

Evaluation and comparison of 15 short tandem repeat loci of south and west Indian population for use in personal identification applications

Prabakaran Mathiyazhagan^{1*}, Thangaraju Palanimuthu¹, Agasthi Padmanathan²

¹Department of Microbiology, SRM Institute of Science and Technology, Chengalpattu, Tamil Nadu, India.

²Department of Mathematics, Tagore Engineering College, Rathinamangalam, Chengalpattu, Tamil Nadu, India.

ARTICLE INFO

Article history:

Received on: March 07, 2023

Accepted on: July 11, 2023

Available online: October 25, 2023

Key words:

Short tandem repeat,
FORSTAT,
Hardy–Weinberg,
Indian population.

ABSTRACT

The objective of this research was to conduct an analysis of the frequency of alleles as well as forensically relevant parameters such as match probability, power of exclusion, polymorphic information content, observed, and expected heterozygosity, homozygosity, paternity index, and discrimination capacity (DC). In all, 209 unrelated people from South Indian groups and 201 people from West Indian populations were examined, and their frequency and statistical attributes were compared. In this research, 15 short tandem repeats (STRs) loci were examined utilizing a multiplex PowerPlex 16 System kit. These loci were CSF1PO, TPOX, D8S1179, TH01, Penta E, D18S51, D21S11, D3S1358, FGA, Vwa, Penta D, D16S539, D7S820, D13S317, and D5S818. With the help of the FORSTAT program, 15 STRs frequencies and forensic significance parameters were computed. Within the population, there was no evidence of any departure from the Hardy–Weinberg equilibrium. The maximal allele frequency at TPOX was found to be 0.4976 in the South Indian population, 0.4925 in the West Indian population, and 0.5040 in the combined Indian population, respectively. Similar to the previous statement, the minimum allele frequency was seen at several loci throughout all three investigations. The Penta E and TPOX were found to be the locus with the highest and the lowest degree of polymorphism in the south and combined Indian population, respectively; whereas the locus D18S51 and TPOX recorded the highest and lowest degree of polymorphism in the West Indian population.

1. INTRODUCTION

India has the greatest polygenetic population of any secular nation. Known to be a cultural melting pot, India is home to people of many different languages, cultural backgrounds, and religious traditions [1]. There are 4693 recognized demographic subgroups in India. They are comprised 2205 recognized communities, 589 subgroups, and 1900 separate administrative regions [2]. South India consists of the Indian regions of Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, and Telangana, as well as the Union territories of Lakshadweep and Puducherry. As of April of 2020, there were 253,051,953 people living in 635,780 km² or 245,480 sq miles [3]. The western part of India includes the states of Maharashtra, Rajasthan, Gujarat, and Goa. As of April 2020, there were 173,343,821 people living in 508,032 km² or 196,152 sq miles [4].

Forensic scientists now commonly utilize DNA analysis based on short tandem repeats (STRs) to determine paternity and to examine

more complex criminal cases, such as those involving rape, murder, and mass rape [5]. Due to their high diversity and widespread distribution throughout the human genome, short tandem repeats (STRs) are useful genetic identifiers for identifying specific people. Multiplex polymerase chain reaction (PCR) has made it possible to amplify numerous loci at once, making STR typing a viable tool for human identification [6]. Autosomal STR loci are widely employed in forensics and paternity testing [7,8]. The odds of two unrelated people having identical STR profiles, or fingerprints, are “about one in 57 trillion,” as shown by the frequencies of the STR alleles [9]. Several nations now have their own forensic DNA databases. To prove the identification of missing individuals, victims, criminals, under trial, unknown dead persons, etc., the “DNA Technologies (Use and Application) Regulation Bill – 2019” was drafted in India. The Bill includes language creating a DNA Regulatory Board (DRB) charged with performing the duties and exercising the authority specified in the Bill. The national forensic DNA database for the purpose of identifying the aforementioned class of individuals will be maintained through the establishment of National and Regional DNA Data Banks, which are also provided for in the bill. This will help improve the scientific level and efficiency of DNA testing in the India [10]. The Indian government is now debating the bill, which is evidence that a comprehensive regional and state-level DNA database is necessary.

*Corresponding Author:

Prabakaran Mathiyazhagan,
SRM Institute of Science and Technology,
Chengalpattu, Tamil Nadu, India.
E-mail: prabakarmathi@yahoo.co.in

studied genetic polymorphism of 15 STR loci in 123 unrelated individuals of Madhya Pradesh population and evaluated autosomal STR diversity and forensic important parameters. They concluded that locus Penta E showed the highest power of discrimination (0.978) the highest polymorphism information content (0.900) and also suggested all loci are highly polymorphic, showed significant genetic diversity, and have the potential for forensic application carried out the research to investigate the genomic diversity and population structure in the Muslim community of Telangana using 23 autosomal STR in 184 randomly selected unrelated healthy individuals and concluded that locus SE33 showed 37 observed alleles, which is the highest number of observed alleles among all the studied loci and SE33 showed the highest number of observed alleles (37) among all the studied loci. The locus SE33 was the most polymorphic and discriminatory locus and may be useful for forensic application. Similar to these studies, lot of research had been done using STR focusing on specific community such as Muslim community of Telangana, Teli population of Maharashtra, Mina, Gujjar, and the admixed population of Rajasthan and Sikh population of central India. In this research, it is attempted to cover two regions of India (South and West) and also compared their genetic parameters and frequency that can be applied in forensic application. In India to prove the identification of missing individuals, victims, criminals, under trial, unknown dead persons, etc., the "DNA Technologies using STR has not been used as prevalent as in other countries. The objective of the research was to determine the allele frequencies and statistical parameters for medicolegal interest of 15 loci Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, Vwa, Penta D, CSF1PO, D16S539, D7S820, D13S317, and D5S818 in subjects with a large population of unrelated south and West Indian population. In this research, 410 buccal swab samples were collected from unrelated individuals of south and West Indian states after getting their informed consent forms. From the collected samples, DNA was extracted and stored in -21°C . The extracted DNA samples were quantified to analyze their DNA concentration by measuring the absorbance at 260 nm (A_{260}) in a spectrophotometer using a quartz cuvette. The quantified DNA was ready to be used for the DNA amplification. Next, amplification process was carried out using PP16 multiplex primer. The amplified PCR samples were engaged into DNA sequencing using ABI Applied BioSystem 3130 sequencer. Once the samples were sequenced, the data were analyzed with the use of GeneMapper software and, subsequently, the frequency and forensic significant parameters calculated using the FORSTAT software. The maximal allele frequency at TPOX was found to be 0.4976 in the South Indian population, 0.4925 in the West Indian population, and 0.5040 in the combined South and West Indian population, respectively. The sequence-based allelic frequencies for south and West Indian population groups studied here with STR would enable equipped laboratories to use the frequencies and forensic important parameters obtained from this study for relationship or forensic identification. This work complements other available sequence-based allelic frequency databases published by other studies carried out in South and West Indian states, but also fills a large data gap for the South, and West Indian populations. In addition, the similarities and differences have been compared between these two populations that would enable forensic and relationship testing industry to use these parameters confidently when analyzing live forensic cases are done using STR for any type of paternity, maternity, missing person analysis, etc. The key limitations from the previous studies were to collect the samples from different parts of the states in India and high expenses to analyze the more number of samples. The advantage of this study aside from being the new approach to compare the South and West

