Int J Legal Med (2010) 124:327–330
DOI 10.1007/s00414-010-0456-2

TECHNICAL NOTE

Genetic variation of 10 X chromosomal STR loci in Indian population

Sanjukta Mukerjee · Tania Ghosh · D. Kalpana · Meeta Mukherjee · Anil Kumar Sharma

Received: 8 January 2010 / Accepted: 31 March 2010 / Published online: 27 April 2010

Abstract X chromosomal short tandem repeats have the potential to complement the analyses of the autosomal, Y chromosomal, and mitochondrial DNA markers in forensics and population genetics, and extensive research on X chromosomal markers is being carried out. In the present study, a decaplex for the co-amplification of ten X chromosomal microsatellite loci (DXS6807, DXS8378, DXS7132, DXS6809, DXS6789, DXS101, DXS7133, GATA172D05, HPRTB, and GATA31E08) was optimized and 749 blood samples of unrelated male individuals from the four major linguistic families of India were analyzed. The number of alleles for the studied loci ranged from 7–16 while the gene diversity values varied from 0.408 to 0.855. Two new alleles were observed for the loci DXS101 and HPRTB. Statistical parameters of forensic interest were calculated and all loci were found to be polymorphic. High power of discrimination was observed for the loci DXS101, DXS6809, and DXS6789. The present study demonstrates the efficacy of these X-linked markers for human identification and kinship analysis.

Keywords X chromosome · Short tandem repeats · Forensic markers · Indian populations · Linguistic families

Introduction

In the recent years, X chromosome short tandem repeats (X-STRs) have drawn attention due to their utility in complex kinship cases, deficiency paternity cases, and identifying the female traces in mixtures of biological fluids [1]. In a father/daughter or mother/son parentage determination scenario, X chromosome markers are more effective than autosomal due to high mean exclusion chance. X chromosome markers are also efficient in the detection of female traces in mixtures with predominant male contribution [2]. Study of X-STR polymorphism in Indian population would enable the use of these markers for forensic casework.

India is a land of genetic, cultural, linguistic, and biological diversity. The biological diversity is mainly due to its location, being placed at the tri-junction of the African, the Northern Eurasian, and the Oriental realm [3]. Indian population can be sub-structured on the basis of ethnicity, linguistic lineages, and social hierarchy. The four major linguistic families of India are Indo-European (IE), Dravidian (DR), Tibeto-Burman (TB), and Austro-Asiatic (AA). The social structure includes castes and tribes. People belonging to various linguistic families inhabit different geographical regions of the country and display considerable genetic diversity [4]. In the present study, ten polymorphic X-STR markers were analyzed in 11 endogamous populations of India. The samples were collected from unrelated males representing all the major linguistic families and geographical regions of the country.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-010-0456-2) contains supplementary material, which is available to authorized users.

S. Mukerjee · T. Ghosh · D. Kalpana · M. Mukherjee · A. K. Sharma (✉)
Directorate of Forensic Science, Ministry of Home Affairs, Govt. of India, Central Forensic Science Laboratory,
30 Gorachand Road,
Park Circus, Kolkata 700014, India
e-mail: aks4n6@gmail.com

Materials and methods

Sample collection

Seven hundred and forty-nine blood samples from unrelated healthy male individuals were collected with informed consent following protocols approved by Institutional Ethics Committee. Blood samples from Balmiki (62), Sakaldwipi Brahmin (65), Kanyakubja Brahmin (78), Konkanastha Brahmin (71), and Mahadev Koli (65) populations belonging to IE group; Iyengar (66), Kurumans (67), and Gond (75) populations belonging to DR group; Tripuri (65) and Riang (67) populations of TB group; and Munda (68) population belonging to AA group were collected from six states of India as summarized in Table 1 (supplementary data).

DNA isolation and quantitation

Genomic DNA was isolated from the blood samples by the phenol-chloroform organic extraction method. DNA was quantified using Quantifiler® Human DNA Quantification Kit according to the manufacturer's protocol employing 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA).

Amplification and genotyping

Approximately 1–2 ng of DNA was amplified for ten microsatellite loci (DXS6807, DXS8378, DXS7132, DXS6809, DXS6789, DXS101, DXS7133, GATA172D05, HPRTB, and GATA31E08). Multiplex amplification was performed in a single PCR in GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, CA) following the protocol of Turrina et al. [5]. The experiments were conducted in accordance with quality control measures and using cell lines 9947A, 9948, and K562 from Promega Corp. (Promega, Madison, USA) as standard reference materials [6]. The amplified products were electrophoresed on an ABI PRISM® 3100 Genetic Analyzer with POP4 polymer using 1 µl of the amplified product, 8.5 µl of formamide, and 0.5 µl of GeneScan™ LIZ-500 size standard (Applied Biosystems, Foster City, CA). The data was analyzed with GeneMapper® ID Software v3.2 (Applied Biosystems, Foster City, CA). Genotypes for the cell lines were recorded according to the recommendations of Szibor et al. and Gusmao et al. [6, 7].

