Indonesian Journal of Biotechnology

VOLUME 29(1), 2024, 18-24 | RESEARCH ARTICLE

Genetic polymorphism and frequency study at 15 short tandem repeat loci in the North and East Indian populations for use in personal identification and applications in India

Prabakaran Mathiyazhagan[1,*], Thangaraju Palanimuthu[1], Agasthi Padmanathan[2]

[1]Faculty of Medicine and Health Sciences, SRM Institute of Science and Technology, Chennai, 603203 India
[2]Department of Mathematics, Tagore Engineering College, Chennai, 600127 India
*Corresponding author: prabakarmathi@yahoo.co.in

SUBMITTED 10 May 2023 **REVISED** 17 November 2023 **ACCEPTED** 29 November 2023

ABSTRACT Allele frequency is a crucial factor in estimating the weight of evidence (WoE) for an individual's involvement in a DNA sample. To determine the allele and genotype frequencies within the populations of the northern and eastern states of India, 15 short tandem repeats (STRs) were used, including Penta E, CSF1PO, D18S51, D7S820, D21S11, TH01, D3S1358, Vwa, FGA, TPOX, D8S1179, D16S539, D13S317, Penta D, and D5S818. The study involved 509 randomly selected individuals, analyzed using the PowerPlex 16 System Kit. Various statistical parameters of forensic significance were calculated using Forensic Statistic Analysis Toolbox (FORSTAT) software, including the typical paternity index (TPI), power of exclusion (PE), matching probability (MP), power of discrimination (PD), polymorphism information content (PIC), and observed (Hobs) and expected heterozygosities (Hexp). The analysis revealed a maximum allele frequency of 0.4282 at TPOX, with a minimum frequency of 0.0009 observed at different loci. FGA was found to be the most polymorphic loci among the 15 loci analyzed in the North and East Indian populations. Furthermore, no divergence from the Hardy-Weinberg equilibrium (HWE) was observed. The results serve as a valuable source of information for establishing a DNA database for North and East Indian populations, providing essential information for population genetics studies and forensic casework in India.

KEYWORDS FORSTAT; Hardy-Weinberg equilibrium; PowerPlex 16; STR; WoE

1. Introduction

India is the second largest populated country in the world and its human diversity is defined by 4,693 inversely recognized population groups which further subdivided into major communities, territorial nits and segments throughout the country. In the pre-notable India, there has been the existence of two ancestral groups namely ancestral North India and ancestral South India. Indian can be broadly divided into eastern, western, southern and northern regions. The genetic structure of Indian population has been influenced by these variables (Chandra et al. 2021). Information about STR genes, genotype frequencies and forensic important parameters are required for the estimation of frequencies for relationship establishment and human identification in North and East Indian population required. Since there has not been relevant studies for these two regions combined attempted, this study was conducted with the intention of determining the allele frequencies and forensic important statistical parameters for medical genetics interest of 15 loci, including Penta E, D18S51, TH01, D3S1358, D8S1179, Vwa, CSF1PO, D16S539, TPOX, D7S820, D13S317, Penta D, D21S11, FGA, and D5S818 in North and East Indian regions.

Haryana, Delhi, Uttar Pradesh, Punjab, Chandigarh, Jammu and Kashmir, Uttarakhand, and Himachal Pradesh are several states that have traditionally been considered to be part of North India. For the purpose of sample collecting, we have taken into consideration those eight states, which together cover an area of approximately 1,420,540 square kilometers (Geological Survey of India 2015).

States in the East Indian area include Arunachal Pradesh, Orissa, Nagaland, West Bengal, Mizoram, Assam, Tripura, Meghalaya, Bihar, Manipur, Sikkim, and Jharkhand. These twelve states represent a total area of approximately 418,323 km^2 that we have considered for sample collection (Geological Survey of India 2015).

