

Genetic Polymorphism of 15 Autosomal STR loci in Population of Madhya Pradesh

Shivani Dixit^{1,3}, Ishwar Prasad Dubey^{1,2}, R.K. Kumawat^{3,4}, I.P. Tripathi²,
Divya Shrivastava³, Muktikanta Panda⁵, Pankaj Shrivastava¹

¹DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar-470001, M.P, India, ²Faculty of Science and Environment, MGCGVV, Chitrakoot, Satna-485334, M.P, India, ³School of Life Sciences, Jaipur National University, Jaipur-302017, Rajasthan, India, ⁴DNA Division, State Forensic Science Laboratory, Rajasthan, Jaipur -302016, India, ⁵Department of Anthropology, Dr. HS Gour Viswavidyalaya, Sagar (MP), India, 470003

Abstract

We report here an autosomal STR DNA database of Population of Madhya Pradesh and evaluation of autosomal STR diversity in the context of population genetics and forensic efficiency parameters. In this study, Allele frequencies of 15 autosomal STR loci were investigated in 123 unrelated individuals (57 Female & 66 Male) from the various geographical regions of Madhya Pradesh using PowerPlex® 16 HS System. All the studied loci appeared the high degree of genetic polymorphism with observed Heterozygosity (Ho= observed heterozygosity) ranged from 0.911 (D21S11) to 0.659 (TPOX). Locus PENTA E showed the highest power of discrimination (0.978) and the highest polymorphism (PIC = 0.90). The combined probability of match (CPM) and combined paternity Index (CPI) for all the 15 STR loci was found to be 2.8×10^{-18} and 1.2×10^6 , respectively. The combined discrimination power (CPD) and combined exclusion power (CPE) for all the tested 15 Loci was observed 1 and 0.9999989 respectively, assign that all the loci are highly polymorphic and have the potential for forensic application.

Keywords: Autosomal, Madhya Pradesh, Forensic DNA Typing, Population study, Short tandem repeats (STRs).

Introduction

Madhya Pradesh is geographically located in the centre of India. It is rich in cultural and linguistic diversity and the second largest state by area in the country. In history, it was the primary route of human migration, which is the major factor behind the rich human genetic diversity. As per census 2011, the population of Madhya Pradesh is 7.27 Crore, which is 6% of the total population of the country⁽¹⁾. In this study, we evaluated the population genetic parameters in population of Madhya Pradesh. The data reported here will contribute

toward the DNA database that can be used as a reference for forensic studies for human identity as well as for population genetic and anthropological purposes.

Materials and Method

Sampling: In this study 123 (57 Female & 66 Male) autosomal short tandem repeat (STR) data generated from blood samples were taken from routine casework analysis performed by the authors who work at the DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar, MP, India. The samples for this study were taken only from unrelated individuals belonging to all regions of Madhya Pradesh. For all the samples prior informed written consent was also obtained from the individuals as per the routine DNA examination procedure of the laboratory and following the Code of Ethics of the World Medical Association (i.e., Declaration of Helsinki)⁽²⁾. No minor was involved in the study.

Corresponding Author:

Pankaj Shrivastava

DNA Fingerprinting Unit, State Forensic
Science Laboratory, Sagar-470001, M.P, India
e-mail: pankaj.shrivastava@rediffmail.com
Phone No.: 9424371946

DNA Extraction and Quantification: Genomic DNA extracted from the blood samples using 12 GC automated DNA extraction system (PSS, Japan). Extracted DNA was quantified using Qubit Fluorometric Quantitation (Applied Biosystems, Foster City, CA, USA- Thermo) with Qubit ds DNA BR Assay Quantification Kit (Thermo) as per the protocol recommended by the manufacturer.

PCR amplification and STR typing: 15 STR loci and a sex determination marker (Amelogenin) were amplified simultaneously using PCR multiplex PowerPlex® 16 HS (Promega) as per procedure recommended by the manufacturer using ABI 9700 thermal cycler (Thermo). The amplified DNA fragments were separated on capillary electrophoresis following the manufacturer’s recommendations using a 3100Genetic Analyzer (Thermo) and sized with GeneScan500-LIZ internal lane size standard (Thermo) following manufacturer’s recommended protocol. After amplification and separation, DNA fragments were genotyped using GeneMapper Software v3.5 (Thermo). The complete procedure and steps were followed as per standardized internal laboratory and kit controls. The authors also have qualified international proficiency test organized by GITAD, Spain (<http://gitad.ugr.es/principal.htm>).

