Int J Hum Genet, 17(4): 151-157 (2017)
DOI:10.31901/24566330.2017/17.04.02

# Genetic Ancestry of Delhi Population Inferred from Autosomal Short Tandem Repeats: Genetic Diversity Analysis

**Bhuvnesh Yadav[1], Ajay Balayan[2], T. D. Dogra[3] and Anupuma Raina[2*]**

*[1]Amity School of Applied Sciences, Amity University, Gurgaon, Haryana*
*[2] DNA Fingerprinting Laboratory, Department of Forensic Medicine and Toxicology, All India Institute of Medical Sciences, New Delhi, India*
*[3]Shree Guru Gobind Singh Tricentenary University, Budhera, Gurgaon, Haryana, India*

**KEYWORDS** Genetic Variation. Population Data. Short Tandem Repeats

**ABSTRACT** Population substructure analysis and ancestry tracing are the critical issues for association studies of health, behaviors and forensic genetics. STR (short tandem repeat) markers are being extensively used to analyze genetic diversity among the populations. In the present study, allele frequencies and statistical parameters were estimated with 15 STR loci from 208 unrelated individuals from Delhi (India). A total of 146 alleles was found with corresponding allelic frequencies ranging from 0.001 to 0.3869. The MP (matching probability), PD (power of discrimination), PIC (polymorphism information content), PE (power of exclusion) and TPI (typical paternity index) ranged from 0.035 to 0.146, 0.854 to 0.965, 0.65 to 0.850, 0.416 to 0.774, and 1.76 to 4.52, respectively. Deviations were observed from the Hardy-Weinberg Equilibrium for D16S539, D18S51, D21S11 and TPOX Markers. The genetic proximity of the studied population was observed with central Indian population. The population's genetic structure analysis of this population can assist to contrive future genetic studies.

## INTRODUCTION

Short tandem repeats are repeated after every ten thousand nucleotides and constitute three percent of the total genome (Butler 2015). Their widespread distribution, relatively low incidence of mutation, ease of multiplexing and high statistical capability of discrimination and individualization made these markers routine for forensic, anthropological and medical studies since last two decades (Osman et al. 2015). These markers are suitable for analyzing degraded, contaminated and minute amounts of human DNA samples (Jha et al. 2012; Brinkmann 1992). Throughout history, society had been ranked on the basis of caste, race, region, ethnicity, gender, age and socioeconomic status. It is ethnicity and racial discrimination that distinguishes one nation from the other (Tamang et al. 2012). The STR based population data is increasing with add-on numbers of laboratories employing this technology (Narkuti et al. 2008; Dubey et al. 2009; Ghosh et al. 2011; Chaudhari and Dahiya 2014; Preet et al. 2016; Yadav et al. 2016; Shrivastava et al. 2017). The studies have been conducted on Delhi random and Rajput population (Chauhan et al. 2015, Raina et al. 2009) and the allelic frequencies were observed for these populations.

The present study was accomplished to determine the allele frequencies for 15 short tandem repeats (STR) loci, forensic and genetic diversity parameters from 208 individuals undergoing paternity testing from the capital region of India. The National Capital Territory of Delhi is located at 28.61°N 77.23°E and lies in northern India. It is the largest city in the country as it covers an area of 1484 $km^2$, of which approximately fifty-three percent area is designated as rural, and rest forty-seven percent is urban (Census of India 2011). Delhi is the second most populated city in the world with 18.25 million people with 79.8 percent Hindu population (Statistical Abstract of Delhi 2016). Approximately 2-3 lakh people settle in Delhi permanently from other states of India per year due to higher employment and education opportunities.

*Address for correspondence*:
Dr. Anupuma Raina
*Scientist,*
DNA Fingerprinting Laboratory,
Department of Forensic Medicine & Toxicology,
All India Institute of Medical Sciences,
New Delhi, India
*Telephone*: 9810462033
*E-mail*: anupumaraina@gmail.com

### Objectives

The aim of this study was two-fold, one was to determine the genetic structure of Delhi population and the second objective of the study was to evaluate the importance of STR loci for forensic genetic analysis purposes.

## METHODOLOGY

### Samples

A total of 208 unrelated (as stated in their identification form) individual casework samples were collected as the source of the study samples.