"Microsatellites," also known as Short Tandem Repeats (STRs), are sequences of DNA that are repeated in tandem short (often between 1 and 6 base pairs). McMahon et al. (2017) estimate that STRs make up about 3% of the human genome and the other 97% is made up of sections of repeats and segmental duplications (50%), as

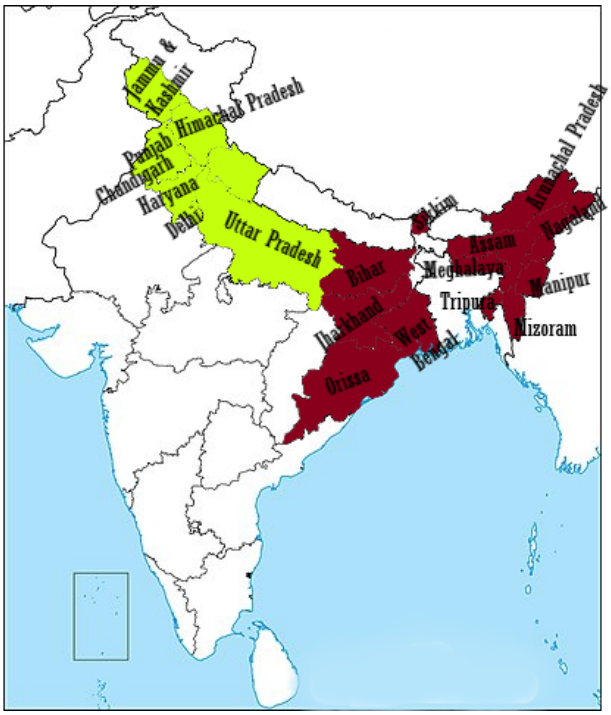

FIGURE 1 North and East Indian states where samples have been collected.

well as some highly conserved regulatory regions. STRs are stretches of DNA that contain five or more sets of two hundred or fewer consecutive nucleotides repeating in tandem and the same orientation. Most STRs are found in non-coding areas of the genome (Herrera and Bertrand 2018). A repeat unit's length is used to categorize STR repeat sequences. Dinucleotide repeats are sequences of two nucleotides that are repeatedly positioned next to one another. Trinucleotides have three nucleotides in the repeat unit, tetranucleotides have four, pentanucleotides have five, and hexanucleotides have six nucleotides in the core repeat. The degree to which STR sequences adhere to an incremental repeat pattern varies, as does the total number of repeats and the repeat unit length (Butler 2014). Genealogical markers like STRs and microsatellites are extremely useful in forensics and medical diagnostics. Since it is so common and so hypervariable across the whole genome, up to 150,000 informative loci have been identified, and this has helped to propel its popularity (Shrivastava et al. 2015). STRs are essential for crime case work and will continue to be so because of the large national DNA databases with STR profiles from criminal offenders and irreplaceable trace samples from old cases (Børsting and Morling 2016). Third-generation genetic markers, single nucleotide polymorphism (SNP), will eventually replace STR in certain applications such as genome mapping. More disciplines of study might benefit from STR analysis with a deeper knowledge of STR mutation and its high informative properties. Today, many forensic cases may be solved and more efficient DNA databases built because of the widespread use of STR profiles. Human identification testing has benefited greatly from the use of these current core loci and will continue to do so (Butler 2007).

Forensic DNA databases serve a crucial part in modern criminal justice systems, serving as valuable resources for investigations. Given the importance of forensic evidence, the European Council released a non-binding recommendation in February 1992 regarding the use of DNA in European criminal justice laws (Santos et al. 2013). A lot of researches have been done in India such as genetic polymorphism of eleven STR in Rajput population of Delhi, with 87 people in 2015; genetic variation of 15 autosomal microsatellite loci in a Tamil population from Tamil Nadu, Southern India in 2010 with 136 individuals; evaluation of the gene flow from East to West among various populations of Rajasthan, with 669 individuals in 2020; genetic polymorphism of 15 autosomal SR loci in population of Madhya Pradesh with 123 individuals in 2020; a study of genomic diversity in populations of Maharashtra, inferred from 20 autosomal STR markers with 227 individuals in 2021; genetic polymorphism study at 15 autosomal locus in central Indian population in 2015 with 582 individuals; allele frequencies for STR loci of the PowerPlex 16 multiplex system in five endogamous population of India with 95 individuals in 2002; genomic diversity of the Muslim population from Telangana inferred from 23 autosomal STRs; population genetic analyses and evaluation of 22 autsomal STRs in Indian populations in 2017 with 357 individuals; and genetic polymorphism of 21 STR markers in the representative sample of Indian population with 437 individuals in 2021. However, none of specific study has been done for North and East Indian population together, therefore this study has been attempted to cover North and East regions of India. In addition from eleven to 23 STRs have been used for these analyses in the previous studies, therefore in this study we have used 16 STRs.

