and D7S820 loci were found to be the most polymorphic markers of the system studied.

Conclusions No deviations from Hardy-Weinberg Equilibrium were found for any of the loci examined. The results indicate that the 9 STR loci system described here is suitable for estimating DNA profile frequencies in human identification and forensic and parentage testing for legal purposes among Sri Lankans.

Introduction

The application of DNA fingerprinting in laboratory medicine is increasing, with uses in paternity/maternity testing, human identification in forensic/criminal case work, validation of genetic disease diagnosis and genetic mapping [1, 2, 3, 4, 5]. In the clinical laboratory, most DNA typing applications have legal and ethical implications and, therefore, there is a particular need for tests that have high reliability and diagnostic efficiency.

The most widely used methodology for DNA typing is polymerase chain reaction (PCR) analysis of various short tandem repeat (STR) loci [6]. STRs are highly polymorphic loci containing tandemly repeated sequences

of 2-7 bp in length, and are abundant in the human genome. Using PCR, STRs can be typed with a high degree of specificity and sensitivity, and genotyping based on these regions is widely used in human identification and parentage testing [7].

The true power of DNA testing for the identification of individuals and estimation of allele frequencies for parentage testing lies in polymorphism at the individual loci and the number of loci tested. In order to estimate multilocus genotype frequencies in this manner, both the alleles at each locus and the alleles among loci must be inherited independently. These conditions of independence must hold true for a population in order to apply individual allele frequencies to produce a composite estimate. Therefore, it is important that any laboratory using STR profiles for parentage testing or forensic practice validates the STR markers by analysing these statistical parameters using a database. STR population genetic databases have been constructed worldwide as a basis for calculating probabilities of a DNA match in the relevant population.

Objectives

The objective of this study was to determine allele frequencies and forensic efficiency parameters of the STRs: CSF1PO, TPOX, THO1, D16S539, D7S820, D13S317, vWA, FESFPS and F13B in a sample population of Sri Lankans with a view to establishing a national DNA databank. The characteristics of the above STRs are given in Table 1. These STR markers were chosen on the basis of their heterozygosity and validated use as forensic markers in other laboratories, and availability of population data for comparison [7]. A previous study has shown that the STR markers described above are highly polymorphic in a randomly selected cohort of Sri Lankans [8]. The importance of a national DNA databank is to assist Sri Lankan investigators in the identification of individuals. A genetic profile of a validated battery of STR markers can be effective for individualisation in a variety of identity testing situations. In this study, the estimation of forensic

¹Molecular Medicine Unit and ²Department of Parasitology, Faculty of Medicine, University of Kelaniya, Thalagolla Road, Ragama, Sri Lanka.

Correspondence: AM, email <areshamanamperi@gmail.com>. Received 30 October 2008 and accepted 7 April 2009. Competing interests: none declared.

efficiency parameters is based primarily on the National Research Council (NRC) Report (1996), "The Evaluation of Forensic DNA Evidence", whose recommendations for such computations are becoming standard [7].

Methods

Test samples: venous blood samples were taken, after obtaining informed written consent, from 305 unrelated Sri Lankan individuals during the period 2002-2006. The number of individuals typed using each of the nine STR loci comprised of: CSF1PO=303; TPOX=305; THO1=305; D16S539=293; D7S820=299; D13S317=299; vWA=297; F13B=299; FESFPS= 299. Test samples consisted of subjects requesting paternity tests and criminal/forensic specimens that were referred by the magistrate/high courts and forensic officials respectively, to the Molecular Medicine Unit, Faculty of Medicine, University of Kelaniya, as well as randomly selected individuals. STR profile data of related individuals (in disputed paternity cases and relations of victims in the case of criminal case work) were excluded from the database for the purpose of computing allele frequencies and forensic efficiency parameters, in order to remove bias. The study subjects originated from all 9 provinces of Sri Lanka and comprised of individuals from all ethnic backgrounds; 80% were Sinhalese.