### DNA Extraction, PCR Amplification and Genotype Determination

The organic extraction method was utilized to extract DNA from blood samples (Sambrook et al. 1989). DNA amplification was carried out by using the Amp FlSTR® Identifiler® PCR Amplification Kit containing 15 Autosomal STR markers (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, THO1, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 and FGA) with amelogenin marker for gender determination. DNA purification, PCR amplification and genotype determination were carried out by following the standard protocols provided by Applied Biosystems (Applied Biosystems 1998).

### Data Analysis

The statistical parameters of population genetics and forensic utility of these markers was evaluated by calculating the allele frequencies, matching probability (MP), polymorphism information content (PIC), typical paternity index (TPI), power of discrimination (PD) and power of exclusion (PE) using Powerstats software version 1.2 (Promega Corporation 2009). The population's genetic structure deviation from Hardy-Weinberg equilibrium (p), observed heterozygosity (Ho) and expected heterozygosity (He) were calculated using methods implemented in the Arlequin V3.5 software (Excoffier et al. 2005).

### Quality Control

Allelic ladder and positive control DNA provided along with AmpFlSTR® Identifiler® PCR Amplification kit were used for quality control.

## RESULTS

Allele frequencies for the 15 STR loci in the AmpFlSTR® Identifiler® and statistical parameters were estimated from a sample of 208 unrelated individuals from Delhi. The observed and expected allele frequencies and statistical parameters based on the 15 STR loci in Delhi population are summarized in Table 1. The allelic variation was observed from 6 (THO1, TPOX, and D5S818) to 17 (D21S11, D18S51 and FGA). These 15 STR markers are found to be highly polymorphic in the populations. A total of 146 alleles was observed at 15 STR loci with accompanying allelic frequencies ranging from 0.0028 to 0.3966 in the Delhi population (Table 2). The number of alleles varied from 7 (CSFIPO) to 14 (D18S51). Though, the number of alleles was high, the number of effective alleles was com-

**Table 1: The observed and expected heterozygosity in the Delhi population of North India**

| *Locus* | *Sample size* | *Obs_Hom* | *Obs_Het* | *Exp_Hom** | *Exp_Het** | *Nei*** | *Ave_Het* |
|---|---|---|---|---|---|---|---|
| D8S1179 | 208 | 0.1285 | 0.8715 | 0.151 | 0.849 | 0.8466 | 0.8466 |
| D21S11 | 208 | 0.162 | 0.838 | 0.1351 | 0.8649 | 0.8625 | 0.8625 |
| D7S820 | 208 | 0.162 | 0.838 | 0.1929 | 0.8071 | 0.8048 | 0.8048 |
| CSF1PO | 208 | 0.3296 | 0.6704 | 0.2852 | 0.7148 | 0.7128 | 0.7128 |
| D3S1358 | 208 | 0.2402 | 0.7598 | 0.2398 | 0.7602 | 0.7581 | 0.7581 |
| THO1 | 208 | 0.2011 | 0.7989 | 0.1911 | 0.8089 | 0.8067 | 0.8067 |
| D13S317 | 208 | 0.1508 | 0.8492 | 0.1841 | 0.8159 | 0.8136 | 0.8136 |
| D16S539 | 208 | 0.1955 | 0.8045 | 0.19 | 0.81 | 0.8078 | 0.8078 |
| D2S1338 | 208 | 0.1061 | 0.8939 | 0.1299 | 0.8701 | 0.8677 | 0.8677 |
| D19S433 | 208 | 0.162 | 0.838 | 0.1641 | 0.8359 | 0.8336 | 0.8336 |
| VWA | 208 | 0.2346 | 0.7654 | 0.2212 | 0.7788 | 0.7766 | 0.7766 |
| TPOX | 208 | 0.2793 | 0.7207 | 0.266 | 0.734 | 0.732 | 0.732 |
| D18S51 | 208 | 0.2291 | 0.7709 | 0.1586 | 0.8414 | 0.839 | 0.839 |
| D5S818 | 208 | 0.2793 | 0.7207 | 0.3008 | 0.6992 | 0.6972 | 0.6972 |
| FGA | 208 | 0.1955 | 0.8045 | 0.1389 | 0.8611 | 0.8587 | 0.8587 |