Indian population with 15 STR loci and it can also be used by the DNA analysts in South and West Indian laboratories for live forensic and relationship casework, as well as for further research outside.

2. MATERIALS AND METHODS

Buccal swab samples of 410 healthy unrelated individuals 209 (123 males and 86 females) and 201 (108 males and 102 females) between the age group of 17 and 23 from South and West Indian states were collected respectively after getting their informed consent [Figure 1]. The work was authorized by the Ethics Committee of SRM Institute of Science and Technology–Chennai. (1887/IEC/2020).

2.1. DNA Extraction and Quantification

Using a swab solution kit that was given by Promega, from Madison, United States, genomic DNA was extracted from the buccal swab.

2.2. Selection of STR Markers

PP16 STRs were chosen for this research due to their polymorphism. The PowerPlex 16 System is a multiplex STR system for use in DNA typing, including paternity testing, forensic DNA analysis, human identity testing, and strain identification. This system allows coamplification and three-color detection of 16 loci (fifteen STR loci and Amelogenin): Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, Vwa, Amelogenin, Penta D, CSF1PO, D16S539, D7S820, D13S317, and D5S818. One primer for each of the Penta E, D18S51, D21S11, TH01, and D3S1358 loci is labeled with fluorescein (FL); one primer for each of the FGA, TPOX, D8S1179, Vwa, and Amelogenin loci is labeled with carboxy-tetramethylrhodamine (TMR); and one primer for each of the Penta D, CSF1PO, D16S539, D7S820, D13S317, and D5S818 loci is labeled with 6-carboxy-4',5'-dichloro-2',7'-dimethoxy-fluorescein (JOE). All 16 loci are amplified simultaneously in a single tube and analyzed in a single injection. The alleles of the STR loci may be distinguished from one another based on the amount of copies of the repeat sequence that is included inside each STR locus. The discriminatory capacity tends to increase with the number of STR loci employed in the typing process. Since it is extremely unlikely that a single individual will share an identical STR profile with another individual drawn at random from the population, the likelihood that a single individual will have an identical STR profile with another individual drawn at random from the population is extremely low [11].

2.3. PCR Amplification

Amplification by multiplexed PCR was carried out with the use of the PowerPlex 16 System (Manufactured by Promega in Madison, United States). PCR was done in a 25ul volume with 2ng template DNA and the following amplification conditions: An initial incubation at 95°C for 11 min, 28 cycles of denaturation at 94°C for 1 min, annealing at 59°C for 1 min, and extension at 72°C for 1 min; and a final extension at 60°C for 45 min, as per manufacturer's instructions.

2.4. Control Samples

For positive control, PowerPlex® 16 System-supplied DNA was used, which produced interpretable results in each run throughout the process.

For negative control, nuclease-free water was used in each run throughout the process.

For quality control, internally processed DNA (a known DNA profile) was used in each run throughout the process.

2.5. STR Typing of PowerPlex-Amplified Samples

Using an ABI Prism 3130 Avant Genetic Analyzer, multicapillary electrophoresis of the amplification product was carried out (Applied BioSystems, Foster City, CA, and USA). The data were examined using the ABI Prism GeneMapper version 3.0 software (Applied BioSystems, and the relative peak area was automatically determined by the program to identify alleles by comparison with the allelic ladder that was given with the kit. The allele nomenclature was established in accordance with the guidelines provided by the International Association for Forensic Hemogenetics [12]. For the purpose of allele identification, the peak detection threshold was set to 50 RFUs. Every step was carried out in the manner prescribed by the laboratory's internal standards as well as the manual and procedure for the relevant kit controls.

2.6. Analysis of the Data

Forensic and population genetics metrics such as allele frequency, matching probability, polymorphism information content, expected and observed heterozygosity and homozygosity, power of exclusion, discriminatory capacity, and paternity index were determined using FORSTAT (Forensic Statistic Analysis Toolbox) FORSTAT web tool which implements the common GENEPOP format to evaluate forensic genetics statistics. The simplicity and ease-of-use are achieved by automation of data entry and calculations which reduce analysis time and labor inputs [13]. In FORSTAT, the analyzed sample profiles were entered into and required data such as frequency and forensic parameters automatically analyzed by the software and results obtained.

3. RESULTS

The allele frequencies calculated for South, West, and combined South and West Indian population are given in Tables 1-3, respectively.

Minimum allele frequency of 0.0024 was found in different alleles which are mentioned below:

D3S1358-20 TH01-5, D21S11-26, 29.2, 33, 34, and 36 D18S51-15.2 and 24 Penta E-15.4, 21 and 22 D13S317-15 D16S539-5 and 15, CSF1PO-15 D18S1179-9 TPOX-6,7, and 13 FGA-19.2, 20.2, 22.2, and 30.

Minimum allele frequency of 0.0025 was found in different alleles which are mentioned below:

D3S1358- 12 TH01-11, D21S11- 25 D18S51-13.2 and 14.2 Penta E-22 and 23 CSF1PO-8 Penta D- 3.2, 6, 7 and 17 Vwa- 11 D8S1179-9 and 18 TPOX-7 FGA-17, 17.2, 22.2, 25.2, 28, 29, and 30.