In India, human DNA Profiling Bill using DNA-based technology was drafted by the Indian government and introduced to parliament in 2017. The bill's intent is to create a National Data Bank for DNA usage in forensics and a DNA Profiling Board. Inmates, witnesses, relatives of missing individuals, and others can all have their DNA profiles saved in this system (https://www.thehindu.com/news/national/everything-about-the-dna-technology-regulation-bill/article32682481.ece). In the near future, when the law is approved by the Government of India, a database of Indian people's STR allele frequencies will be produced, potentially containing the DNA profiles of millions of individuals to help with forensic and other types of DNA cases across the country. In light of these considerations, it is important to establish an Indian DNA database for use in forensics and in conducting credible statistical studies of the genetics of human populations. In this study, allele frequencies of North and East Indian population have been analysed, and in addition to that alleles observation in each locus was performed, the different alleles noted and its total numbers have also been presented (Table 1).

2. Materials and Methods

2.1. Samples and donors

The samples as well as the data of their nativity were only collected from the recruited donors who lives in the respective address for the last three generation, and agree to be involved in this study by signing informed consent. In order to identify the samples from respective states, the samples were clearly marked with unique individual numbers. Ethical permission was received from SRM Institute of Science and Technology in Chennai, India (1887/IEC/2020).

2.2. Buccal swab sample collection

Buccal swab samples from 509 randomly selected participants from the states of Uttarakhand, Haryana, Himachal Pradesh, Uttar Pradesh, Delhi, Chandigarh, Jammu and Kashmir, Punjab, West Bengal, Assam, Bihar, Jharkhand, Orissa, Mizoram, Sikkim, Arunachal Pradesh, Tripura, Meghalaya, and Nagaland were collected. It was performed by scraping the cells from the mouth usingcotton swabs on plastic sticks, and the impregnated ends of the swabs were stored in a 15 mL plastic tube supplied by the maker of buccal swabs.

2.3. Extraction and Quantification of DNA

DNA was extracted from buccal swabs using a swab solution kit distributed by Promega, Madison, USA.

2.4. Selection of STR Markers

The PP16 Multiplex STR was used for this investigation. The results of research have proven that the discrimination value increases with the number of STR loci that are utilised for typing. This is due to the fact that the chance of two people selected at random from a population holding precisely the same amount of repeat units for all of the STR that are being assessed is very low. According to Stanley et al. (2020), the magnitude of these variations may vary from person to person without having an effect on the individual's genetic health.

2.5. PCR amplification

PowerPlex 16 System was utilised for multiplex PCR amplification of 16 autosomal STR loci (Promega, Madison, USA). One injection was used to amplify and analyse all 16 loci at once in the 25 µL tube with the following amplification conditions; an initial incubation at 95 ºC for 11 min, 8 cycles of denaturation at 94 ºC for 1 min, annealing at 59 ºC for 1 min and extension at 72 ºC for 1 min; and a final extension at 60 ºC for 45 min. The PCR was run using a Gene Amp PCR system 970thermocycler (Applied Biosystems, Foster City, USA).

2.5.1 Control Samples

The manufacturer provided control DNA for positive control, while the nuclease free water was used for negative control. For quality control, a laboratory control DNA was used. The PowerPlex® 16 System allows co-amplification and three-color detection of sixteen loci (fifteen STR loci and Amelogenin). The system contains the loci Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, vWA, Amelogenin, Penta D, CSF1PO, D16S539, D7S820, D13S317 and D5S818. One primer specific for Penta E, D18S51, D21S11, TH01 and D3S1358 is labelled with fluorescein (FL); one primer specific for FGA, TPOX, D8S1179, vWA and Amelogenin is labeled with carboxy-tetramethylrhodamine (TMR); and one primer specific for Penta D, CSF1PO, D16S539, D7S820, D13S317 and D5S818 is labelled with 6-carboxy-4′,5′-dichloro-2′,7′-dimethoxy-fluorescein (JOE). All sixteen loci are amplified simultaneously in a single tube and analysed in a single injection or gel lane (https://www.promega.in/resources/protocols/technical-manuals/d0/powerplex-16-system-protocol/).