Statistical and phylogenetic analysis: Allele frequencies were computed using GenAIEx version 6.5⁽³⁾. The Hardy–Weinberg P-values, observed heterozygosity (Ho) and expected heterozygosity (He), population differentiation test were calculated using Arlequin Version3.5⁽⁴⁾. Genetic parameters of forensic interest viz. Power of Discrimination (PD), Matching Probability (PM), Polymorphic Information Content (PIC), Power of Exclusion (PE), and Typical Paternity Index (TPI) were calculated using the PowerStatsv1.2⁽⁵⁾.

Allele frequencies obtained from the studied population were compared with data of other published populations on the basis of p-value of the exact test for Hardy–Weinberg equilibrium (calculated using the Arlequin software version 3.5⁽⁴⁾). Also, population differentiation tests Fst genetic distances for comparison with other published populations were carried out using Arlequin version 3.5⁽⁴⁾. A phylogenetic tree of the studied population with previously reported 15 populations was constructed with the Neighbour-joining method by software package Poptree2⁽⁶⁾ comparing 13 common STR loci showing the inter-population relationship.

Findings: Distribution of allele frequencies and statistical evaluations of the 15 autosomal STR loci in the studied population are reported in Supplementary data. All loci demonstrated the high level of genetic polymorphism with observed Heterozygosity (Ho) ranged from 0.911 (D21S11) to 0.659 (TPOX), the Expected Heterozygosity (HE) ranged from 0.904 (PENTA E) to 0.701 (TPOX). PENTA E demonstrated the greatest power of discrimination (0.978) in the studied population, while TPOX showed the lowest (0.863). The power of exclusion (PE) varies between 0.367 (TPOX) and 0.817 (D21S11). Among all the studied 15 loci, PENTA E had the highest polymorphism (PIC = 0.90), whereas TPOX had the lowest (PIC = 0.65). The most polymorphic and discriminatory STR locus of the studied population is PENTA E with the value of 0.978 (power of discrimination) and 0.90 (polymorphic information content), respectively. The combined power of discrimination (CPD) and the combined power of Exclusion (CPE) were observed with a value of 1 and 0.0999989, respectively. The combined probability of match (CPM) and combined paternity Index (CPI) for all the 15 STR loci was found to be 2.8×10^{-18} and 1.2×10^6 , respectively.

Table 1: Observed allele frequency and forensic parameters for 15 Autosomal STRs in studied Population (N=123)

| Allele/n | D3S1358 | THO1 | D21S11 | D18S51 | PENTA E | D5S818 | D13S317 | D7S820 | D16S539 | CSF1PO | PENTA D | vWA | D8S1179 | TPOX | FGA |
|----------|---------|-------|--------|--------|---------|--------|---------|--------|---------|--------|---------|-----|---------|-------|-----|
| 4 | | | | | 0.004 | | | | | | | | | | |
| 5 | | | | | 0.073 | | | | | | | | | | |
| 6 | | 0.329 | | | | | | | | | 0.004 | | | | |
| 7 | | 0.134 | | | 0.069 | | 0.008 | 0.024 | | | 0.004 | | | | |
| 8 | | 0.102 | | | | 0.008 | 0.215 | 0.207 | 0.102 | 0.008 | 0.033 | | 0.024 | 0.309 | |
| 9 | | 0.305 | | 0.004 | 0.012 | 0.045 | 0.110 | 0.053 | 0.179 | 0.024 | 0.215 | | 0.004 | 0.175 | |