Minimum allele frequency of 0.0010 was found in different alleles which are mentioned below:

D3S1358- 20 TH01-5, 11, D21S11-29.2, 30.3 and 34.2 D18S51-24, D16S539-5,15 Penta D-3.2, 17 Vwa-12, D8S1179-18, TPOX-13, and FGA 17,19,2,25,2,30,2, and 33.2.

By comparing the alleles in the population groups, the following alleles 15.2, 15.4, 18.2, 19.2, 20.2, 29.2, 33.1, 34, 35, and 36 were observed in South Indian while they were not found in the West Indian population; whereas the alleles 3.2, 13.2, 14.2, 17.2, 20.3, and 25.2

were recorded in the West Indian population but were not observed in the South Indian population. On the other hand, the following alleles 10.3, 15.3, 16.2, 21.2, and 23.2 were found in the combined population but were not seen either in South or West Indian population.

A total of 157, 159, and 182 alleles were recorded in the South, West, and combined South and West Indian populations, respectively. Alleles were counted, and the corresponding numbers were 46, 42, and 54. In TPOX, the allele frequencies in the South Indian population varied from 0.0024 to 0.4976, those in the West Indian population from 0.0024 to 0.4925, and those in the combined Indian population from 0.0010 to 0.5040. TPOX locus had the highest frequency in all three sample populations. Allele frequencies varied from 6 (D7S820) to 18 (Penta E) in the South, 7 (TH01, D7S820, D16S539, and TPOX), to 19 (Penta E) in the West, and 8 (TH01, D13S317, D7S820, and TPOX) to 23 (FGA) in the South and West Indian population combined. The West Indians had the highest frequency of alleles at the FGA locus (19/42=45.2%), followed by the combined South and West Indian population also at the same FGA locus (23/54=42.6%), whereas the South Indians had at the Penta E locus (18/46 = 39.1%).

Analyses of the range and number of alleles at each locus for the South, West, and combined Indian population revealed that the maximum range at the same locus Penta E was 5-22, 10-25, and 5-25 and the number of alleles were 18 and 19 for South and West Indian population, respectively, whereas the FGA locus recorded the maximum of 23 alleles in the combined Indian population.

Most common allele (MCA) and least common allele (LCA) observed at different loci are given in Table 4.

The least common alleles occurred in many of the loci only one time.

The polymorphism values of STRs expressed by various statistical parameters are presented in Tables 5 for south, west, and combined Indian population.

According to the above data, the lowest match probability was 0.0444, 0.0710, and 0.0343 in PENTA E in the South, West, and combined Indian populations, respectively, while the corresponding highest match probability was 0.1906, 0.2005, and 0.1805 at TPOX. The highest values of discrimination capacity were found at the locus Penta E (0.9556, 0.9290, and 0.9659) and the lowest values at the locus TPOX (0.8094, 0.7995, and 0.8157) in South, West, and combined Indian populations, respectively.

The lowest polymorphic information content (PIC) values were 0.6039, 0.5935, and 0.6074 at locus TPOX in the South, West, and combined Indian populations, respectively, and the corresponding highest values were 0.8940 and 0.8916 at Penta E in the south and combined Indian populations, and 0.8516 at locus D18S51 in the west.

In the South, West, and combined Indian populations, the lowest expected heterozygosity (H_{exp}) values were 0.6503, 0.6475, and 0.6499, respectively, at the locus TPOX, and the maximum values of 0.9017 and 0.8988 at locus Penta E were recorded in South and combined Indian populations and 0.8658 in the locus D18S51 in the West Indian population. The minimum observed heterozygosity (H_{obs}) values were 0.6464, 0.6550, and 0.6510 in the South, West, and combined Indian populations, respectively; whereas the corresponding maximum observed heterozygosity values were 0.9071 and 0.8857 in the South and combined Indian populations and 0.8600 at FGA locus in the west. The minimum homozygosity values of 0.0929 and 0.1143 were found at locus Penta E in the South and combined

Table 1: Allele frequencies in South Indian population.

Alleles	D3S1358	TH01	D21S11	D18S51	Penta_E	D5S818	D13S317	D7S820	D16S539	CSF1PO	Penta_D	Vwa	D8S1179	TPOX	FGA
2.2											0.0048				
5		0.0024			0.0909				0.0024						
6		0.1889									0.0072			0.0024	
7		0.2751			0.1651						0.0048			0.0024	
8		0.1507			0.0096	0.0478	0.1268	0.0143	0.0215	0.0048	0.0215		0.0072	0.4976	
9		0.1722			0.0167	0.0454	0.0861	0.0550	0.1627	0.0167	0.2320		0.0024	0.1507	
9.3		0.1986													
10		0.0119													
11				0.0072	0.0885	0.0718	0.0598	0.3206	0.0933	0.2177	0.0957		0.0669	0.0454	
12				0.0143	0.0909	0.2368	0.3133	0.3038	0.3014	0.3301	0.1316		0.0694	0.2536	
13	0.0048			0.0981	0.2033	0.3421	0.2344	0.2655	0.2153	0.3517	0.2153		0.1316	0.0454	
14	0.0837			0.0981	0.0837	0.2368	0.1316	0.0407	0.1794	0.0694	0.2129		0.2608	0.0024	
15	0.3086			0.1411	0.0717	0.0143	0.0454		0.0215	0.0071	0.0598	0.0933	0.2440		
15.2				0.1722	0.0359	0.0048	0.0024		0.0024	0.0024	0.0143	0.1651	0.1698		
15.4				0.0024											
16	0.2871			0.1411	0.0431							0.2297	0.0431		
17	0.1962			0.1411	0.0478							0.2320	0.0048		
18	0.1124			0.0813	0.0239							0.1842			0.0119
18.2															0.0072
19	0.0048			0.0406	0.0143							0.0813			0.0526
19.2															0.0024
20	0.0024			0.0383	0.0072							0.0072			0.0885
20.2															0.0024
21				0.0096	0.0024							0.0072			0.1555
22				0.0096	0.0024										0.2009
22.2															0.0024
23				0.0239											0.1531
24				0.0024											0.1531
25															0.0933
26		0.0024													0.0454
27		0.0383													0.0287
28		0.1746													
29		0.1794													
29.2		0.0024													

(Contd...)