**Table 2: The allelic frequency and forensic parameters for the Delhi population of North India**

| *Allele* | *D8S1179* | *D21S11* | *D7S 820* | *CSFIPO* | *D3S1358* | *THO1* | *D13S317* | *D16S539* | *D2S1338* | *D19S433* | *VWA* | *TPOX* | *D18S51* | *D5S818* | *FGA* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4 | | | | | | 0.0028 | | | | | | | | | |
| 5 | | | | | | 0.0028 | | | | | | 0.0112 | | | |
| 5.3 | | | | | | 0.0084 | | | | | | | | | |
| 6 | | | | | | 0.2626 | | | | | | 0.3869 | | 0.0028 | |
| 6.3 | | | | | | 0.0112 | | | | | | | | | |
| 7 | 0.006 | | 0.0251 | | | 0.1508 | 0.0056 | 0.0028 | | | | 0.1872 | | 0.0056 | |
| 7.3 | | | | | | 0.014 | | | | | | | | | |
| 8 | 0.003 | | 0.1844 | 0.0084 | | 0.1173 | 0.1676 | 0.0726 | | | | 0.0978 | | 0.0084 | |
| 8.3 | | | | | | 0.0168 | | | | | | | | | |
| 9 | 0.008 | | 0.0866 | 0.0196 | | 0.2486 | 0.1425 | 0.1257 | | 0.0028 | | 0.2709 | | 0.0168 | |
| 9.3 | | | | | | 0.1592 | | | | | | | | | |
| 10 | 0.162 | | 0.2291 | 0.2821 | | 0.0056 | 0.1117 | 0.0782 | | 0.0084 | 0.0028 | | | 0.0587 | |
| 10.2 | | | 0.0028 | 0.0056 | | | | | | 0.014 | | | | | |
| 11 | | | 0.2598 | 0.2961 | | | 0.2039 | 0.2877 | | 0.0056 | | | 0.0112 | 0.3408 | |
| 11.2 | | | | | | | | | | 0.0223 | | | | | |
| 12 | 0.101 | | 0.1788 | 0.3436 | 0.014 | | 0.2793 | 0.2179 | | 0.0587 | 0.0056 | 0.0335 | 0.0531 | 0.3966 | |
| 12.2 | | | | | | | | 0.0028 | | 0.0056 | | | | | |
| 13 | 0.07 | | 0.0335 | 0.0363 | 0.0112 | | 0.0754 | 0.1844 | | 0.2626 | 0.0056 | | 0.1397 | 0.1592 | |
| 13.2 | | | | | | | | | | 0.0223 | | | | | |
| 14 | 0.184 | | | 0.0084 | 0.0391 | | 0.014 | 0.0279 | | 0.2402 | 0.1089 | | 0.2961 | 0.0112 | |
| 14.2 | | | | | | | | | | 0.0559 | | | | | |
| 15 | 0.212 | | | | 0.3464 | | | | | 0.1564 | 0.0419 | | 0.1536 | | |
| 15.2 | | | | | | | | | | 0.0698 | | | | | |
| 16 | 0.17 | | | | 0.2598 | | | | 0.0112 | 0.0391 | 0.2486 | | 0.1145 | | |
| 16.2 | | | | | 0.0028 | | | | | 0.0307 | | | | | |
| 17 | 0.062 | | | | 0.1983 | | | | 0.0642 | 0.0056 | 0.338 | | 0.1061 | | |
| 18 | 0.011 | | | | 0.1145 | | | | 0.1425 | | 0.1704 | | 0.0279 | | |
| 19 | 0.008 | | | | 0.0112 | | | | 0.2011 | | 0.067 | | 0.0391 | | 0.0112 |
| 19.2 | | | | | | | | | | | | | | | 0.0028 |
| 20 | 0.003 | | | | 0.0028 | | | | 0.0866 | | 0.0112 | | 0.014 | | 0.1145 |
| 20.2 | | | | | | | | | | | | | | | |
| 21 | | | | | | | | | 0.0391 | | | | 0.014 | | 0.1536 |
| 21.2 | | | | | | | | | | | | | | | 0.0084 |
| 22 | | | | | | | | | 0.0922 | | | | 0.0056 | | 0.1257 |
| 22.2 | | | | | | | | | | | | | | | 0.0084 |
| 23 | | | | | | | | | 0.1927 | | | | 0.0028 | | 0.2067 |
| 23.2 | | | | | | | | | | | | | | | 0.0056 |
| 24 | | | | | | | | | 0.0894 | | | | | | 0.1788 |
| 24.2 | | | | | | | | | | | | | | | 0.0168 |
| 25 | | | | | | | | | 0.067 | | | | | | 0.1089 |
| 25.2 | | | | | | | | | | | | | | | 0.0028 |