Table 1 Percentage of total variation at three levels of population hierarchy in different groups of Indian populations

Population group	Number of groups	Among groups	Percentage of variation	
			Among populations within groups	Within populations
Linguistic groups (IE, DR, TB, AA)	4	1.05	1.38	97.57
Geographical regions (north, central, east, northeast, west, south)	6	0.77	1.47	97.76

Sequencing of alleles

Three to six alleles of each locus were sequenced for confirmation of repeat number and structure. In this study, new alleles were observed for the loci DXS101 and HPRTB. Few rare intermediate alleles were also observed for the locus HPRTB. The repeat number and structure of the alleles at these loci were determined as described previously [8, 9]. The alleles were confirmed by bidirectional sequencing using Big Dye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) and published primers.

Statistical analysis

Allele frequencies for the ten X-STR loci were calculated by Genepop program [10]. Arlequin program v3.1 was used to compute the gene diversity, p values of exact test for linkage disequilibrium, population pairwise genetic distances (F_{ST} and R_{ST}), and analysis of molecular variance (AMOVA) among the studied populations [11]. For the evaluation of forensic efficiency of the decaplex, various statistical parameters were calculated with Chromosome X website software (http://www.chrx-str.org) [12].

Results and discussion

Allele frequency data for the 10 X-STRs in 11 endogamous populations of India is reported in Table 2 (supplementary data). The number of alleles ranged from 7–16 for the studied loci and 91 alleles in total were observed. A new allele 35 was observed for the locus DXS101 in a sample of Iyengar population. Alleles 31, 32, and 35 observed for this locus in the Indian population have not been reported in the Asian populations [13–16]. Another new allele 10.2 for the locus HPRTB was observed in a sample of Kurumans population. Rare intermediate alleles 11.2 and 12.2 for HPRTB locus were observed in few samples of the IE, DR, and AA populations. These intermediate alleles have also been reported in the German, Italian, Spanish, and Thai populations [17–20]. All populations showed high frequency for allele 11 of the locus DXS6807. Allele 11 of the locus DXS7132 was observed exclusively in three tribal

populations. Allele 8 of the locus DXS8378 was observed only in the Balmiki population of Punjab, similar to the observation for the Punjabi population of Pakistan [21]. For the locus DXS7133, TB populations showed lesser frequency of allele 11 than other populations where it was frequent. The exact test for linkage disequilibrium was performed for all pairs of the 10 X-STR loci. The marker pair DXS6809 and DXS6789 belongs to the haplotype cluster DXS6801-DXS6809-DXS6789 on Xq21 and is reported to be in linkage disequilibrium. Haplotyping of this STR cluster can provide useful information for kinship analysis [22]. For the 45 pairwise comparisons, exact test *p* values were above the significance level of 0.0011 (after Bonferroni correction) indicating that no association among the ten markers could be detected in the studied populations. Haplotype frequencies for the markers DXS6809 and DXS6789 in 11 populations are presented in Table 3 (supplementary data). Fifteen haplotypes were observed exclusively in the tribal populations. Konkanastha Brahmin population showed the presence of four unique haplotypes. The haplotype "31–19" was observed only in TB populations showing their differentiation from the remaining populations.

Results for population pairwise genetic distances (F_{ST} and R_{ST}) are shown in Table 4 (supplementary data). The tribal populations of DR, TB, and AA groups had significant genetic distances from the caste populations as well as from each other. The considerable genetic heterogeneity among castes and tribes of India may be due to their different origins and geographical locations. The people belonging DR family inhabit southern parts of India, speakers of TB language are confined to northeastern parts of India while people of AA family reside mostly in east and central India [23]. The genetic diversity explored by AMOVA (Table 1) showed that irrespective of any grouping, about 98% variation was found within the studied populations, highlighting their diverse origins and histories. A higher degree of variation was observed among the populations when grouped based on linguistic affinity (1.05%) rather than geographically (0.77%). These results indicate that Indian populations show high genetic diversity at the linguistic level though geography also plays an important role [23]. Forensic parameters for the decaplex are presented in Table 5 (supplementary data). All loci in the studied decaplex were polymorphic for Indian population. Gene diversity values ranged from 0.408 to 0.855. Locus DXS101 was found to be the most polymorphic and heterozygous in all populations, followed by the loci DXS6809 and DX6789. The powers of discrimination for these loci were also high. Significantly high values were obtained for combined MEC_T, MEC_D, and PD_M showing that this decaplex can be efficiently used for kinship analysis, especially in trio cases with a female child.