DNA extraction and STR analysis: DNA extractions were performed from whole blood using chelex-100 [9]. PCR amplification was performed in a 9700 GeneAmp PCR system (Perkin Elmer, USA) using locus specific primers and reagents provided in the GenePrint STR system kit (Promega Corp. Madison, WI, USA) according to the

technical manual [10]. The 9 loci were amplified in two triplex PCR systems representing the CTT (CSF1PO, TPOX and THO1) loci, the SILVER III (D16S539, D7S820 and D13S317) loci and three independent assays for vWA, FESFPS and F13B loci. For each batch of PCR analysis, positive controls supplied by Promega Corp., USA (K562) and negative controls (no DNA) were employed. The amplified products were separated in 6% denaturing polyacrylamide gels and visualised by silver staining [11]. The alleles were assigned by comparison with appropriate allelic ladders according to the standard nomenclature [12]. The final DNA profiles were analysed using PowerStats software version 12 (Promega Corp., Madison, WI, USA) for allele frequencies [13], homozygosity, heterozygosity, forensic efficiency parameters [matching probability (MP) [14], power of discrimination (PD) [15], and polymorphic information content (PIC) [16]] and paternity indices [power of exclusion (PE) [14] and typical paternity index (TPI) [14]]. Possible divergence from Hardy-Weinberg Equilibrium (HWE) was tested using the chi-square test [17]. When the chi-square test was applied, alleles were pooled so that each allelic event could be under the "rule of five" [18]. For the exact test, alleles were also pooled with adjacent alleles in all loci, so that no alleles would have less than 5 events [19]. To estimate the probability of the exact test, we used the PEPI404X software (Sagebrush Press, 225, 10th Avenue, Salt Lake City, Utah 84103-2529, USA; www.sagebrushpress.com). A computerised integrated database with all the genotypic and allelic information was created for this data.

Approval for the study was obtained from the Ethics Committee of the Faculty of Medicine, University of Kelaniya, Sri Lanka.

Table 1. **Characteristics of the STR loci: CSF1PO, TPOX, THO1, D16S539, D7S820, D13S317, vWA, FESFPS and F13B**

<i>STR</i>	<i>Locus definition</i>	<i>Location</i>	<i>Repeat element</i>
CSF1PO	Proto-oncogene for the CSF-1 receptor	5q33.3-34	AGAT
TPOX	Human thyroid peroxidase gene	2p25.1-pter	AATG
THO1	Human tyrosine hydroxylase gene	11p15.5	AATG
D16S539	Not applicable	16q24-qter	AGAT
D7S820	Not applicable	7q11.21-22	AGAT
D13S317	Not applicable	13q22-q31	AGAT
vWA	Human von Willebrand factor gene	12p12-pter	AGAT
FESFPS	Human c-fes/fps proto-oncogene	15q25-qter	AAAT
F13B	Coagulation factor VIII subunit b gene	1q31-q32.1	AAAT

Table 2: Allele frequencies, forensic efficiency parameters, paternity statistics, and other statistical parameters for Hardy-Weinberg equilibrium for the nine STR loci in the given Sri Lankan sample population