**Table 2: Contd....**

| *Allele* | *D8S1179* | *D21S11* | *D7S 820* | *CSFIPO* | *D3S1358* | *THO1* | *D13S317* | *D16S539* | *D2S1338* | *D19S433* | *VWA* | *TPOX* | *D18S51* | *D5S818* | *FGA* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 26 | | 0.0084 | | | | | | | 0.014 | | | | | | 0.0391 |
| 26.2 | | 0.0084 | | | | | | | | | | | | | |
| 27 | | 0.0196 | | | | | | | | | | | | | 0.0084 |
| 27.2 | | 0.0028 | | | | | | | | | | | | | |
| 28 | | 0.1173 | | | | | | | | | | | | | |
| 29 | | 0.2067 | | | | | | | | | | | | | |
| 29.2 | | 0.0028 | | | | | | | | | | | | | |
| 30 | | 0.1844 | | | | | | | | | | | | | |
| 30.2 | | 0.0363 | | | | | | | | | | | | | 0.0028 |
| 31 | | 0.0475 | | | | | | | | | | | | | |
| 31.2 | | 0.0978 | | | | | | | | | | | | | |
| 32.2 | | 0.1564 | | | | | | | | | | | | | 0.0056 |
| 33.2 | | 0.0922 | | | | | | | | | | | | | |
| 34.2 | | 0.0196 | | | | | | | | | | | | | |
| N | 208 | 208 | 208 | 208 | 208 | 208 | 208 | 208 | 208 | 208 | 208 | 208 | 208 | 208 | 208 |
| Na | 13 | 9 | 8 | 7 | 10 | 9 | 8 | 9 | 11 | 9 | 10 | 8 | 14 | 10 | 11 |
| Ne | 6.704 | 6.146 | 5.024 | 3.432 | 4.177 | 3.53 | 5.439 | 5.217 | 7.522 | 4.357 | 4.755 | 3.808 | 5.995 | 3.382 | 6.592 |
| I | 2.049 | 1.922 | 1.731 | 1.357 | 1.599 | 1.415 | 1.8 | 1.782 | 2.159 | 1.662 | 1.743 | 1.511 | 2.048 | 1.43 | 2.003 |
| F | -0.04 | -0.005 | -0.056 | 0.023 | -0.018 | -0.114 | -0.043 | -0.011 | -0.026 | -0.073 | 0.014 | 0.015 | 0.071 | -0.017 | 0.054 |

N-Sample Size, Na- No. Alleles, No. Ne- Effective Alleles, I- Information Index, F- Fixation Index

paratively low for a few loci, that is, TH01, D5S818 and D3S1358. The expected heterozygosity was significantly high in case of D8S1179, FGA, D2S1338 whereas it was quite low for D5S818, CSFIPO. The expected heterozygosity and the power of discrimination calculated from allele frequencies obtained from Delhi population revealed that in combination, the 15 autosomal STRs have a high forensic efficacy. There was not much variation between observed and expected heterozygosity. The MP, PD, PIC, PE and TPI ranged from 0.035 to 0.146, 0.854 to 0.965, 0.65 to 0.850, 0.416 to 0.774, and 1.76 to 4.52, respectively. Nei's unbiased genetic distances (Nei 1978) were measured and the population was compared with previously studied Indian populations for genetic distance (Ashma and Kashyap 2002; Sarkar and Kashyap 2002; Tandon et al. 2002; Agrawal and Khan 2005; Mohapatra et al. 2004; Mohpatra et al. 2016; Yadav et al. 2016; Shrivastava et al. 2017). The clustering of the Delhi population was observed with central Indian populations as the least genetic distance was observed in these populations.

## DISCUSSION

Great variation in the total number of alleles and effective number of alleles was observed in the studied population. The higher value of combined matching probability in the present work also proved to be suitable for forensic applications. The observed and expected heterozygosity was comparable with the previously studied Delhi population (Raina et al. 2009) but variation was observed from Delhi's Rajput population (Chauhan et al. 2015). When compared to the previously studied Indian populations, it was observed that the population is having least genetic distance with Madhya Pradesh's population (Sarkar and Kashyap 2002) and the highest genetic distance was observed from the Patel community of Gujarat (Mohapatra et al. 2004). The variation was revealed by computing DA distances and represented in the form of the Neighbor Joining Tree (Fig.1). Among the studied populations, Delhi's population was found clustered along with Central Indian population (Sarkar and Kashyap 2002). However, the tribal