Table 1: (Continued).

Alleles	D3S1358	TH01	D21S11	D18S51	Penta_E	D5S818	D13S317	D7S820	D16S539	CSFIPO	Penta_D	Vwa	D8S1179	TPOX	FGA
30			0.2416												0.0024
30.2			0.0311												
31			0.0718												
31.2			0.0885												
32			0.0143												
32.2			0.1124												
33			0.0024												
33.1			0.0048												
33.2			0.0239												
34			0.0024												
35			0.0072												
36			0.0024												
Observed Alleles 46															
Range of Alleles	13-20	5-10	26-26	10-24	5-22	8-15	8-15	8-13	5-15	8-15	2.2-15	14-21	8-17	6-13	18-30
Recorded Alleles Each Locus	8	7	17	16	18	8	8	6	9	8	11	8	10	8	15
															Total 157

Table 2: Allele frequencies in West Indian population.

Alleles	D3S1358	TH01	D21S11	D18S51	Penta_E	D5S818	D13S317	D7S820	D16S539	CSFIPO	Penta_D	Vwa	D8S1179	TPOX	FGA
2.2											0.0199				
3.2											0.0025				
5					0.0423						0.0124				
6	0.2363										0.0025			0.0099	
7	0.2686				0.1219	0.0149		0.0199		0.0149	0.0025			0.0025	
8	0.1094				0.0223	0.0124	0.1219	0.1666	0.0149	0.0025	0.0199		0.0149	0.4925	
9	0.2189				0.0124	0.0622	0.1642	0.0970	0.2015	0.0199	0.2487		0.0025	0.0995	
9.3	0.1567														
10	0.0075			0.0049	0.0945	0.0448	0.0920	0.3035	0.1468	0.2413	0.1517		0.0821	0.0472	
11	0.0025			0.0049	0.0945	0.3433	0.2214	0.2562	0.2637	0.2786	0.1542	0.0025	0.0721	0.2463	
12	0.0025			0.0970	0.1866	0.3557	0.2413	0.1293	0.2264	0.3756	0.1666		0.1194	0.1019	
13	0.0075			0.1194	0.092	0.1567	0.0920	0.0274	0.1169	0.0622	0.1393	0.0049	0.2910		
13.2				0.0025											
14	0.1019			0.1716	0.0597	0.0099	0.0622		0.0298	0.0049	0.0721	0.1094	0.2214		
14.2				0.0025											
15	0.3159			0.1492	0.0771		0.0049				0.0049	0.0721	0.1318		

(Contid...)

Table 3: Allele frequencies in combined South and West Indian population.

Allele	D3S158	TH01	D21S11	D18S51	Penta_E	D5S818	D13S317	D7S820	DI6S539	CSF1PO	Penta_D	Vwa	D8S1179	TPOX	FGA
2.2											0.0101				
3.2											0.0010				
5		0.0010			0.0753				0.0010		0.0030				
6		0.2067									0.0061			0.0050	
7		0.2749			0.1201	0.0132		0.0091		0.0101	0.0030			0.0030	
8		0.1334			0.0142	0.0285	0.1374	0.0926	0.0162	0.0040	0.0193		0.0091	0.5040	
9		0.1578			0.0152	0.0478	0.1262	0.0814	0.1802	0.0264	0.2423		0.003	0.1171	
9.3		0.2169													
10		0.0081		0.0040	0.0845	0.0600	0.0794	0.3136	0.1272	0.227	0.1374		0.0763	0.0448	
11		0.0010		0.0081	0.1018	0.2942	0.2688	0.2790	0.2790	0.2953	0.1517	0.0020	0.0692	0.2545	
12				0.0926	0.1782	0.3553	0.2342	0.1894	0.2219	0.3594	0.1853	0.0010	0.1201	0.0702	
13	0.0071			0.1181	0.0794	0.1873	0.1038	0.0336	0.1466	0.0682	0.1680	0.0030	0.2698	0.0010	
13.2				0.0020											
14	0.0936			0.1537	0.0631	0.0112	0.0448		0.0264	0.0071	0.0590	0.1008	0.2474		
15	0.3146			0.1680	0.0631	0.0020	0.0030		0.0010	0.0020	0.0101	0.1232	0.1507		
16	0.2810			0.1323	0.0509				0.0020		0.0020	0.2454	0.0468		0.0010
17	0.1914			0.1446	0.0549				0.0010		0.0010	0.2464	0.0061		0.0020
17.2															
18	0.1048			0.0794	0.0386							0.1771	0.0010		0.0142
18.2															0.0040
19	0.0061			0.0427	0.0224							0.0814			0.0570
19.2															0.0010
20	0.0010			0.0264	0.0152							0.0142			0.0896
20.2	0.1249														0.0020
21				0.0132	0.0061							0.0050			0.1517
22				0.0122	0.0061										0.1914
22.2															0.0020
23				0.0020	0.0050										0.1517
24				0.0010											0.1527
25					0.0030										0.0916
25.2															0.0010
26															0.0498
27			0.0285												0.0264
28			0.1527												0.0030
29			0.2036												0.0020

(Contid...)

Table 3: (Continued).

Allele	D3S158	TH01	D21S11	D18S51	Penta_E	D5S818	D13S317	D7S820	D16S539	CSF1PO	Penta_D	Vwa	D8S1179	TPOX	FGA
29.2			0.0010												
30			0.2545												0.0030
30.2			0.0234												0.0010
30.3			0.0010												
31			0.0692												
31.2			0.0865												
32			0.0173												
32.2			0.1059												
33			0.0061												0.0010
33.2			0.0376												
34			0.0020												
34.2			0.0010												
35			0.0061												
36			0.0050												
Observed Alleles	13-20.2	5-11	27-36	10-24	5-25	7-15	8-15	7-13	5-15	7-15	2.2-17	11-21	8-18	6-13	17-30.2
Recorded Alleles Each Loci	9	8	17	16	19	9	8	7	9	9	15	11	11	8	22
Total	182														

Indian populations, respectively, while 0.1400 was found at locus FGA in the west. Among the South, West, and combined Indian populations, the highest homozygosity values were 0.3536, 0.3450, and 0.3489, respectively at the locus TPOX. The Power of Exclusion (PE) minimum values of 0.3649, 0.3769, and 0.3675 were recorded at locus TPOX in the South, West, and combined Indian populations, respectively, whereas the corresponding maximum values of 0.8107 and 0.7731 were noted at Penta E in the south and combined Indian populations, respectively, and in the west, it was 0.7161 at FGA locus. The paternity index (PI) minimum values of 1.5106, 1.5802, and 1.5058 were observed at locus TPOX in South, West, and combined Indian populations, respectively, while the maximum values of 6.3467 and 4.5763 were observed in Penta E in the South and combined Indian populations, respectively, while in the west, it was 4.2000 in Penta D locus. In both populations, no divergence from Hardy–Weinberg equilibrium was detected ($P > 0.05$).