Allele	CSF1PO	TPOX	TH01	D16S539	D7S820	D13S317	FESFPS	vWA	F13B
5			(0.000)	(0.000)					
6	(0.000)**	(0.000)	0.228	(0.000)	(0.000)				0.104
7	(0.000)	(0.000)	0.120	(0.000)	0.040	0.008	(0.000)		0.017
8	(0.000)	0.315	0.151	0.094	0.234	0.229	0.002		0.216
9	0.003	0.133	0.343	0.167	0.057	0.102	0.002		0.236
9.3			0.133**						
10	0.188	0.100	0.020	0.068	0.216	0.089	0.191		0.423
11	0.285	0.433	0.007	0.287	0.231	0.227	0.344	(0.000)**	0.005
12	0.411	0.016		0.232	0.191	0.229	0.304		(0.000)**
13	0.076	0.003		0.138	0.027	0.085	0.147	0.003	
14	0.008			0.014	0.005	0.030	0.010	0.148	
15	0.002			(0.000)		(0.000)		0.109	
16								0.209	
17								0.271	
18								0.158	
19								0.089	
20								0.012	
21								(0.000)**	
MP	0.134	0.154	0.084	0.068	0.071	0.062	0.122	0.065	0.139
MP for 9 loci					3.67x10 ⁻¹⁰				
MP expressed as 1 in..... for each locus	7.4	6.5	11.9	14.7	14.0	16.2	8.2	15.4	7.2
PD	0.866	0.846	0.916	0.932	0.929	0.938	0.878	0.935	0.861
PD for 9 loci (1-MP _{Combined})					0.999999				
PIC	0.660	0.630	0.740	0.780	0.770	0.790	0.680	0.790	0.660
PE	0.428	0.377	0.522	0.565	0.648	0.668	0.464	0.692	0.480
TPI	1.66	1.50	2.06	2.29	2.88	3.05	1.80	3.30	1.87
TPI for 9 loci					1,146.1				
Probability of paternity for 9 loci [PI/(PI+1)]					0.999				
Total no. of individuals	303	305	305	293	299	299	299	297	299
Heterozygous individuals	212	203	231	229	247	250	216	253	219
Observed Heterozygosity	0.700	0.666	0.757	0.782	0.826	0.836	0.722	0.852	0.732
Observed Homozygosity	0.300	0.334	0.243	0.218	0.174	0.164	0.278	0.148	0.268
Expected Homozygosity	0.290	0.318	0.229	0.201	0.190	0.173	0.264	0.171	0.287
X ² Test (P value after allele pooling)	0.876	0.727	0.792	0.535	0.999	0.997	0.795	0.897	0.836
Exact Test (P value after allele pooling)	0.976	0.929	0.762	0.780	0.941	0.958	0.839	0.947	0.823
Total Alleles	606	610	610	586	598	598	598	594	598

MP: matching probability, PD: power of discrimination, PIC: polymorphic information content, PE: power of exclusion, TPI: typical paternity index. Alleles are identified by number of repeats. Alleles not detected in the sample population are given in parenthesis and rare alleles observed in the human population are denoted by **.

Table 3. Most common genotypes observed for the nine STR loci in the Sri Lankan test sample population

<i>CSF1PO</i>	<i>TPOX</i>	<i>THO1</i>	<i>D16S539</i>	<i>D7S820</i>	<i>D13S317</i>	<i>FESFPS</i>	<i>vWA</i>	<i>F13B</i>
11,12 (23%)	8,11 (27%)	6,9 (16%)	11,12 (14%)	8,11/8,12 (11%)	11,12/8,11 (12%)	11,12 (19%)	16,17/17,18 (13%)	8,10 (21%)
12,12 (18%)	11,11 (18%)	9,9 (12%)	11,11/12,13 (9%)	10,11 (10%)	8,12 (10%)	10,11 (16%)	14,17 (9%)	9,10 (18%)
10,11/10,12 (13%)	8,8/9,11 (12%)	6,8 (9%)	11,13 (6%)	8,10 (9%)	9,12 (5%)	11,13 (12%)	16,18 (7%)	10,10 (17%)

Results

The allele frequencies of the nine STRs obtained with this system in 305 unrelated Sri Lankan subjects are given in Table 2. As a rule of thumb, the higher the number of alleles at a locus, the higher the potential number of heterozygotes, and thus the more powerful this system can be for resolving small differences in the distribution of allele frequencies in populations [13]. In our study, up to several alleles in the allelic ladders were observed for each locus, but only few were relatively common; other alleles were rare, thus contributing little to heterozygosity. Further, none of the alleles in the 9 STR loci exceeded 50% frequency, reflecting the usefulness and validity of these loci in calculating paternity indices and discriminating individuals [13]. However, relatively high allele frequencies were shown by locus CSF1PO for allele 12 (41.1%), TPOX locus for allele 11 (43.3%), and F13B locus for allele 10 (42.3%), indicating relatively low power of discrimination for these loci. The most frequent alleles were 12 and 11 of CSF1PO, 11 and 8 of TPOX, 9 and 6 of THO1, 11 and 12 of D16S539, 8 and 11 of D7S820, 8 and 12 of D13S317, 11 and 12 of FESFPS, 17 and 16 of vWA, and 10 and 9 for F13B. Of the rare alleles in the human population (Table 1), only 9.3 of THO1 was present in this study population and exhibited an allele frequency of 13.3%. The most common genotypes for the 9 STR loci shown by individuals of this study are given in Table 3. A comparison was also made (Genetic Identity Population Data, www.promega.com) between the allele frequencies observed in this study, with those of other major populations that have been studied (Caucasian American, African American, Chinese, Hispanic American), and a broad agreement was observed for the most common alleles (data not shown).

