Population Genetics for Autosomal STR Loci in Sikh Population of Central India

Dogra D¹, Shrivastava P², Chaudhary R³, Gupta U³ and Jain T²

¹Department of Biotechnology, Barkatullah University, Bhopal 462023, Madhya Pradesh, India
²DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar 470001, Madhya Pradesh, India
³Biotechnology Division, Department of Zoology, Government Motilal Vignan Mahavidyalaya, Bhopal 462023, Madhya Pradesh, India

*Corresponding author: Pankaj Shrivastava, DNA Fingerprinting Unit, State Forensic Science Laboratory, Sagar-470001, Madhya Pradesh, India, Tel: 94243 71946, E-mail: pankaj.shrivastava@rediffmail.com

Rec date: February 4, 2015, Acc date: February 23, 2015, Pub date: February 26, 2015
Copyright: © 2015 Dogra D, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

This study is an attempt to generate genetic database for three endogamous populations of Sikh population (Arora, Jat and Ramgariha) of Central India. The analysis of eight autosomal STR loci (D16S539, D7S820, D13S317, FGA, CSF1PO, D21S11, D18S51, and D2S1338) was done in 140 unrelated Sikh individuals. In all the three studied populations, all loci were in Hardy-Weinberg equilibrium except at locus FGA in Ramgarhiha Sikh and locus D16S539 in Arora Sikh. An analysis of molecular variance (AMOVA) showed 1% variation among the three studied populations. The close genetic relationship between Jat and Ramgarhiha Sikh population were confirmed in the MDS Plot generated from the pairwise genetic distances.

Keywords: Ramgarhiha dikh; Jat sikh; Arora sikh; Central India; Short tandem repeats; Population data

Introduction

Microsatellite markers are most suited for the genetic structural assessment of a population due to ease of use, co-dominant inheritance, high polymorphism and mutation rate [1,2]. Use of multiplex polymerase chain reaction based technology has made the task of identification easier and this has emerged as the dominant conclusive identification method in forensic investigation and for anthropological studies. India is a country rich in ethnic, cultural and linguistic variant groups. Although a considerable amount of information on polymorphism at microsatellite loci in humans is now available, but such studies are confined to limited groups [3-13]. Like most other Indians, Sikh is endogamous by caste and exogamous by sub caste [14]. Sikhism is India’s fourth-largest religion and has existed over 500 years, beginning with the birth of its founder Guru Nanak dev in the late 15th century C.E. in the Punjab region of what is today in India and Pakistan. The Sikhs community has a stronghold in the state of Punjab; roughly 60% of the population belongs to the Sikh faith. The state of Madhya Pradesh (MP) comprises about 1.9% of the Sikh population [15]. There is necessary to fill a big lacuna with information about the genetic diversity of Sikh Population. A very few number of genetic studies on Sikh population have been carried out around the world [16-23]. However, no STR marker based study on Sikh populations of central India has been reported in the literature till date. Therefore the present data would be used in the forensics and individual identification for these selected population groups and these genetic data would enrich the genetic informational resource. In the present study 140 unrelated individuals of three studied Sikh population Arora (n=40), Ramgarhia (n=50) and Jat (n=50) were taken for analysis from MP, India on the nine microsatellite loci which are D13S317 (13q22-31), D7S820 (7q11.21-22), Amelogenin (Xp22.1-22.3:yp11.2), D2S1338 (2q35-37.1), D21S11 (21q11.2-q21), D18S59 (16q24-qter), D18S51 (18q21.3), CSF1PO (5q33.3-34) and FGA (4q28). These loci were chosen for two reasons. Firstly, they consist of repetitions of tetrancleotide repeat units and are therefore less prone to slippage of polymerase during enzymatic amplification [24]. Secondly they are located on different chromosomes so there is no possibility of mitotic recombination as they are present far apart from each other. All studied loci are substantially unlinked, which make them ideal tools to study genomic variation.

Materials and Methods

Sample collection

Venous blood from a total of 140 unrelated healthy individuals from three endogamous groups (50 Ramgarhiha, 50 Jat sikh and 40 Arora sikh) of Sikh population from Bhopal and Raiesen district of Madhya Pradesh, India were taken on FTA card.