| Allele/n | D3S1358 | THO1 | D21S11 | D18S51 | PENTA E | D5S818 | D13S317 | D7S820 | D16S539 | CSFIPO | PENTA D | vWA | D8S1179 | TPOX | FGA |
|--|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 9.3 | | 0.126 | | | | | | | | | | | | | |
| 10 | | 0.004 | | | 0.033 | 0.110 | 0.069 | 0.211 | 0.122 | 0.207 | 0.187 | | 0.142 | 0.081 | |
| 11 | | | | 0.016 | 0.146 | 0.362 | 0.232 | 0.321 | 0.236 | 0.333 | 0.236 | | 0.089 | 0.407 | |
| 12 | | | | 0.069 | 0.159 | 0.337 | 0.260 | 0.167 | 0.220 | 0.341 | 0.102 | | 0.110 | 0.028 | |
| 13 | | | | 0.102 | 0.045 | 0.126 | 0.085 | 0.012 | 0.130 | 0.081 | 0.126 | | 0.163 | | |
| 14 | 0.057 | | | 0.211 | 0.085 | 0.012 | 0.020 | 0.004 | 0.012 | 0.004 | 0.073 | 0.102 | 0.122 | | |
| 15 | 0.329 | | | 0.228 | 0.098 | | | | | | 0.020 | 0.110 | 0.191 | | |
| 16 | 0.268 | | | 0.154 | 0.085 | | | | | | | 0.248 | 0.126 | | |
| 17 | 0.224 | | | 0.089 | 0.094 | | | | | | | 0.236 | 0.028 | | |
| 18 | 0.102 | | | 0.053 | 0.053 | | | | | | | 0.207 | | | 0.004 |
| 19 | 0.020 | | | 0.049 | 0.008 | | | | | | | 0.081 | | | 0.077 |
| 20 | | | | | 0.016 | | | | | | | 0.016 | | | 0.106 |
| 20.2 | | | | | | | | | | | | | | | 0.004 |
| 21 | | | | 0.024 | 0.004 | | | | | | | | | | 0.134 |
| 21.2 | | | | | | | | | | | | | | | 0.004 |
| 22 | | | | | 0.008 | | | | | | | | | | 0.150 |
| 22.2 | | | | | | | | | | | | | | | 0.008 |
| 23 | | | | | 0.008 | | | | | | | | | | 0.171 |
| 23.2 | | | | | | | | | | | | | | | 0.004 |
| 24 | | | | | | | | | | | | | | | 0.122 |
| 24.2 | | | | | | | | | | | | | | | 0.004 |
| 25 | | | | | | | | | | | | | | | 0.134 |
| 26 | | | | | | | | | | | | | | | 0.077 |
| 27 | | | 0.020 | | | | | | | | | | | | |
| 28 | | | 0.126 | | | | | | | | | | | | |
| 29 | | | 0.215 | | | | | | | | | | | | |
| 30 | | | 0.195 | | | | | | | | | | | | |
| 30.2 | | | 0.020 | | | | | | | | | | | | |
| 31 | | | 0.037 | | | | | | | | | | | | |
| 31.2 | | | 0.102 | | | | | | | | | | | | |
| 32 | | | 0.008 | | | | | | | | | | | | |
| 32.2 | | | 0.195 | | | | | | | | | | | | |
| 33 | | | 0.004 | | | | | | | | | | | | |
| 33.2 | | | 0.077 | | | | | | | | | | | | |
| PD | 0.896 | 0.898 | 0.943 | 0.957 | 0.978 | 0.878 | 0.929 | 0.906 | 0.939 | 0.869 | 0.936 | 0.934 | 0.961 | 0.863 | 0.967 |
| PIC | 0.715 | 0.72 | 0.82 | 0.83 | 0.90 | 0.68 | 0.78 | 0.74 | 0.80 | 0.67 | 0.81 | 0.78 | 0.85 | 0.65 | 0.86 |
| PE | 0.520 | 0.520 | 0.817 | 0.718 | 0.654 | 0.493 | 0.549 | 0.686 | 0.593 | 0.379 | 0.751 | 0.479 | 0.654 | 0.367 | 0.670 |
| PI | 2.05 | 2.05 | 5.59 | 3.62 | 2.93 | 1.92 | 2.20 | 3.24 | 2.46 | 1.50 | 4.10 | 1.86 | 2.93 | 1.46 | 3.08 |
| Ho | 0.756 | 0.756 | 0.911 | 0.862 | 0.829 | 0.740 | 0.772 | 0.846 | 0.797 | 0.667 | 0.878 | 0.732 | 0.829 | 0.659 | 0.837 |
| He | 0.756 | 0.754 | 0.843 | 0.850 | 0.904 | 0.725 | 0.808 | 0.778 | 0.822 | 0.722 | 0.830 | 0.811 | 0.865 | 0.701 | 0.874 |
| P-value | 0.30839 | 0.66032 | 0.33562 | 0.47931 | 0.41001 | 0.91278 | 0.4168 | 0.62697 | 0.48614 | 0.04369 | 0.2906 | 0.06243 | 0.39733 | 0.32536 | 0.52679 |
| Pm | 0.104 | 0.102 | 0.057 | 0.043 | 0.022 | 0.122 | 0.071 | 0.094 | 0.061 | 0.131 | 0.064 | 0.066 | 0.039 | 0.137 | 0.033 |
| F | -0.001 | -0.002 | -0.080 | -0.013 | 0.083 | -0.020 | 0.044 | -0.087 | 0.031 | 0.077 | -0.058 | 0.097 | 0.041 | 0.061 | 0.042 |
| PD- Power of discrimination, PIC- Polymorphism information content, PE- Power of exclusion, PI- Paternity index, Ho- Observed heterozygosity, He- Expected Heterozygosity, P value- HWE test, Pm- Matching Probability, F = Fixation Index = (He - Ho)/He = 1 - (Ho/He) | | | | | | | | | | | | | | | |
| Values in bold represent lowest and highest values of forensic parameters | | | | | | | | | | | | | | | |