Among tables presented in this study, the frequency tables of south and west (1 and 2) Indian frequencies are the most important ones as they have been highlighted the frequency of South and West Indian population separately which can be used to potential relationship analysis when the live samples analyzed from those states.

4. DISCUSSION

The accuracy of the findings is generally improved by gathering data from a larger sample size. Prior research revealed that calculating the probability using just 100–150 tested people per population [14]. For this reason, a total of 410 people from the South and West Indian populations were evaluated to determine allele frequency analyses and statistical parameters of forensic relevance.

4.1. Alleles and Allele Frequencies

In South, West, and combined Indian population, the number of actual observed alleles were 46, 42, and 54, respectively. Analyzed genetic diversity of autosomal STRs in 11 population of India and found 43 and 38 alleles in South and West Indian population, respectively, which is very close to our results [15]. In South, West, and combined Indian population, the total alleles recorded were 157, 159, and 182, respectively. The percentages of alleles were 75.1% (157/209) in South, 79.1% (159/201) in West Indian population. Carried out population genetic analyses of 22 autosomal STRs in Indian populations with 357 individuals from 11 states across India and recorded 275 alleles which worked out to 77% (275/357) [16] which corroborated to the values of this study. The maximum of 18 alleles in south and 19 in the west were observed in Penta E locus and 23 alleles were found in FGA in the combined Indian population indicated the most polymorphic nature of Penta E and FGA. Similarly, researched on genomic diversity in Maharashtra population with 20 autosomal markers highlighted that locus Penta E had the maximum number of alleles in the admixed and Teli population [17]. The maximum allele frequency of 0.4976, 0.4925, and 0.5040 were observed at TPOX in all three populations groups, respectively, whereas the minimum allele frequency was observed in several loci in South, West, and combined Indian population. Similarly, had studied the genetic variation of 15 autosomal microsatellite loci in Tamil population and observed the maximum frequency of 0.415 at TPOX which is more or less similar to our study. Allele 8 was recorded the maximum frequency in all the three populations [18]. Likewise, worked on genetic polymorphism of eleven STR loci in Rajput population of Delhi and stated that allele 8 of locus TPOX showed the highest frequency 0.425 [19]. In line with [17], our study showed

Table 4: MCA and LCA alleles in South, West, and combined Indian populations.

STR Loci	South Indian population		West Indian population		Combined South and West Indian population	
	MCA	LCA	MCA	LCA	MCA	LCA
D3S1358*	15 (129)	20 (1)	15 (127)	12 (1)	15 (309)	20 (1)
TH01	9.3 (83)	5 (1)	7 (108)	11 (1)	7 (270)	5, 11 (1)
D21S11*	30 (101)	26, 29.2, 33 34, 36 (1)	30 (105)	25 (1)	30 (250)	29.2, 30.3, 34.2 (1)
D18S51	15 (72)	15.2, 24 (1)	14 (69)	13.2 (1), 14.2	15 (165)	24 (1)
Penta_E*	12 (85)	15.4, 21, 22 (1)	12 (75)	22, 23 (1)	12 (175)	19.4 (2)
D5S818*	12 (143)	15 (2)	12 (143)	14 (4)	12 (349)	15 (2)
D13S317	11 (131)	15 (1)	12 (97)	15 (2)	11 (264)	15 (3)
D7S820*	10 (134)	8 (6)	10 (122)	7 (8)	10 (308)	7, 8 (9)
D16S539*	11 (126)	5, 15 (1)	11 (106)	8 (6)	11 (274)	5, 15 (1)
CSF1PO*	12 (147)	15 (1)	12 (151)	8 (1)	12 (353)	15 (2)
PentaD	12 (90)	2.2, 7 (2)	9 (100)	3.2, 6, 7, 17 (1)	9 (238)	3.2, 17 (1)
Vwa*	17 (97)	20, 21 (3)	17 (108)	11 (1)	17 (242)	12 (1)
D8S1179*	13 (109)	9 (1)	13 (117)	9, 18 (1)	13 (265)	18 (1)
TPOX*	8 (208)	6, 7, 13 (1)	8 (198)	7 (1)	8 (495)	13 (1)
FGA*	22 (84)	19.2, 20.2, 22.2, 30 (1)	22 (75)	17, 17.2, 22.2, 25.2, 28, 29, 30 (1)	22 (188)	17, 19.2, 25.2, 30.2, 33.2 (1)
	1739	30	1701	27	4145	25
Total	1769		1728		4170	

*The alleles in the 11 marked loci occurred maximum number of times in South, West, and in the combined Indian population (D16S539, CSF1PO, Vwa, D3S1358, D21S11, D8S1179, TPOX, Penta E, D5S818, D7S820, and FGA). The allele 12 is the most predominant in South, West, and combined populations occurring 465/1769 times (26.28%) in South and 467/1728 counts (27.02%) in West and 877/4170 counts (21.03%) in the combined Indian population.

that locus TPOX was the least polymorphic in all three population as 8, 7, and 8 alleles were noted in South, West, and combined Indian population, respectively.

4.2. Most and Least Common Alleles (MCA and LCA)

After an examination of the MCA as well as the LCA, it was discovered that the greatest number of alleles had been recorded in 11 of the 15 loci that had been investigated across all three populations. The allele 12 has the highest frequency throughout all three distinct populations, with a distribution that is as follows: 26.28% in the south, 27.02% in the west, and 21.03% in the combined Indian population. Similar findings were found by who conducted research on population genetic analysis for autosomal STR loci in the Sikh community of central India [20]. They gave a similar study, which found that allele 12 is the most prevalent allele in all three ethnicities.