2.6. Genotyping of Powerplex-Amplified Samples

NextGen sequencer from Applied Biosystems, Foster City, CA, USA was used to sequence the amplified samples.

2.7. Statistical Evaluation

FORSTAT (Forensic Statistic Analysis Toolbox) software was used to calculate the importance of population genetics and forensic parameters such as matching probability (MP), typical paternity index (TPI), discrimination capacity (DC), power of exclusion (PE), polymorphism information content (PIC), observed (Hobs) and expected (Hexp) heterozygozities (Ristow and D'Amato 2017).

3. Results and Discussion

The allele frequencies and statistical parameters were calculated of the North and East Indian population as given in Table 1.

The autosomal STRs studied contributed to a better knowledge of population organisation and genetic diversity in the north and east India. Fifty six (56) distinct alleles were found in the group under study. Average frequency of 0.0999 (D3S158), 0.1250 (TH01), 0.0499 (D21S11), 0.0526 (D18S51), 0.0526 (Penta E), 0.1111 (D5S818), 0.1428 (D13S317), 0.0999 (D7S820), 0.1428 (D16S539), 0.1250 (CSF1PO), 0.0665 (Penta D), 0.0999 (Vwa), 0.0909 (D8S1179), 0.1110 (TPOX) and 0.0434 (FGA) were noted. A maximum frequency of 0.4282 at TPOX and a minimum frequency of 0.0009 at various loci were detected across 15 loci. In the studied population, a total of 186 alleles were found, with an average of 0.0942 per locus. A maximum of 23 alleles were found at FGA, followed by 20 at D21S11 and a minimum of 7 at D13S317 and D16S539. In this study, the presence of specific two alleles for the majority of the STRs D3S1358 (15 and 16), TH01(7 and 9.3), D21S11 (30 and 29), D18S51 (15 and 16), Penta E (12 and 7), D5S818 (12 and 11), D13S317 (12 and 11), D7S820 (10 and 11), D16S539 (11 and 12), CSF1PO (12 and 11), Penta D (9 and 11), Vwa (17 and

TABLE 1 North and East Indian population allele frequencies of fifteen STR loci.