Population differentiation test was carried out through Arlequin software⁽⁴⁾. In this test, allelic frequencies of studied population (Madhya Pradesh) was compared with geographically neighbouring regions for which published data is available, namely Balmiki (Punjab)⁽⁷⁾, Sakaldwipi-Brahmin (Jharkhand)⁽⁷⁾, Mahadev-Koli (Maharashtra), Iyengar (Tamilnadu)⁽⁷⁾, Kurumans (Tamilnadu)⁽⁷⁾, Lambadi(Andhra Pradesh)⁽⁸⁾, Yerukula (Andhra Pradesh)⁽⁸⁾, Santal (Chota Nagpur)⁽⁹⁾, Bhil (Madhya Pradesh)⁽¹⁰⁾, Bhil (Gujarat)⁽¹¹⁾, Central Indian population,⁽¹²⁾ Gond (Madhya Pradesh)⁽¹³⁾, Kahar population(Uttar Pradesh)⁽¹⁴⁾, population of Jharkhand⁽¹⁵⁾, population of Uttar Pradesh⁽¹⁶⁾. After applying the Bonferroni correction⁽¹⁷⁾ (at the significant level of 95% $p < 0.003$), the population differentiation tests showed that studied population had significant differences with Balmiki (Punjab) in 2 out of 13 loci, with Sakaldwipi_Brahmin (Jharkhand) in 4 out of 13 loci, with Mahadev_Koli (Maharashtra) in 5 out of 13 loci, with Iyengar (Tamilnadu) in 1 out of 13 loci, with Kurumans (Tamilnadu) in 1 out of 13 loci, with Lambadi (Andhra Pradesh) 3 out of 13 loci, with Yerukula (Andhra Pradesh) 9 out of 13 loci, with Santal (Chotanagpur) 6

out of 13 loci, with Bhil (Madhya Pradesh) 3 out of 13 loci, with Bhil (Gujarat) 2 out of 13 loci, with central Indian population 1 out of 13 loci, Gond (Madhya Pradesh) 6 out of 13 loci, with population of Jharkhand 1 out of 13 loci and there is no significant difference with UP population and Kahar population (Uttar Pradesh).

To confirm the results of the population differentiation test (Fst-P), Neighbour Joining (NJ) Tree (Fig.1), and Principal Component Analysis (PCA) (Fig.2) were performed for the studied population along with previously reported populations. NJ tree revealed the genetic relatedness with the previously reported population of Madhya Pradesh, central Indian population, and clustered with Population of Uttar Pradesh along with some outlier population (Fig.1). In Principal Component Analysis (PCA), component 1 and component 2 explained 67.918 % variance in the studied population. In the PCA plot, geographically close population pooled in one cluster (Fig.2). Neighbour Joining Tree and PCA plot showed consistency with each other.