DNA extraction

A 1.2 mm punch from a dried sample spot on FTA paper was taken in a PCR tube. FTA purification reagent (200 μl) was added to PCR tube, incubated for 5 minutes at room temperature and then continuously agitated by using a pipette. This process was repeated thrice with FTA purification reagent and twice with 100 μl TE-buffer. Finally the entire unspent TE buffer was removed and discarded by pipetting and the disc was allowed to dry at room temp for overnight and was directly used for PCR amplification.

PCR amplification

Multiplexed PCR amplifications of the 9 STR loci: D16S539, D13S317, D7S820, CSF1PO, FGA, D21S11, D21S1338, D18S51 and amelogenin was performed using AmpFISTR® MiniFiler™ PCR amplification kit (Applied Biosystem, Foster city, CA, USA). The PCR reagents have been standardized in the laboratory for consistency of results. PCR was performed by taking the ½ reaction volume of the manufacturer’s recommended protocol [25] by using 9700 thermal
cycler (Applied Biosystems, USA). For one 1.2 mm washed punch of FTA paper the PCR mix was comprised of Reaction Buffer - 5.0 µL, Primers - 2.5 µL, MQ water - 5.0 µL to make final volume 12.5 µL.

Genotyping of amplified fragments

The PCR products were genotyped using multicapillary electrophoresis with POP-4 polymer in ABI Prism Avant 3100 Genetic Analyzer (Applied Biosystem, Foster city, CA, USA) according to the manufacturer’s protocol provided with the kit and the data was analyzed using Gene Mapper Software v3.5 (Applied Biosystem, Foster city, CA, USA) to designate alleles by comparison with the allelic ladder supplied with the kit. Peak detection threshold was set to 50 RFUs for allele designation. All steps were according to the laboratory internal standards and respective kit controls.

Analysis of the data

Allele frequencies of the 8 STR loci were calculated by GenAlEx 6.5 software [26]. Several Statistical parameters of forensic importance like the power of discrimination (PD), polymorphism information content (PIC), matching probability (PM) and power of exclusion (PE) were calculated using the Excel PowerStats spreadsheet program [27].

PD is the probability that two randomly chosen persons would not have matching DNA profiles [28] and CPD is used to prove that selected loci can be safely used to establish DNA based database for the studied population. CPM describes the possibility of finding two individuals with the same genotype in the population is almost null. Heterozygosity a measure of genetic variation within a population. High heterozygosity values for a breed may be due to long term natural selection for adaptation, to the mixed nature of the breeds or to historic mixing of strains of different populations. A low level of heterozygosity may be due to isolation with the subsequent loss of unexploited genetic potential. Observed heterozygosity is defined as the % of loci heterozygous per individual. Low observed heterozygosity values indicate inbreeding and may be a departure from HWE too. Observed and expected heterozygosities and Hardy-Weinberg equilibrium (HWE) using exact test were calculated using Arlequin software [29]. The same package of software was used for calculating AMOVA among the three endogamous studied populations. The genetic affinities among the studied three populations were observed by plotting dendogram using POPTREE software [30]. Nei’s genetic distances for the studied populations with other published Indian population data [31-33] was calculated by using POPTREE software [30] which were graphically summarized using Principal Component Analysis (PCA) plot generated using Past v3.02a software [34] to visualize population affinities.

Results and Discussion

The genetic variation in allele frequency distribution at 8 STR loci and statistical analysis of forensic parameters for Arora, Jat and Ramgarhi sikh population are shown in Tables 1, 2 and 3 respectively. The common pattern of allele distribution was observed at all the studied loci which may be due to their practice of endogamy as a social rule. The distribution of the most common allele (MCA) and least common allele (LCA) in the three endogamous caste is presented in Table 4, when the significance level was corrected by the Bonferroni method [35] and P values <0.05/8 = 0.00625 were considered statistically significant only two deviation persisted one at locus FGA for Ramgarhi sikh and other for locus D16S539 in Arora sikh. Molecular variance analysis (AMOVA) was conducted to understand the intra and interpopulation variations in the three Sikh endogamous populations (Figure 1).

![AMOVA](image1)

The genetic relation among them is shown in (Figure 2). Jat and Ramgarhi sikh showed close affinity than the Arora sikh. PCA plot (Figure 3) shows comparison of the three studied populations to other published population of India. The study shows that all the three populations showed significant differences from the tribal population. This finding is similar to the early reports [33-34] on caste and tribal population, which indicate that social stratification has played a major role in shaping the genetic diversity of India.