Allele	D3S158	TH01	D21S11	D18S51	Penta_E	D5S818	D13S317	D7S820	D16S539	CSF1PO	Penta_D	Vwa	D8S1179	TPOX	FGA	
2.2											0.0776					
3.2											0.0058					
5		0.0019									0.0294			0.0009		
5.2					0.0785											
6		0.1866			0.0009			0.0009			0.0068			0.0599		
7		0.3104			0.1296	0.0127		0.0127		0.0373	0.0304			0.0117		
8		0.1159			0.0805	0.0117	0.0520	0.1876	0.0304	0.0392	0.0785		0.0088	0.4282		
9		0.1787			0.0333	0.0422	0.0500	0.1247	0.1463	0.0304	0.1817		0.0068	0.1620		
9.3		0.2003														
10		0.0049		0.0058	0.0628	0.0638	0.0353	0.3045	0.0943	0.2406	0.1129		0.0795	0.0746		
10.3								0.0009								
11		0.0009		0.0068	0.0785	0.3251	0.3212	0.2062	0.3055	0.2583	0.1601		0.0618	0.2288		
12	0.0009			0.1021	0.1669	0.3732	0.3713	0.1326	0.273	0.3212	0.1522	0.0009	0.1611	0.0324		
13	0.0019			0.0952	0.0943	0.1581	0.1296	0.0275	0.1306	0.0628	0.1159	0.0049	0.2760	0.0009		
13.2				0.0019												
14	0.0952			0.1051	0.0658	0.0098	0.0402	0.0009	0.0196	0.0098	0.0373	0.0893	0.2269			
14.2				0.0009												
15	0.2838			0.1719	0.0609	0.0029					0.0058	0.1424	0.1286			
15.2	0.00098			0.0009												
15.4					0.0009											
16	0.2888			0.1453	0.055						0.0019	0.2298	0.0471			
16.2															0.0009	
17	0.2121			0.1385	0.0451						0.0019	0.2671	0.0019			
18	0.1060			0.0923	0.0225							0.1689	0.0009		0.0166	
18.2															0.0049	
19	0.0088			0.0687	0.0108							0.0844			0.0599	
19.2															0.0019	
20	0.0009			0.0432	0.0117							0.0108			0.0864	
21				0.0098	0.0009							0.0009			0.1473	
21.2				0.0009											0.0029	
22				0.0078											0.1944	
22.2															0.0078	
22.3															0.0009	
23				0.0009											0.1522	
23.2															0.0039	
24				0.0009											0.1286	
24.2			0.0009												0.0009	
25			0.0009		0.0009										0.108	
25.2			0.0009													
26			0.0009												0.054	
27			0.0491												0.0166	
28			0.1817												0.0058	
29			0.2043												0.0019	
29.2			0.0029													
30			0.2337												0.0009	
30.2			0.022												0.0009	
31			0.0844													
31.2			0.0746												0.0009	
32			0.0068													
32.2			0.0805													
33			0.0058													
33.2			0.0294													
34			0.0029													
34.2			0.0029													
35			0.0098													
36			0.0039													
Total 56																
Recorded Alleles in each locus	10	8	20	19	19	9	7	10	7	8	16	10	11	9	23	Total 186

16), D8S1179 (13 and 14), TPOX (8 and 11) and FGA (22 and 23) were also noted.

By analyzing the most common and least common alleles, it was shown that allele 8 at TPOX has occurred 436 times, followed by the 12th allele of D5S818, which has been seen 380 times. In several of the loci, the rarest alleles only appeared once. Particularly, allele 20 in D3S158, allele 11 in THO1, allele 7 in TPOX 5, allele 13 in FGA, alleles 16.2,22.3, 24.2, 30, 30.2, and 31.2 were identified once.

Discrimination Capacity (DC), Homozygosity (H), Expected Heterozygosity (Hexp), Observed Heterozygosity (Hobs), Homozygosity (H), and Hardy-Weinberg Equilibrium (HWE) matching probability (MP), typical paternity index (TPI), power of exclusion (PE), polymorphism information content (PIC)

The highest polymorphic information content (PIC) value of 0.8954 was found at the Penta E locus, followed by 0.8629 at D18S51, 0.8511 at FGA, and 0.6733 at D5S818 locus. Penta E had the greatest value of expected heterozygosity (Hexp) (0.9032), tailed by D18S51 (0.8755), and FGA (0.8663). Heterozygosity (Hobs) was

TABLE 2 MCA and LCA of North and East Indian population.

STR Locus	MCA	LCA
D3S158	15 (289)	20 (1)
TH01	7 (316)	11 (1)
D21S11	30 (238)	24.2, 25, 25.2, and 26 (1)
D18S51	15 (175)	14.2 (1)
Penta E	12 (170)	6, 15.4, 22, and 25 (1)
D5S818	12 (380)	15 (3)
D13S317	12 (378)	14 (41)
D7S820	10 (310)	6, 8.1 and 14 (1)
D16S539	11 (311)	14 (20)
CSF1PO	12 (327)	14 (10)
Penta_D	9 (185)	13.4 (1)
Vwa	17 (272)	12 and 21 (1)
D8S1179	13 (281)	18 (1)
TPOX	8 (436)	5 and 13 (1)
FGA	22 (198)	16.2, 22.3, 24.2, 30, 30.2, and 31.2 (1)

measured to be between 0.7080 (D13S317) to 0.8940 (FGA). Homozygosity levels were measured to be anywhere from 0.1060 (FGA) to 0.2920 (D13S317). In terms of discriminating capacity, the values of 0.8626 (D5S818) and 0.9666 (Penta E) were discovered to be the lowest and highest, respectively. TPOX was shown to have the highest matching probability among the sampled population (0.1276). Power of exclusion was measured to be between 0.4759 (TPOX) to 0.7839 (FGA). In FGA, the paternity index reached a peak of 5.7589. There was no significant departure from the Hardy-Weinberg equilibrium ($p > 0.05$), as indicated in Table 3.