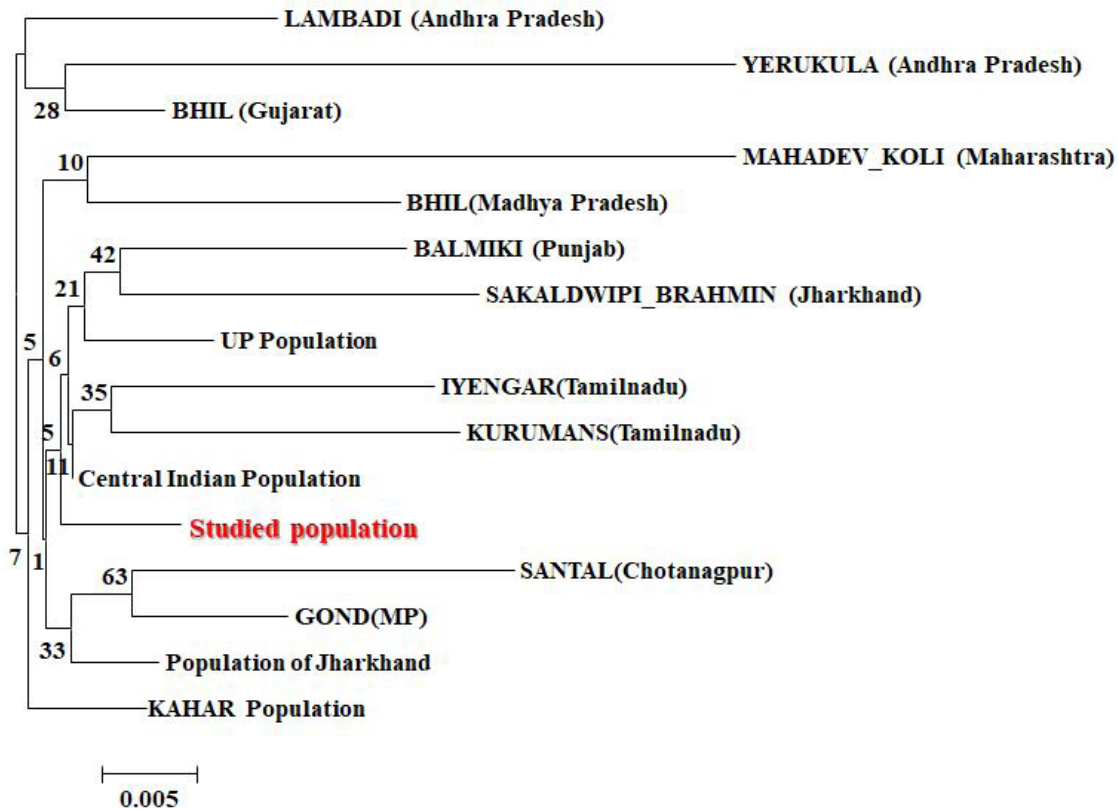


Fig. 1: NJ tree of the studied population showing the genetic relatedness with the previously reported populations.