![Genealogy](image2)

The genetic relation among them is shown in (Figure 2). Jat and Ramgarhi sikh showed close affinity than the Arora sikh. PCA plot (Figure 3) shows comparison of the three studied populations to other published population of India. The study shows that all the three populations showed significant differences from the tribal population. This finding is similar to the early reports [33-34] on caste and tribal population, which indicate that social stratification has played a major role in shaping the genetic diversity of India.
Figure 3: PCA plot for comparison of studied populations with other published populations based on Nei’s genetic distance

| Allele/n | D13S317 | D7S820 | D2S1338 | D21S11 | D16S539 | D18S51 | CSF1PO | FGA |
|---------|---------|--------|---------|--------|---------|--------|--------|-----|
| 7       | 0.02    |        |         |        |         |        |        |     |
| 8       | 0.19    | 0.34   |         |        |         |        |        |     |
| 9       | 0.17    | 0.06   | 0.18    |        |         |        | 0.02   |     |
| 10      | 0.11    | 0.22   | 0.11    | 0.01   | 0.28    |        |        |     |
| 11      | 0.13    | 0.18   | 0.32    | 0.02   | 0.28    |        |        |     |
| 12      | 0.31    | 0.18   | 0.24    | 0.12   | 0.34    |        |        |     |
| 13      | 0.03    | 0.01   | 0.13    | 0.08   | 0.07    |        |        |     |
| 14      | 0.04    | 0.01   | 0.02    | 0.22   | 0.01    |        |        |     |
| 15      |         |        |         |        |         |        |        | 0.27|
| 16      |         | 0.02   |         |        |         | 0.12   |        |     |
| 17      |         | 0.07   |         |        |         | 0.07   |        |     |
| 18      |         | 0.17   |         |        |         | 0.03   | 0.02   |     |
| 19      |         | 0.19   |         |        |         | 0.01   | 0.06   |     |
| 20      |         | 0.14   |         |        |         |        | 0.13   |     |
| 21      |         | 0.08   |         |        |         | 0.02   | 0.09   |     |
| 22      |         | 0.07   |         |        |         | 0.02   | 0.11   |     |
| Allele/n | D13S317 | D7S820 | D2S1338 | D21S11 | D16S539 | D18S51 | CSF1PO | FGA |
|----------|---------|--------|---------|--------|---------|--------|--------|-----|
| 7        | 0.020   |        |         |        |         |        |        |     |
| 8        | 0.190   | 0.340  |         |        |         |        |        |     |
| 9        | 0.170   | 0.060  | 0.180   | 0.010  | 0.280   |        |        |     |
| 10       | 0.110   | 0.220  | 0.110   | 0.010  | 0.280   |        |        |     |
| 11       | 0.130   | 0.180  | 0.320   | 0.020  | 0.280   |        |        |     |
| 12       | 0.310   | 0.180  | 0.240   | 0.120  | 0.340   |        |        |     |
| 13       | 0.030   | 0.010  | 0.130   | 0.080  | 0.070   |        |        |     |
| 14       | 0.040   | 0.010  | 0.020   | 0.220  | 0.010   |        |        |     |
| 15       |         |        |         |        | 0.270   |        |        |     |
| 16       |         |        |         |        | 0.020   | 0.120  |        |     |
| 17       |         |        |         |        | 0.070   | 0.070  |        |     |
| 18       |         |        | 0.170   | 0.030  | 0.020   |        |        |     |

PM - matching probability, Power of Discrimination, PIC- Polymorphism Information Content, PE- Power of Exclusion, Ho- Observed Heterozygosity, He- Expected Heterozygosity, P-values- Hardy–Weinberg equilibrium (P< 0.00625)
| Allele/n | D13S317 | D7S820 | D2S1338 | D21S11 | D16S539 | D18S51 | CSF1PO | FGA |
|---------|---------|--------|---------|--------|---------|--------|--------|-----|
| 7       | 0.031   | 0.052  |         |        |         |        |        |     |
| 8       | 0.333   | 0.271  | 0.094   |        |         |        |        |     |
| 9       | 0.135   | 0.021  | 0.229   |        |         |        |        |     |
| 10      | 0.167   | 0.198  | 0.042   | 0.219  |         |        |        |     |
| 11      | 0.073   | 0.24   | 0.302   | 0.01   | 0.26    |        |        |     |
| 12      | 0.146   | 0.219  | 0.24    | 0.083  | 0.417   |        |        |     |
| 13      | 0.094   | 0.083  | 0.083   | 0.083  |         |        |        |     |
| 14      | 0.021   | 0.01   | 0.28    | 0.021  |         |        |        |     |