4.3. Range of Alleles

Penta E was the only locus in which the maximum range was seen across all three populations, and the greatest numbers of alleles observed were 18 and 19 in the South and West Indian population, with the exception of locus FGA, which recorded the maximum of 23 alleles in the combined Indian population. Furthermore, at locus D7S820, all three populations shared a minimum range and number of alleles. Consistent with our findings, investigated the genetic makeup of 227 unrelated persons from Maharashtra and found that the highest number of alleles, 19, was located at the locus Penta E [17].

4.4. Statistical Parameters

Penta E had the lowest match probability and the highest discriminating capability across all three population types. In their study of 22 autosomal STRs in the Indian population, showed that Penta E has the

highest value of power of discrimination [16]. Penta E has the highest observed and expected heterozygosity in south and combined Indian populations, whereas the highest expected and observed heterozygosity were recorded in the D18S51 and FGA locus, respectively, in the West Indian population. According to a summary by who studied the genetic diversity of Gorkhas, the most polymorphic and discriminatory locus in the population was FGA [21]. According to the results of the above analysis, the locus Penta E was the most variable of the studied loci, displaying the highest values of discrimination capacity (0.9659), expected heterozygosity (0.8988), observed heterozygosity (0.8857), polymorphic information content (0.8916), power of exclusion (0.7731), and paternity index (4.5763) and the lowest values of match probability and homozygosity. However, locus TPOX recorded the highest match probability (0.1805) and homozygosity (0.3489), and the lowest polymorphic information content (0.6074), expected heterozygosity (0.6499), observed heterozygosity (0.6510), power of exclusion (0.3675), discrimination capacity (0.8157), and paternity index (1.5058) which indicated the less informativeness of the TPOX locus in these populations.

This study may be considered for STR-based analysis of Indian people in the future since the frequency and forensic parameters values are consistent with earlier STR research on the Indian population. If and when the Indian government passes "The DNA Technologies (Use and Application) Regulation Bill – 2019," these findings may serve as a foundation for further research and policy development. The key limitations from the previous studies were to collect the samples from different parts of the states in India and high expenses to analyze the more number of samples. The biggest challenge in this study was to convince the participants to provide the samples. In addition, the multiplex of PP16 system has a low chance of success rate for degraded DNA. The advantage of this study aside from being the new approach to compare the South and West Indian population with

Table 5: Statistical parameters of STR loci for South, West, and combined South and West Indian population.

LOCI	D3S1358	TH01	D21S11	D18S51	Penta_E	D5S818	DI3S317	D7S820	DI6S539	CSF1PO	Penta_D	Vwa	D8S1179	TPOX	FGA
MP (South)	0.1166	0.1089	0.0816	0.0597	0.0444	0.1660	0.0847	0.1148	0.0987	0.1497	0.0869	0.0961	0.0878	0.1906	0.0670
MP (West)	0.1350	0.1075	0.0825	0.0730	0.0710	0.1380	0.1195	0.1580	0.1170	0.1840	0.0955	0.1085	0.1055	0.2005	0.0830
MP (Combined)	0.1075	0.0899	0.0627	0.0454	0.0343	0.1253	0.0784	0.1054	0.0855	0.1370	0.0705	0.0817	0.0731	0.1805	0.0512
PIC (South)	0.7134	0.7427	0.8034	0.8554	0.8940	0.6666	0.7875	0.7423	0.7635	0.6721	0.7888	0.7695	0.7745	0.6039	0.8408
PIC (West)	0.7061	0.7445	0.8085	0.8516	0.8494	0.7014	0.7500	0.6668	0.7429	0.6368	0.7785	0.7669	0.7708	0.5935	0.8323
PIC (Combined)	0.7191	0.7520	0.8173	0.8635	0.8916	0.6954	0.7854	0.7310	0.7650	0.6725	0.7959	0.7826	0.7847	0.6074	0.8471
Hexp (South)	0.7529	0.7783	0.8247	0.8692	0.9017	0.7126	0.8130	0.7756	0.7942	0.7179	0.8135	0.7979	0.8010	0.6503	0.8570
Hexp (West)	0.7472	0.7790	0.8288	0.8658	0.8630	0.7417	0.7805	0.7174	0.7757	0.6933	0.8053	0.7965	0.7990	0.6475	0.8495
Hexp (Combined)	0.7579	0.7842	0.8334	0.8757	0.8988	0.7341	0.8162	0.7688	0.7939	0.7211	0.8211	0.8081	0.8111	0.6499	0.8610
Hobs (South)	0.7393	0.7643	0.8426	0.8857	0.9071	0.7214	0.8321	0.8250	0.7821	0.7429	0.8393	0.7785	0.7964	0.6464	0.8607
Hobs (West)	0.7100	0.7550	0.8400	0.8450	0.8500	0.7450	0.8150	0.7300	0.7600	0.7450	0.8500	0.8250	0.7900	0.6550	0.8600
Hobs (Combined)	0.7286	0.7592	0.8367	0.8673	0.8857	0.7265	0.8224	0.7816	0.7714	0.7408	0.8469	0.8020	0.7918	0.6510	0.8612
Homozygosity (South)	0.2607	0.2357	0.1571	0.1143	0.0929	0.2786	0.1678	0.1750	0.2179	0.2571	0.1607	0.2214	0.2036	0.3536	0.1393
Homozygosity (West)	0.2714	0.2408	0.1633	0.1327	0.1143	0.2735	0.1776	0.2184	0.2286	0.2592	0.1531	0.1979	0.2082	0.3489	0.1388
Homozygosity (Combined)	0.2714	0.2408	0.1633	0.1327	0.1143	0.2735	0.1776	0.2184	0.2286	0.2592	0.1531	0.1979	0.2082	0.3489	0.1388
Paternity Index (South)	1.9856	2.3872	4.4600	5.4500	6.3467	2.0406	3.1556	3.0233	2.6305	2.1195	3.8300	2.5856	2.6656	1.5106	4.2600
Paternity Index (West)	1.8135	2.3262	3.4500	3.7333	4.0833	2.4027	2.9833	2.1159	2.4679	2.812302	4.2000	2.9500	2.6167	1.5802	4.1666
Paternity Index (Combined)	1.8662	2.1523	3.2134	4.3439	4.5763	1.9380	2.9749	2.4625	2.3176	2.0800	3.4942	2.6018	2.4821	1.5058	3.7888
PE (South)	0.4962	0.5419	0.6852	0.7686	0.8107	0.4792	0.6668	0.6481	0.5752	0.5064	0.6783	0.5663	0.5966	0.3649	0.7192
PE (West)	0.4522	0.5258	0.6777	0.6889	0.6975	0.5194	0.6316	0.4977	0.5396	0.5202	0.7002	0.6476	0.5852	0.3769	0.7161
PE (Combined)	0.4755	0.5283	0.6707	0.7341	0.7731	0.4859	0.6492	0.5693	0.5425	0.4839	0.6968	0.6055	0.5768	0.3675	0.7143
DC (South)	0.8834	0.8910	0.9184	0.9403	0.9556	0.8339	0.9153	0.8852	0.9013	0.8502	0.9130	0.9038	0.9122	0.8094	0.9329
DC (West)	0.8650	0.8925	0.9175	0.9270	0.9290	0.8620	0.8805	0.8420	0.8830	0.8160	0.9045	0.8915	0.8945	0.7995	0.9170
DC (Combined)	0.8925	0.9099	0.9354	0.9552	0.9659	0.8726	0.9232	0.8951	0.9142	0.8620	0.9298	0.9161	0.9275	0.8157	0.9489
HWE (South)	0.0737	0.8976	0.9599	0.6155	0.8946	0.7984	0.3816	0.7192	0.5088	0.7444	0.5944	0.6743	0.9233	0.0867	0.9432
HWE (West)	0.0693	0.8454	0.9354	0.6019	0.8765	0.7681	0.3756	0.7090	0.4989	0.7311	0.5859=8	0.6646	0.9122	0.0767	0.9313
HWE (Combined)	0.0811	0.8721	0.9476	0.8087	0.8857	0.7933	0.3788	0.7136	0.5486	0.7376	0.5902	0.6696	0.9178	0.0827	0.9373