The common forensic efficiency parameters useful for application of the nine STR markers to forensic and paternity tests are given in Table 2: matching probability (MP), power of discrimination (PD), polymorphic information content (PIC), paternity statistics [power of exclusion (PE), typical paternity index (TPI), and allele information (homozygotes, heterozygotes). The high level of heterozygosity observed for this 9 STR system is an indication that the Sri Lankan population has a high level of genetic variation, and this could be successfully utilised in discriminating between individuals (Table 2).

For forensic purposes, HWE conformity is desired in the estimation of statistical parameters, as the calculations rely heavily on the condition of random genetic assortment. The estimated allelic distribution for all 9 STR loci studied here did not deviate from the HWE proportions based on chi-square or exact tests after the grouping of alleles, so that all the grouped alleles had more than 5 events (Table 2).

Discussion

In this study, we examined a panel of 9 STR loci in a sample population of Sri Lankans and estimated the allele and genotype frequencies for each, in order to evaluate the usefulness of this system in cases of forensic investigation and paternity testing. Statistical parameters such as the observed and expected homozygosity values, the polymorphic information content (PIC), the power of discrimination (PD), and the matching probability (MP), which are important in applying DNA typing methods for identity testing demonstrated the usefulness of this system in forensic practice.

PIC values >0.5 for all 9 STR loci (range 0.630 - 0.790) indicate that the analysed system is informative and useful for identification purposes. The typical paternity index (TPI) which is assigned to a particular locus shows values TPI>1, indicative of relatedness (range 1.5-3.05). Further, this can be expressed as the combined TPI for all loci of the system, and for this study the combined TPI was calculated to be 1,146.1. Another parameter used to evaluate the strength of a locus to exclude falsely accused individuals is the power of paternity exclusion (PE). It represents the percentage of individuals in the relevant population who would not share the same DNA profile presented in a paternity case [14]. The higher the PE value, the more non-fathers are excluded. The high power of exclusion (PE) values and the typical paternity index (TPI) values of the 9 STR loci indicate the diagnostic potential for paternity testing, and these results are indicative of practically conclusive results using this method. The combined matching probability (MP) of the 9 STR system was 3.67×10^{-10} , indicating that the system has a much stronger inter-individual discrimination power. The 9 loci showed a combined power of discrimination (PD) of

$1-3.67 \times 10^{-10}$ (0.999999). Practically, this degree of MP and PD means that no other individual with the same profile for all 9 loci could exist in Sri Lanka, as these two parameters express the probability of two random DNA profiles matching at the loci tested: MP expressed as 1 in ... individuals for 9 STR loci is ~ 3 billion (2,777,777,777) vs a population of over 18 million in the country.

The D13S317, vWA and D7S820 loci are the most polymorphic and informative loci of the system studied, with the highest observed PE, PD, PIC and TPI values. Further, as Table 1 shows, the other loci in the silver III system and THO1 locus of the CTT system can also be particularly useful in discriminating individuals. These observations are in agreement with previous reports which have confirmed these STR loci to be particularly useful in paternity testing and the individualisation of biological evidence, among the routinely used STRs [20]. The STR loci TPOX and F13B recorded the lowest values for forensic efficiency parameters and paternity indices, indicating that these markers are of lesser value in medical forensic practice.