PM - matching probability, PD - Power of Discrimination, PIC - Polymorphism Information Content, PE - Power of Exclusion, Ho - Observed Heterozygosity, He - Expected Heterozygosity, P-values - Hardy-Weinberg equilibrium (P< 0.00625)

**Table 2:** Allele frequency distribution and forensic parameters for 8 autosomal STR loci investigated in Jat Sikh population of Central India
|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 15 |   |   | 0.281 |   |   |
| 16 |   |   | 0.135 |   |   |
| 17 | 0.073 | 0.219 | 0.063 |   |   |
| 18 | 0.25 |   | 0.042 |   |   |
| 19 | 0.125 |   | 0.021 | 0.063 |   |
| 20 |   | 0.052 | 0.01 | 0.063 |   |
| 21 | 0.021 |   | 0.01 | 0.115 |   |
| 22 |   | 0.135 |   | 0.167 |   |
| 23 | 0.073 | 0.042 |   | 0.26 |   |
| 24 |   | 0.01 |   | 0.146 |   |
| 25 |   |   | 0.063 |   |   |
| 26 | 0.01 | 0.115 |   | 0.021 |   |
| 27 |   | 0.208 |   |   |   |
| 28 |   | 0.01 |   |   |   |
| 29 | 0.219 |   | 0.021 |   |   |
| 29.2 |   | 0.01 |   |   |   |
| 30 |   | 0.01 |   |   |   |
| 31 |   | 0.052 |   |   |   |
| 31.2 |   | 0.021 |   |   |   |
| 32 |   | 0.021 |   |   |   |
| 32.2 |   | 0.313 |   |   |   |
| 33.2 |   | 0.042 |   |   |   |
| PM | 0.077 | 0.098 | 0.055 | 0.083 | 0.09 | 0.077 | 0.161 | 0.057 |
| PD | 0.923 | 0.902 | 0.945 | 0.917 | 0.91 | 0.923 | 0.839 | 0.943 |
| PIC | 0.782 | 0.743 | 0.821 | 0.764 | 0.748 | 0.791 | 0.652 | 0.834 |
| PE | 0.622 | 0.475 | 0.546 | 0.475 | 0.271 | 0.703 | 0.546 | 0.35 |
| Ho | 0.813 | 0.729 | 0.771 | 0.729 | 0.583 | 0.854 | 0.771 | 0.846 |
| He | 0.814 | 0.787 | 0.849 | 0.801 | 0.789 | 0.823 | 0.711 | 0.859 |
| P-value | 0.268 | 0.222 | 0.293 | 0.359 | 0.015 | 0.493 | 0.579 | 0.001 |

**Table 3:** Allele frequency distribution and forensic parameters for 8 autosomal STR loci investigated in Ramgariya Sikh population of Central India

**Conclusion**

The present study involves the development of forensic databases for indigenous populations in Central India. The autosomal STRs indicate general allelic homogeneity in the three endogamous populations and it can be suggested that a single database based on combined random samples may be sufficient in addressing the forensic needs.

**Acknowledgement**

Authors are thankful to Director, State Forensic Science Laboratory, Sagar, MP, India for providing genotyping facility. The study was...
supported by the I.M.B.I.B.E Award Fellowship grant from M.P. Biotechnology Council, Bhopal, India sanctioned to Devika Dogra.