MP: Matching Probability, PIC: Polymorphism Information Content, Hexp: Expected Heterozygosity, Hobs: Observed Heterozygosity, h: homozygosity, P: Paternity Index, PE: Power of Exclusion, DC: Discrimination Capacity and HWE stands for Hardy-Weinberg Equilibrium.

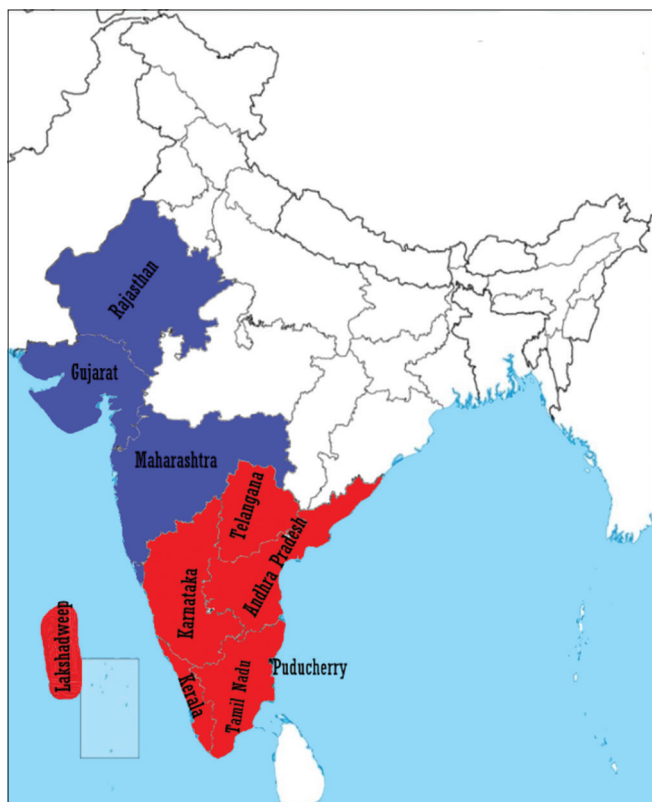


Figure 1: South and West Indian states from which the samples were collected.

15 STR it can also be used by the DNA analysts in South and West Indian laboratories for live forensic and relationship casework, as well as for further research outside. The statistical characteristics of allele frequency and forensic importance parameters derived from this work can be utilized for forensic identification and DNA relationship testing when analyzing live samples from South and West Indian population. This approach will undoubtedly help the existing forensic casework in South and West India employing allele frequency databases for STR markers. This effort will be supplemented by the development of a panel of genetic markers, specifically for the Indian population. Here is an example how the findings of this result can provide the accurate result for paternity case for example locus CSF1PO:

Mother: 11, 12

Child: 10, 11

Alleged Father: 9, 10

Matching allele with father: 10

Calculation of RMNE (Random Mating Not Excluded) = $P^2 + 2p(1-p)$, $q = 1-p$

Example

$$= (0.264)^2 + 2 * 0.264 * 0.736$$

$$= 0.0697 + 0.3886 = 0.4583$$

RMNE in several populations

	European	African	Spanish	South Indian
CSF1PO*10	0.264	0.255	0.291	0.2177

$$\begin{aligned} \text{This study results (E.g. South Indian)} &= (0.2177)^2 + 2 * 0.2177 * 0.7823 \\ &= 0.04739 + 0.3406 = 0.3880 \end{aligned}$$

Similar finding can be established for all 15 STR when relationship analysis carried for South and West Indian population samples analyzed. This result can be defendable in court of law in any country if the person in question (this case alleged father) samples processed.

It is possible to infer that the findings of this study add to the present Indian DNA industrial frequency dataset, as well as provide insight into variances, similarities, and genetic distances among the South and West Indian population. The presented statistical parameters in this study corresponded with the findings of the previous studies conducted in different parts of India. The authors recommend the future researchers to cover the entire Indian population for sample collection analyzing with more STR loci such as PowerPlex (PP21) system which has been introduced after PP16. The future researchers can also focus on SNPs as it has an advantage over STRs which even in minute quantities of DNA can provide valuable information about individualization. The use of the software for the determination of various phenotypic features is recent and, hence, is an area with a wide research scope.