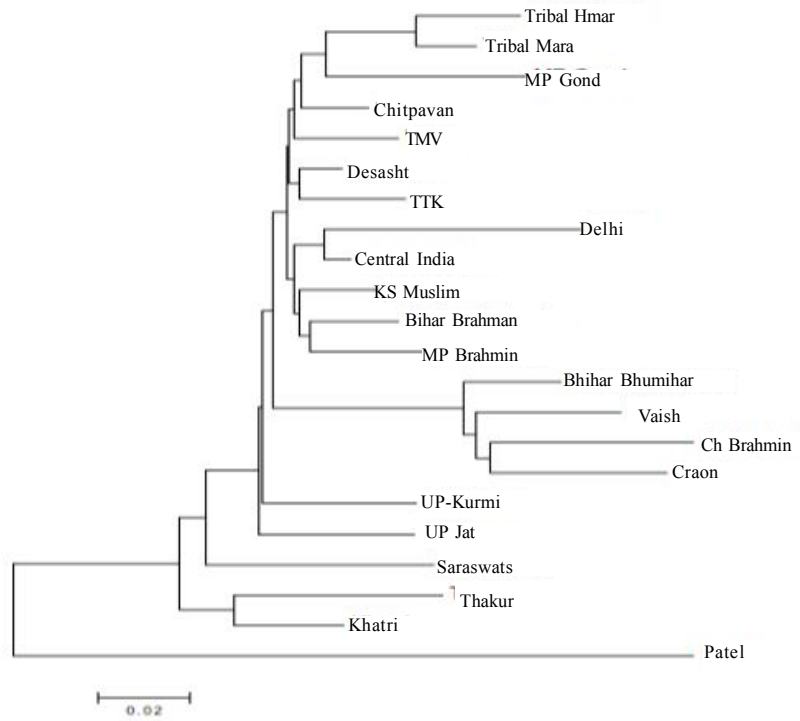


**Fig. 1. Phylogenetic relationships of Delhi Population with other Indian populations conducted with MEGA6**

populations were far apart from this population (Shrivastava et al. 2017). Populations from Uttar Pradesh (Kurmi, Jat, Thakur) (Tandon et al. 2002) and Saraswat Brahmins (Yadav et al. 2016) and Kasmiri Muslims (Mohpatra et al. 2016) were observed to have comparatively high genetic distance from Delhi population, indicating less genetic mixing in these populations. The Delhi population was observed to be far away from Patels of Gujarat as weighted against other populations.

## CONCLUSION

The STR markers used in the current study are highly polymorphic for the population and these can be used in forensic identity establishment. The population has significant diversity and genetic affinity was observed with central Indian population. The population has higher genetic distance, to neighbouring states, indicating less genetic admixture.

## RECOMMENDATIONS

The studied population database can be used for further genetic diversity analysis. The Delhi population may be further analyzed by dividing it in the subgroups.

## REFERENCES


Agrawal S, Khan F 2005. Reconstructing recent human phylogenies with forensic STR loci: A statistical approach. *BMC Genetics,* 6: 47.

*AmpFLSTR Profiler Plus PCR Amplification Kit User's Manual*. Foster City, CA: PE Applied Biosystems, 1998.

Ashma R, Kashyap VK 2002. Genetic study of 15 important STR loci among four major ethnic groups of Bihar, India. *J Forensic Sci USA,* 47: 1139-1142.

Brinkmann B 1992. The Use of STRs in Stain Analysis. In: *Proceedings From The Third International Symposium on Human Identification*. Promega Corporation, Madison, USA, pp. 357–373.

Butler JM 2015. *Advanced Topics in Forensic DNA Typing: Interpretation*. San Diego: Elsevier Academic Press.

Census of India 2011. From <http://censusindia.gov.in> (Retrieved on 4 December 2016).

Chaudhari RR, Dahiya MS 2014. Genetic diversity of 15 autosomal short tandem repeats locus using the AmpFLSTR® Identifiler™ kit in a Bhil tribe population from Gujarat state, India. *Indian J Hum Genet,* 20(2): 148–152.

Chauhan T, Kushwaha KPS, Chauhan V 2015. Genetic Polymorphism of Eleven STR Loci in Rajput Population of Delhi, India. *Forensic Res Criminol Int J*, 1(5): 00031.