3.1. Discussion

The data from the Indian Genome Variation Consortium demonstrates that the Indian population is genetically more diverse than the global average. Previous genetic research (Kumawat et al. 2020) shown that India is a repository of genetic material admixed from people all over the globe. The Indian subcontinent's forensic and DNA testing sector needs to understand the nature and degree of genetic variability, which is important for establishing a large database of distinct polymorphic DNA. Unfortunately, a small amount of data on DNA markers is accessible to the Indian population, despite the fact that extensive data on different enzyme polymorphic markers, blood type markers, proteins, HLS, and so on is already available (Shrivastava et al. 2015). These results have revealed a more specific study as we examined the polymorphism and frequency analysis of 15 STR loci from the north and east Indian populations and compared them to previous research.

In this study, frequencies in the north and east Indian population ranged from 0.0009 (many alleles) to 0.4282 (TPOX). Singh and Nandineni (2017) conducted a study of genetic analysis and evaluation with 357 individuals using 22 autosomal STRs in Indian populations from 11 states across India and discovered that the frequencies ranged from 0.0010 (multiple alleles) to 0.3780 (TPOX), similar to our study, TPOX had the highest allele frequency and the lowest range observed at different loci. Preet et al. (2016) worked on Genetic Diversity in Gorkhas an Autosomal STR Study with 98 people, and observed frequencies ranging from 0.0050 to 0.4690, which is closely similar to our results. In our investigation, allele 8 of TPOX (436 out of 1018) and allele 12 of D5S818 (380 out of 1018) were shown to be dominant alleles. Allele 8 (0.4282) at the TPOX locus had the highest frequency, and it was also the most common allele in the north and east. Similarly, Balamurugan et al. (2010) conducted research in southern India on the Tamil community from Tamil Nadu with 15 autosomal microsatellite loci for analysing genetic diversity and reported that TPOX loci had the highest frequency of 0.4150, which is very close to our findings. Penta E has the highest polymorphism information content (PIC) score of 0.8954 and the highest discrimination capacity value of 0.9666. Similarly, Singh and Nandineni (2017) conducted a study of genetic testing and evaluation with 357 people utilising 22 autosomal STRs in Indian populations from 11 states throughout India and discovered that Penta E had the highest PIC and PD with reported values of 0.9100 and 0.9850. The eight most prevalent alleles found in this study are similar to those found in prior research; for example, this study found allele frequencies of 15 at D3S1358, 30 at D21S11, 12 at Penta E, 12 at D13S317, 10 at D7S820, 11 at D16S539, 12 at CSF1PO, and 17 at Vwa. Genomic diversity of the muslim population in Telangana (India) inferred from 23 autosomal STRs was studied by Srivastava et al. (2020), and they found that eight most common alleles (15 at D3S1358, 30 at D21S11, 12 at Penta E, 12 at D13S317, 10 at D7S820, 11 at D16S539, 12 at CSF1PO, 17 at Vwa) were noted. The results of our investigation revealed that all of the analysed loci exhibited a high degree of genetic polymorphism, with observed heterozygosity (Hobs) ranging from 0.7080 (D13S317) to 0.8940 (FGA) in the studied population. Dixit et al. (2020) examined the Madhya Pradesh population for Genetic Polymorphism of 15 Autosomal STR loci and discovered that observed heterozygosity (Hobs) varied from 0.659 (TPOX) to 0.911 (D21S11) in the analysed population, which is remarkably similar to our findings. TPOX had the maximum polymorphism with twenty-three distinct alleles, whereas markers like D13S317 and D16S539 had the least polymorphism with only seven alleles apiece. Because many parameter values are similar to prior research on the Indian population, the findings of this study can be used to construct regional or countrywide DNA databases.