5. CONCLUSION

The highest range of allele frequency was noted at TPOX for all three populations. In South, West, and combined Indian population, the observed alleles were 46, 42, and 54, respectively. The corresponding total alleles were 157, 159, and 182, respectively. The range of alleles was from 6 to 23 in all three populations. These three parameters showed the South and West Indian population had similar range of observed alleles, frequencies, and number of alleles. The locus Penta E was the most variable of the studied loci, displaying the highest values of discrimination capacity (0.9659), expected heterozygosity (0.8988), observed heterozygosity (0.8857), polymorphic information content (0.8916), power of exclusion (0.7731), and paternity index (4.5763) and the lowest values of match probability and homozygosity. However, locus TPOX recorded the highest match probability (0.1805) and homozygosity (0.3489), and the lowest polymorphic information content (0.6074), expected heterozygosity (0.6499), observed heterozygosity (0.6510), power of exclusion (0.3675), discrimination capacity (0.8157), and paternity index (1.5058) indicating the least variable locus in these population. The current data may serve as a suitable starting point for building the Indian population's DNA database. The information gathered here may supplement existing databases from other STR-based studies of Indian people. These 15 STR loci are useful for personal identification due to their specificity and polymorphism. Thus, the authors advocate for further genetic or forensic examination of polymorphism in the Indian population substructure utilizing short tandem repeat [STR]. People from the states of Lakshadweep, Pondicherry, Kerala, Tamil Nadu, Telangana, Andhra Pradesh, Karnataka, Gujarat, Rajasthan, Maharashtra, and Goa should be considered a genetically representative sample of the South and West Indian population for the purposes of genetic analysis or the establishment of a DNA database for the entire Indian population.

6. AUTHORS' CONTRIBUTIONS

All authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval

of the version to be published; and agreed to be accountable for all aspects of the work. All the authors are eligible to be an author as per the International Committee of Medical Journal Editors (ICMJE) requirements/guidelines.

7. FUNDING

There is no funding to report.

8. CONFLICTS OF INTEREST

The authors report no financial or any other conflicts of interest in this work.

9. ETHICAL APPROVALS

Ethical committee of SRM Institute of Science and Technology in Chennai, India has approved this study (1887/IEC/2020).

10. DATA AVAILABILITY

Data can be obtained from the corresponding author upon valid request.

11. PUBLISHER'S NOTE

This journal remains neutral with regard to jurisdictional claims in published institutional affiliation.

REFERENCES

- Jain S, Panigrahi I, Sheth J, Agarwal S. STR markers for detecting heterogeneity in Indian population. *Mol Biol Rep* 2011;39:461-5.
- Singh KS. India's Communities People of India. National Series. Vol. 4. India: Oxford University Press; 1998.
- Available from: https://en.wikipedia.org/wiki/south_india [Last accessed on 2023 Jan 25].
- Available from: https://en.wikipedia.org/wiki/western_india [Last accessed on 2023 Jan 13].
- Martin PD. National DNA database-practice and practicability. A forum for discussion. *Int Congr Ser* 2004;1261:1-8.
- Lee YS, Kennedy WD, Yin YW. Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell* 2009;139:312-24.
- Butler JM. Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* 2006;51:253-65.
- Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am J Hum Genet* 1994;55:175-89.
- Reilly P. Legal and public policy issues in DNA forensics. *Nat Rev Genet* 2001;2:313-7.
- Available from: <https://thewire.in/government/dna-technology-regulation-bill-seen-to-harm-minorities-hurt-privacy> [Last accessed on 2023 Feb 02].
- Panneerchelvam S, Norazmi MN. Forensic DNA profiling and database. *Malays J Med Sci* 2003;10:20-6.
- Shi Y, Li X, Ju D, Li Y, Zhang X, Zhang Y. Genetic polymorphisms of short tandem repeat loci D13S305, D13S631 and D13S634 in the Han population of Tianjin, China. *Exp Ther Med* 2015;10:773-7.
- D'Amato E, Ristow PG. Forensic statistics analysis toolbox (FORSTAT): A streamlined workflow for forensic statistics. In: *Forensic Science International: Genetics Supplement Series*. Vol. 6. Netherlands: Elsevier; 2017. p. e52-4.
- Projic P, Skaro V, Samija I, Pojskic N, Durmic-Pasic A, Kovscecic L, *et al.* Allele frequencies for 15 short tandem repeat loci in representative sample of Croatian population. *Croat Med J* 2007;48:473-7.
- Ghosh T, Kalpana D, Mukerjee S, Mukherjee M, Sharma AK, Nath S, *et al.* Genetic diversity of autosomal STRs in eleven populations of India. *Forensic Sci Int Genet* 2011;3:259-61.
- Singh M, Nandineni MR. Population genetic analyses and evaluation of 22 autosomal STRs in Indian populations. *Int J Leg Med* 2017;131:971-3.
- Badiye A, Kpoor N, Kumawat RK, Dixit S, Mishra A, Dixit A, *et al.* A study of genomic diversity in populations of Maharashtra, India, inferred from 20 autosomal STR markers. *BMC Res Notes* 2021;14:69.
- Balamurugan K, Kanthimathi S, Vijaya M, Suhasini G, Duncan G, Tracey M, *et al.* Genetic variation of 15 autosomal microsatellite loci in a Tamil population from Tamil Nadu, Southern India. *Leg Med (Tokyo)* 2010;12:320-3.
- Chauhan T, Kushwaha KP, Chauhan V. Genetic polymorphism of eleven STR loci in Rajput population of Delhi, India. *Forensic Res Criminol Int J* 2015;1:192-6.
- Shrivastava P, Jain T, Ben Trivedi V. Genetic polymorphism study at 15 autosomal locus in central Indian population. *Springerplus* 2015;4:566.
- Preet K, Malhotra S, Shrivastava P, Jain T, Rawat S, Varte LR, *et al.* Genetic diversity in Gorkhas: An autosomal STR study. *Sci Rep* 2016;6:32494.

How to cite this article:

Mathiyazhagan P, Palanimuthu T, Padmanathan A. Evaluation and comparison of 15 short tandem repeat loci of south and West Indian population for use in personal identification applications. *J App Biol Biotech*. 2023;11(6):216-227. DOI: 10.7324/JABB.2023.145640