Dubey B, Meganathan PR, Eaaswarkhanth M, Vasulu TS, Haque I 2009. Forensic STR profile of two endogamous populations of Madhya Pradesh, India. *Legal Med,* 11(1): 41–44.

Excoffier L, Laval G, Schneider S 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online,* 1: 47–50.

Gaikwad S, Kashyap VK 2002. Polymorphism at fifteen hypervariable microsatellite loci in four populations of Maharashtra, India. *Forensic Sci Int,* 126: 267-271.

Ghosh T, Kalpana D, Mukerjee S, Mukherjee M, Sharma AK, Nath S, Rathod VR, Thakar MK, Jha GN 2011. Genetic diversity of autosomal STRs in eleven populations of India. *Forensic Sci Int Genet,* 5(3): 259–261.

Jha DK, Gonzalez LJM, Rijal JP, Tuladhar BS, Chhetri NT 2012. Allele frequencies of 15 AmpFLSTR Identifiler loci in the Nepalese population. *Scientific World,* 10(10): 20-23.

Mohapatra BK, Trivedi R, Mehta AK, Vyas JM, Kashyap VK 2004. Genetic diversity at 15 fluorescent-labeled Short Tandem Repeat loci in the Patel and other communities of Gujarat, India. *Am J For Med and Pathol,* 25(2): 108-112.

Mohapatra BK, Chauhan K, Thakur US, Yadav B, Raina A 2016. Genetic analysis and evolutionary relationship of Jammu & Kashmir Muslim population with Short Tandem Repeat Loci. *Int J Current Res*, 8(8): 36398-36401.

Narkuti V, Vellanki RN, Oraganti NM, Mangamoori LN 2008. Paternal exclusion: Allele sharing in microsatellite testing. *Clin Chem Lab Med,* 46(11): 1586–1588.

Nei M 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics,* 89: 583-590.

Osman AE, Alsafar H, Tay GK, Theyab JB, Mubasher M et al. 2015. Autosomal Short Tandem Repeat (STR) variation based on 15 loci in a population from the Central Region (Riyadh Province) of Saudi Arabia. *J Forensic Res,* 6: 1.

Preet K, Malhotra S, Shrivastava P, Jain T, Rawat S et al. 2016. Genetic diversity in Gorkhas: An autosomal STR Study. *Scientific Reports,* 6: 32494.

Promega Corporation Website 2009. Powerstats version 1.2. From <http://www.promega.com/geneticidtools/powerstats> (Retrieved on 10 December 2016).

Raina A, Yadav B**,** Bhat KV, Dogra TD 2009. Genetic Polymorphism at 15 STR Loci in Delhi's (India) population. *Internet J Indian Congress Forensic Med Toxicol*, 7: 1.

Sambrook J, Fritch EF, Maniatis T 1989. *Molecular Cloning. A Laboratory Manual.* 2nd Edition. New York, NY: Cold Spring Harbour Laboratory.

Sarkar N, Kashyap VK 2002. Genetic diversity at two pentanucleotide and thirteen tetranucleotide STR loci by multiplex PCR in four predominant population groups of Central India. *Forensic Sci Int UK,* 128: 196–201.

Shrivastava P, Jain T, Trivedi VB 2015. Genetic polymorphism study at 15 autosomal locus in central Indian population. *Springerplus,* 4: 566.

Shrivastava P, Jain T, Trivedi VB 2017. Structure and genetic relationship of five populations from central India based on 15 autosomal STR loci. *Ann Hum Biol,* 44(1): 74-86

Statistical Abstract of Delhi 2016. Government of National Capital Territory of Delhi Directorate of Economics and Statistics. From <http://www. delhi. gov.in> (Retrieved on 2 October 2017).

Tamang R, Thangaraj K 2012. Genomic view on the peopling of India. *Investigative Genet,* 3(20): 2014-2223.

Tandon M, Trivedi R, Kashyap VK 2002. Genomic diversity at fifteen fluorescent labelled short tandem repeat loci in few important populations of states of Uttar Pradesh, India. *Forensic Sci Int UK,* 123: 190–195.

Yadav B, Raina A, Dogra TD 2016. Genetic proximity of Saraswat Brahmin community of northern India based on autosomal STR markers. *J Bio Innovations,* 5(5): 764-785.

**Paper received for publication on December 2016**
**Paper accepted for publication on October 2017**