Conclusions

The present study shows that the studied markers are highly polymorphic in the Indian population and can be used as an informative tool for human identification along with the autosomal markers. The data generated from the present study will aid in constructing database of X-STRs for Indian populations for forensic purposes and further studies in population genetics and evolutionary biology.

Acknowledgements The study was funded by Directorate of Forensic Science, Ministry of Home Affairs, Govt. of India, New Delhi. The authors are thankful to all the blood sample donors for their cooperation.

References

1. Szibor R, Krawczak M, Hering S, Edelmann J, Kuhlisch E, Krause D (2003) Use of X-linked markers for forensic purposes. Int J Legal Med 117:67–74
2. Szibor R (2007) X-chromosomal markers: Past, present and future. Forensic Sci Int 1:93–99
3. Gadgil M, Joshi NV, Shambhu Prasad UV, Manoharan S, Patil S (1998) Peopling of India. In: Balasubramanian D, Appaji Rao N (eds) The Indian Human Heritage. Universities Press, Hyderabad, pp 100–129
4. Ghosh T, Kalpana D, Mukerjee S, Mukherjee M, Sharma AK (2010) Genetic diversity of autosomal STRs in eleven populations of India. Forensic Sci Int. doi:10.1016/j.fsigen.2010.01.005
5. Turrina S, Atzei R, Fillipini G, De Leo D (2007) Development and forensic validation of a new multiplex PCR assay with 12 X-chromosomal short tandem repeats. Forensic Sci Int 1:201–204
6. Szibor R, Edelmann J, Hering S (2003) Cell line DNA typing in forensic genetics- the necessity for reliable standards. Forensic Sci Int 138:37–43
7. Gusmao L, Sanchez-Diz P, Alves C et al (2009) A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. Int J Legal Med 123:227–234
8. Edelmann J, Szibor R (2001) DXS101: a highly polymorphic X-linked STR. Int J Legal Med 114:301–304
9. Szibor R, Edelmann J, Hering S, Gomes I, Gusmao L (2009) Nomenclature discrepancies in the HPRTB short tandem repeat. Int J Legal Med 123:185–186
10. Raymond M, Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. J Hered 86:248–249
11. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform (Online) 1:47–50
12. Szibor R, Hering S, Edelmann J (2005) A new website compiling forensic chromosome X research is now online. Int J Legal Med 120:252–254
13. Shin SH, Yu JS, Park SW, Min GS, Chung KW (2005) Genetic analysis of 18 X-linked short tandem repeat markers in Korean population. Forensic Sci Int 147:35–41
14. Hwa HL, Chang YY, Lee JC, Yin HY, Chin YH, Tseng LH, Su YN, Ko TM (2009) Thirteen X-chromosomal short tandem repeat loci multiplex data from Taiwanese. Int J Legal Med 123:263–269
15. Liu Q, Li S (2006) Patterns of genetic polymorphisms at the 10 X-chromosome STR loci in Mongol Population. Forensic Sci Int 158:76–79

16. Zeng X, Rakha A, Li S (2009) Genetic polymorphisms of 10 X-chromosome STR loci in Chinese Daur ethnic minority group. Legal Med 11:152–154
17. Becker D, Rodig H, Augustin G, Edeelmann J, Gotz F, Hering S, Szibor R, Brabetz W (2008) Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit. Forensic Sci Int 2:69–74
18. Toni C, Presciuttini S, Spinetti I, Domenici R (2003) Population data of four X-chromosome markers in Tuscany, and their use in a deficiency paternity case. Forensic Sci Int 137:215–216
19. Aler M, Sanchez-Diz P, Gomes I, Gisbert M, Carracedo A, Amorim A, Gusmao L (2007) Genetic data of 10 X-STRs in a Spanish population sample. Forensic Sci Int 173:193–196
20. Tabbada KA, De Ungria MC, Faustino LP, Athanasiadou D, Stradmann-Bellinghausen B, Schneider PM (2005) Development of a pentaplex X-chromosomal short tandem repeat typing system and population genetic studies. Forensic Sci Int 154:173–180
21. Tariq MA, Ullah O, Riazuddin SA, Riazuddin S (2008) Allele frequency distribution of 13 X-chromosomal STR loci in Pakistani population. Int J Legal Med 122:525–528
22. Szibor R, Hering S, Kuhlisch E, Plate I, Demberger S, Krawczak M, Edelmann J (2005) Haplotyping of STR cluster DXS6801-DXS6809-DXS6789 on Xq21 provides a powerful tool for kinship testing. Int J Legal Med 119:363–369
23. Majumdar PP (2008) Genomic inferences on peopling of south Asia. Current Opinion Genet Development 18:280–284