Table 4: Distribution of the most common allele (MCA) and least common allele (LCA) in the three endogamous Sikh populations of Central India

| Locus  | Arora Sikh | Jat Sikh | Ramgarhiya Sikh |
|--------|------------|----------|-----------------|
|        | MCA        | LCA      | MCA             | LCA            | MCA   | LCA   |
| CSF1PO | 11         | 9        | 12              | 14             | 12    | 14    |
| D2S1338| 19         | 16,26    | 19              | 26             | 19    | 27    |
| D7S820 | 8          | 7,13     | 8               | 13,14          | 8     | 9     |
| D13S317| 12         | 15,16    | 12              | 7              | 8     | 14    |
| D18S51 | 13,15      | 9        | 15              | 10,19,23       | 15    | 11,21,22 |
| D21S11 | 28         | 27,29,2,31 | 32,2           | 31             | 32,2  | 27,29,2,31 |
| D16S539| 11         | 10       | 11              | 14             | 11    | 14    |
| FGA    | 23,24      | 18       | 24              | 18,26,30       | 24    | 27,30 |

References

1. Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res 12: 4127-4138.
2. Weber JL, May PE (1989) Abundant class of Human DNA polymorphisms which can be typed using the polymerase chain reaction. Amer J Human Genet 44: 388-396.
3. Deka R, Shriver MD, Yu LM, Ferrell RE, Chakraborty R (1995) Population genetic study of three VNTR loci (D2S44, D7S2 and D12S11) in five ethnically defined populations of the Indian subcontinent. Hum Biol 68: 819-835.
4. Mukherjee N, Majumder PP, Roy B, Roy M, Dey B, et al. (1999) Genetic variation at twenty-three microsatellite loci in sixteen human populations. J Genet 78: 99-121.
5. Mountain JL, Hebert JM, Bhattacharyya S, Underhill PA, Ottoleghi C, et al. (1995) Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56: 979-992.
6. Papia SS, Mastana SS, Purandare CA, Jayasekara R, Chakraborty R (1996) Population genetic study of three VNTR loci (D2S44, D7S2 and D12S11) in five ethnically defined populations of the Indian subcontinent. Hum Biol 68: 819-835.
7. Singh P, Singh M, Bhatnagar DP, Kaur T, Mastana S (2003) Apolipoprotein C3 (SstI) Gene variability in Northwest India: A Global Perspective. Int J Hum Genet 8: 51-60.
8. Chattopadhyay PK, Cleef S, Duelmer M, Henke L, Henke J (2004) Human Y-chromosome short tandem repeat loci in four population groups of India. J Forensic Sci 49: 111-113.
9. Babita K, Usha D (2003) HLA Antigen distribution in Sikhs from Punjab, India. Int J Hum Genet 3: 45-50.
10. Babita K, Usha D (2004) HLA Antigen distribution in Sikhs from Punjab, India. Int J Hum Genet 4: 111-113.
11. Feng H, He L, Morris S, Lennon A, St Clair DM, Porteous DJ, et al. (1996) Genome-wide scan to identify loci associated with type 2 diabetes and blood lipid phenotypes in the Sikh Diabetes Study. PLoS One 6: e21188.
12. He L, Morris S, Lennon A, St Clair DM, Porteous DJ, et al. (1996) A genome-wide search for linkage in a large bipolar family: comparison of genotyping accuracy using di-and tetranucleotide repeat microsatellite markers. Psychiatr Genet 6: 123-129.
26. Peakall R, Smouse PE (2006) GENALEX6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes 6: 288-95.
27. Tereba A (1999) Tools for analysis of population statistics. Profiles DNA 2: 14-16.
28. Fisher RA (1951) Standard calculations for evaluating a blood-group system. Heredity (Edinb) 5: 95-102.
29. Excoffier L, Laval G, Schneider S (2007) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1: 47-50.
30. Takezaki N, Nei M, Tamura K (2010) POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol Biol Evol 27: 747-752.
31. Ghosh T, Kalpana D, Mukerjee S, Mukherjee M, Sharma AK, et al. (2011) Genetic diversity of autosomal STRs in eleven populations of India. Forensic Sci Int Genet 5: 259-261.
32. Dubey B, Meganathan PR, Easwarkhanth M, Vasulu TS, Haque I (2009) Forensic STR profile of two endogamous populations of Madhya Pradesh, India. Leg Med (Tokyo) 11: 41-44.
33. Easwarkhanth M, Dubey B, Ramakodi Meganathan P, Noor S, Haque I (2009) Microsatellite diversity delineates genetic relationships of Shia and Sunni Muslim populations of Uttar Pradesh, India. Hum Biol 81: 427-445.
34. Hammer O, Harper DAT, Ryan PD (2001) PAST: Palaeontological Statistics software package for education and data analysis. Palaeontologia Electronica 4: 9.
35. Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. BMJ 310: 170.