Conclusions

Results of this preliminary study indicate that the alleles in this 9 STR loci system are in conformity with the Hardy-Weinberg Equilibrium. The loci are also highly polymorphic and discriminating, permitting the 9 STR loci system to be used in human identification and parentage testing for legal purposes in Sri Lanka. The vWA and Silver III systems have been confirmed as particularly useful in paternity testing and individualisation of biological evidence. The present database, therefore, validates the use of the 9 STR loci for forensic applications.

Acknowledgements

We thank Ms. Nayomi Raymond of the Molecular Medicine Unit, for secretarial assistance. Financial support from the International Atomic Energy Agency (IAEA) TC Project SRL 06/028 is acknowledged.

References

- Schmitt C, Benecke M. Five cases of forensic short tandem repeat DNA typing. *Electrophoresis* 1997; **18**: 690-4.
- Kimpton CP, Oldroyd NJ, Watson SK, *et al.* Validation of highly discriminating multiplex short tandem repeat amplification system for individual identification. *Electrophoresis* 1996; **17**: 1283-93.
- Mao L, Schoenberg MP, Scicchitano M, *et al.* Molecular detection of primary bladder cancer by microsatellite analysis. *Science* 1996; **271**: 659-62.
- Scharf SJ, Smith AG, Hansen JA, *et al.* Quantitative determination of bone marrow transplant engraftment using fluorescent polymerase chain reaction primers for human identity markers. *Blood* 1995; **85**: 1954-63.
- Gyapay G, Morissette J, Vignal A, *et al.* The 1993-94 Genethon human genetic linkage map. *Nature Genetics* 1994; **7**: 246-9.
- Alford RL, Hammond HA, Coto I, Caskey CT. Rapid and efficient resolution of parentage by amplification of short tandem repeats. *American Journal of Human Genetics* 1994; **55**: 190-5.
- Hammond HA, Jin L, Zhong Y, *et al.* Evaluation of 13 short tandem repeat loci for use in personal identification application. *American Journal of Human Genetics* 1994; **55**: 175-89.
- Fernandopulle N. Development of DNA markers for the identification of individuals in Sri Lankan human population and its applications in forensic case work. Ph.D. thesis, University of Colombo, Sri Lanka, 2001.
- Walsh PS, Matzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 1991; **10**: 506-13.
- Technical manual of the GenePrintR, PowerPlex™ 2.1 System. Promega Corporation, Madison, 2001.
- Bassam BJ, Caetano-Anolles G, Gresshoff PM. Fast and sensitive silver staining of DNA in polyacrylamide gels. *Analytical Biochemistry* 1991; **196**: 80-3.
- Bar W, Brinkmann B, Budowle B, *et al.* DNA recommendations: further report of the DNA Commission of the ISFH regarding the use of short tandem repeat system. *International Journal of Legal Medicine* 1997; **110**: 175-6.
- Weir BS. In: Genetics Data Analysis II. Sinauer Associates Inc., Sunderland, MA, USA, 1996.
- Brenner C, Morris JW. Paternity index calculations in single locus hypervariable DNA probes validation and other studies. Proceedings from International Symposium of Human Identification. Promega Corp, 1989: 23-53.
- Jones DA. Blood samples: probabilities of discrimination. *Journal of the Forensic Science Society* 1972; **12**: 355-9.
- Botstein D, White RL, Skolnick M, Davies RW. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics* 1980; **32**: 314-31.
- Edwards A, Hammond HA, Jin L, *et al.* Genetic Variation of five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 1992; **12**: 241-53.
- Cochran WG. Some methods for strengthening the common chi 2 test. *Biometrics* 1954; **10**: 417-51.
- National Research Council II Report. The evaluation of forensic DNA evidence, statistical issues – alleles with low frequency. In: National Research Council ed, National Academy Press, Washington DC, 1996; 148.
- He P, Nata M, Kanetake J. Polymorphism of short tandem repeat (STR) loci PLA2, D3S2459, D8S315 (kw38), cyp19, D3S1359, FGA and D8S1132 in the Japanese and Chinese populations. *Tohoku Journal of Experimental Medicine* 1997; **183**: 251-62.