Allele frequency information and knowledge of population substructure are critical to forensic DNA analysis and will be especially important as India develops the capability to do forensic DNA analysis on a regular basis. This work will undoubtedly help the existing forensic casework in India employing allele frequency databases for STR markers. This effort will be supplemented by the development of a panel of genetic markers, specifically for

TABLE 3 The forensic importance parameters of 15 STRs expressed by various statistical parameters.

STR LOCUS	D3S1358	TH01	D21S11	D18S51	Penta E	D5S818	D13S317	D7S820	D16S539	CSF1PO	Penta D	Vwa	D8S1179	TPOX	FGA
PIC	0.7234	0.7366	0.8233	0.8629	0.8954	0.6733	0.6883	0.7583	0.7447	0.7197	0.8554	0.7802	0.7867	0.6838	0.8511
Hexp	0.7613	0.7722	0.8417	0.8755	0.9032	0.7193	0.7301	0.7888	0.7773	0.7581	0.8691	0.8069	0.8114	0.7215	0.8653
Hobs	0.7780	0.7440	0.8360	0.8580	0.8680	0.7120	0.7080	0.8040	0.7780	0.7640	0.8700	0.8280	0.7780	0.7280	0.8940
H	0.2220	0.2560	0.1640	0.1420	0.1320	0.2880	0.2920	0.1960	0.2220	0.2360	0.1300	0.1720	0.2220	0.2720	0.1060
DC	0.8915	0.8998	0.9457	0.9550	0.9666	0.8626	0.8769	0.9087	0.9007	0.8930	0.9524	0.9198	0.9301	0.8724	0.9518
PE	0.5609	0.5050	0.6680	0.7127	0.7326	0.4552	0.4444	0.6112	0.5679	0.5351	0.7364	0.6543	0.5637	0.4759	0.7839
PM	0.1085	0.1002	0.0543	0.0449	0.0333	0.1374	0.1231	0.0912	0.0993	0.1069	0.0476	0.0802	0.0699	0.1276	0.0482
TPI	23.353	20.511	31.171	39.474	44.108	18.753	17.555	28.529	25.667	21.478	48.919	31.489	24.250	18.859	57.589
HWE	0.0580	0.7794	0.6803	0.5041	0.4123	0.7070	0.1624	0.9702	0.1861	0.0744	0.1171	0.7362	0.0590	0.0574	0.1532

the Indian population. 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 Indian population. Despite the fact that forensics and relationship testing play an important part in today's judicial system, their complexities and ethical problems are of concern to law enforcement, scientific, and legal institutions. Finally, the data gathered in this study may be used to create a DNA database for the north and east Indian populations.

The limitation of this study was to collect the samples from different parts of the states in India and high expenses to analyze a greater number of samples. STRs can be insufficient when DNA samples are degraded due to environmental exposure after the sample collection. In addition, not many nationwide data to reflect India are available to compare or validate the results. Recognizing and addressing limitations in research is essential for ensuring the rigor and scientific process of the scientific inquiry. Future researchers must navigate these limitations thoughtfully to effectively advance our understanding of the subject matter.

4. Conclusions

To conclude, the obtained results suggest that all the tested 15 loci are polymorphic and important markers for human identification and relationship testing. The statistical characteristics of allele frequency and forensic importance parameters derived from this work can be utilised for forensic identification and DNA relationship testing when analysing live samples from the north and east Indian populations. This study should also provide the groundwork for the creation of a national DNA database using STR for the entire country of India. Furthermore the data obtained in this study is expected to increase the DNA databank in India.

Acknowledgments

This research was supported by the Faculty of Medical and Health Sciences of SRM Institute of Science and Technology-Chennai-India.

Authors' contributions

PM designed the whole work and performed sample collection, extraction of DNA, polymerase chain reaction, DNA sequencing, and wrote the article. TP reviewed, interpreted, and corrected the article. AP analyzed the data and offered calculation support. All the authors read and approved the final article.

Competing interests

No competing financial interests for this study and self-funded by author PM and not supported by any external organization.