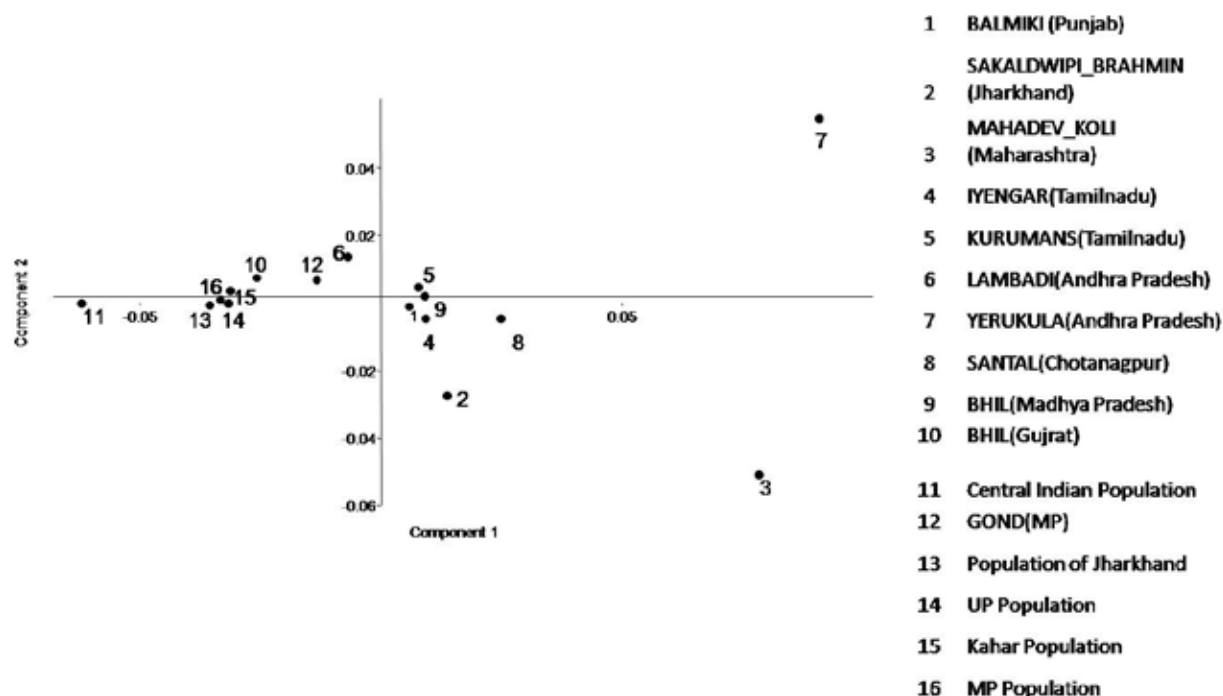


Fig. 2: Principal component analysis (PCA) plot based on Nei’s Da distance matrix showing the distance pattern of studied population with previously reported populations.

Conclusion

The data set generated from this study will enrich the DNA data bank. The studied 15 STR markers are polymorphic, informative and can be used in forensic application as well as anthropological studies for Indian population. The studied population has significant genetic diversity.

Conflict of Interest: Authors declared that they have no conflict of interest.

Source of Funding: Self funding

Ethical Clearance: This study was approved by the Institutional Ethical Committee of the Jaipur National University, Jaipur vide letter no. JNUMSRC/IEC/2018/45 dated 20.07.2018.

Acknowledgements: Authors acknowledge the support and motivation from Director, State Forensic Science Laboratory, Sagar, MP, India.

References

1. INDIA POFMPIN. Census of India 2011 Provisional Population Totals. Office of the Registrar General and Census Commissioner New Delhi; 2011.

2. Rickham PP. Human experimentation. Code of ethics of the world medical association. Declaration of Helsinki. Br Med J. 1964;2(5402):177.

3. Peakall ROD, Smouse PE. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 2006;6(1):288–95.

4. Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinforma. 2005;1:117693430500100000.

5. Tereba A. Tools for analysis of population statistics. Profiles DNA. 1999;2:14–6.

6. Takezaki N, Nei M, Tamura K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol Biol Evol. 2009;27(4):747–52.

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;5(3):259–61.

8. Hima Bindu G, Trivedi R, Kashyap VK. Genotypic polymorphisms at fifteen tetranucleotides and two

- pentanucleotide repeat loci in four tribal populations of Andhra Pradesh, southern India. *J Forensic Sci.* 2005;50:978–83.
9. Banerjee J, Trivedi R, Kashyap VK. Polymorphism at 15 Short Tandem Repeat AmpF ℓ STR \circledR IdentifilerTM Loci in Three Aboriginal Populations of India: An Assessment in Human Identification. *J Forensic Sci.* 2005;50(5):JFS2005151-6.
 10. Shrivastava P, Jain T, Gupta U, Trivedi V Ben. Genetic variation at 15 autosomal STR loci in Bhil tribal population of Central India. *Ann Hum Biol.* 2016;43(1):81–4.
 11. Chaudhari RR, Dahiya MS. Genetic diversity of 15 autosomal short tandem repeats loci using the AmpFLSTR \circledR IdentifilerTM kit in a Bhil Tribe Population from Gujarat state, India. *Indian J Hum Genet.* 2014;20(2):148.
 12. Shrivastava P, Jain T, Trivedi V Ben. Genetic polymorphism study at 15 autosomal locus in central Indian population. *Springerplus.* 2015;4(1):566.
 13. Shrivastava P, Jain T, Trivedi V Ben. Structure and genetic relationship of five populations from central India based on 15 autosomal STR loci. *Ann Hum Biol.* 2017;44(1):74–86.
 14. Batham MS, Kushwaha KPS, Chauhan T, Kumawat RK, Shrivastava P. Autosomal STR allele frequencies in Kahars of Uttar Pradesh, India, drawn with PowerPlex \circledR 21 multiplex system. *Int J Legal Med.* 2019;1–3.
 15. Imam J, Reyaz R, Singh RS, Bapuly AK, Shrivastava P. Genomic portrait of population of Jharkhand, India, drawn with 15 autosomal STRs and 17 Y-STRs. *Int J Legal Med.* 2018;132(1):139–40.
 16. Shrivastava A, Kumawat R, Dixit S, Kaitholia K, Shrivastava D, Yadav VK, et al. Genetic data for PowerPlex 21TM autosomal and PowerPlex 23 Y-STRTM loci from population of the state of Uttar Pradesh, India. *Int J Legal Med.* 2019;1–3.
 17. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *Bmj.* 1995;310(6973):170.