References

Balamurugan K, Kanthimathi S, Vijaya M, Suhasini G, Duncan G, Tracey M, Budowle B. 2010. Genetic variation of 15 autosomal microsatellite loci in a Tamil population from Tamil Nadu, Southern India. Leg. Med. 12(6):320–323. doi:10.1016/j.legalmed.2010.07.004.

Børsting C, Morling N. 2016. Genomic applications in forensic medicine. Med. Heal. Genomics p. 295–309. doi:10.1016/B978-0-12-420196-5.00022-8.

Butler J. 2007. Short tandem repeat typing technology used in human identity testing. BioTechniques 43:ii–v. doi:10.2144/000112582.

Butler JM. 2014. Advanced topics in forensic DNA typing: Interpretation. Elsevier Science. doi:10.1016/C2011-0-07649-4.

Chandra D, Mishra VC, Anthwal A, Gupta N, Bhardwaj AK, Raina A, Kumari S, Dey N, Raina V. 2021. Genetic polymorphism of 21 STR markers in the representative sample of Indian population. Gene Reports 25:101411. doi:10.1016/j.genrep.2021.101411.

Dixit S, Dubey I, Kumawat R, Tripathi I, Shrivastava D, Panda M, Shrivastava P. 2020. Genetic polymorphism of 15 autosomal STR loci in population of Madhya Pradesh. Medico-Legal Updat. 20(4):2074–2079. doi:10.37506/mlu.v20i4.2150.

Geological Survey of India. 2015. Northern Region. Geological Survey of India, MOI, Government of India. Archived from the original on 24 September 2015. Retrieved 2 May 2015.

Herrera RJ, Bertrand RG. 2018. The nature of evolution. London: Elsevier.

Kumawat R, Shrivastava P, Shrivastava D, Mathur G, Dixit S, Chaubey G. 2020. Evaluating the gene flow from East to West among various populations of Rajasthan, India. Preprints p. 2020060272. doi:10.21203/rs.3.rs-18353/v1.

McMahon K, Paciorkowski AR, Walters-Sen LC, Milunsky JM, Bassuk A, Darbro B, Diaz J, Dobyns WB, Gropman A. 2017. Neurogenetics in the genome era. Swaiman's Pediatr. Neurol. Princ. Pract. Sixth Ed. p. 257–267. doi:10.1016/B978-0-323-37101-8.00034-5.

Preet K, Malhotra S, Shrivastava P, Jain T, Rawat S, Varte LR, Singh S, Singh I, Sarkar S. 2016. Genetic diversity in Gorkhas: An autosomal STR study. Sci. Rep. 6:32494. doi:10.1038/srep32494.

Ristow PG, D'Amato ME. 2017. Forensic statistics analysis toolbox (FORSTAT): A streamlined workflow for forensic statistics. Forensic Sci. Int. Genet. Suppl. Ser. 6:e52–e54. doi:10.1016/j.fsigss.2017.09.006.

Santos F, Machado H, Silva S. 2013. Forensic DNA databases in European countries: is size linked to performance? Life Sci. Soc. Policy 9:12. doi:10.1186/2195-7819-9-12.

Shrivastava P, Jain T, Trivedi VB. 2015. Genetic polymorphism study at 15 autosomal locus in central Indian population. Springerplus 4(1):566. doi:10.1186/s40064-015-1364-1.

Singh M, Nandineni MR. 2017. Population genetic analyses and evaluation of 22 autosomal STRs in Indian populations. Int. J. Legal Med. 131(4):971–973. doi:10.1007/s00414-016-1525-y.

Srivastava V, Surekha Rani H, Kumawat R, Chaubey G, Shrivastava P. 2020. Genomic diversity of the Muslim population from Telangana (India) inferred from 23 autosomal STRs. Ann. Hum. Biol. 47(7-8):652–658. doi:10.1080/03014460.2020.1822915.

Stanley UN, Khadija AM, Bukola AT, Precious IO, Davidson EA. 2020. Forensic DNA profiling: Autosomal short tandem repeat as a prominent marker in crime investigation. Malays. J. Med. Sci. 27(4):22–35. doi:10.21315/mjms2020.27.4.3.