Structure and polymorphism of novel X-chromosome short tandem repeat loci in a Chinese Han population

Y.S. Zhu[1,2], H. Wu[3] and J.H. Lai[1,2]

[1]College of Forensic Science, Xi'an Jiaotong University,
Key Laboratory of Ministry of Public Health for Forensic Science,
Xi'an, Shaanxi, China
[2]Key Laboratory of Environment and Genes Related to Diseases,
Xi'an Jiaotong University, Ministry of Education, Xi'an, Shaanxi, China
[3]Department of Neurology, Xi'an Children's Hospital, Xi'an, Shaanxi, China

Corresponding author: J.H. Lai
E-mail: zhuyongsheng3000@aliyun.com

Genet. Mol. Res. 14 (4): 15044-15049 (2015)
Received June 11, 2015
Accepted September 30, 2015
Published November 24, 2015
DOI http://dx.doi.org/10.4238/2015.November.24.12

ABSTRACT. Recently, 5 novel X-chromosome short tandem repeat (X-STR) loci with high degrees of polymorphism were examined. In this study, we investigated the genetic distribution of these loci in a Chinese Han population. The 5 X-STR loci were successfully examined by polyacrylamide gel electrophoresis in a total of 200 unrelated Shaanxi Han individuals (100 males and 100 females). Hardy-Weinberg equilibrium tests revealed no significant deviation from expected values ($P > 0.05$) for all 5 X-STR loci in the Shaanxi Han population.The loci were named DXS-p11.3, DXS-q12, DXS-q13.3, DXS-q22.1, and DXS-q25 and were found to contain 6, 8, 7, 7, and 5 alleles, respectively. In addition, 17, 21, 18, 19, and 11 genotypes, respectively, were detected in the female samples. The heterozygosities of the 5 X-STR loci were 0.75, 0.74, 0.74, 0.72, and 0.56, respectively. The polymorphic information contents of the 5 X-STR loci were 0.70, 0.69, 0.69, 0.68, and 0.51, respectively. The individual

discrimination values of the 5 X-STR loci were 0.88, 0.86, 0.88, 0.87, and 0.74, respectively. Five new X-chromosome STR loci with high degrees of polymorphism were observed in our lab. The results of this study are important for forensic individual identification, paternity identification, and population genetics research.

Key words: Han population; Short tandem repeats; X-chromosome

INTRODUCTION

Short tandem repeats (STRs) have been widely used in human disease-related gene-based cloning, anthropology, population genetics, and other areas (Botstein et al., 1980; Desmarais et al.,1998; Cuc et al., 2008). STRs of the X chromosome have enormous value compared to other genetic markers such as normal chromosomes, Y chromosomes, and mitochondria because of the unique genetic characteristics of the X chromosome. X-chromosome STRs (X-STRs) are important for individual identification and paternity testing, which play important roles in forensic science (Ayres and Powley, 2005; Silveira et al., 2007; Datta et al., 2012). A study by Edward et al. (1991) identified the HumAR locus as the first X-chromosome STR loci used. Their study paved the way for the study of X-STRloci in human genetics and forensic science. Currently, approximately 40X-STR loci have been identified (Machado and Medina-Acosta, 2009), which not sufficient for practical applications, such as genetics and forensic science. Therefore, identifying highly polymorphic X-STR loci is very important.

In our study, the X chromosome sequences of 200 samples were used for polymorphic screening and identification. As a result, 5 new highly polymorphic X-STR loci were identified by our lab. Based on this finding, we investigated the genetic distribution of these X-STR loci in the Han population in Xi'an, China.

MATERIAL AND METHODS

Subjects

The control group consisted of 200 unrelated healthy subjects (mean age ± SD: 35.4 ± 5.3 years; 100 men and 100 women) who underwent health examinations at the Medical Examination Center of the First Affiliated Hospital of Xi'an Jiaotong University (Xinxiang, China) from October 2006to September 2013. All participants completed a family history questionnaire and were self-identified as Han Chinese from the Shaanxi province for at least 3 generations. Exclusion criteria were as follows: a relative was included in the study; participation in other studies; mixed ancestry; chronic brain disease; prescription medications that could affect the central nervous system; history of seizures, hematological diseases, or severe liver or kidney impairment; and pregnancy. No familial relationships between the study participants were identified. Written informed consent was obtained from all participants. The study protocol was approved by the Ethical Committee of the Medical College, Xi'an Jiaotong University.

Allelic detection and genotyping

Peripheral blood was collected from the enrolled subjects in tubes coated with EDTA. Genomic DNA was extracted from blood leukocytes using the EZNA™ Blood DNA Midi Kit (Omega

Bio-Tek, Norcross, GA, USA) according to the manufacturer protocol. All samples were stored at -20°C until use. SSHunter software was used for sequencing the X-chromosome, and Primer 5.0 software (Premier Biosoft, Palo Alto, CA, USA) was used to design primers.Briefly, the 12-μL polymerase chain reaction (PCR) included the following reagents: 50-200 ng genomic DNA, 2X PCR buffer, 15 mM $MgCl_2$, 5 μM primers, and 0.5 U Taq polymerase (Tiangen Biotech, Beijing, China). The primer sequences and PCR amplification cycle parameters for the 5lociare listed in Table 1. For allele typing, 6% polyacrylamide gel electrophoresis and silver staining were used, and the alleles were named according to the number of repeats in the sequences based on the recommendations of the International Society of Forensic Genetics.

Table 1. PCR primer sequences and parameters of 5 X-STR loci in this study.

Accession	Name	Motif	Allele number (N)	Repeat number (N)	Product size (bp)	Annealing temperature	Primer sequence
FN557527	DXS-p11.3	TAGA	6	10-15	139-159	64.2°C, 40 s	F: 5'-GCAGGCATCATCCAATCCAC 3' R: 5'-TGTCGCCCAGGCTTGAGTAC-3'
FN557528	DXS-q12	GCA	8	17-29	329-365	69°C, 40 s	5'-AGTTAGGGCTGGGAAGGGTCT-3' 5'-CGGCTGTGAAGGTTGCTGTT-3'
FN557529	DXS-q13.3	TCTA	7	12-20	216-248	62°C, 40 s	5'-AATAACAGAAACTATGTCCCACG-3' 5'-TCTCCCTAAATGAAAGGCAAAA-3'
FN557530	DXS-q22.1	TAGA	7	9-15	159-183	62.5°C, 40 s	5'-ATTCCAGCCTTGACGACAGT-3' 5'-TCTTCCCAAAGCACTATTCA-3'
FN557531	DXS-q25	TAGA	5	9-13	189-205	58.5°C, 40 s	5'-GGGTTATGCTGAATCTTAGT-3' 5'-CCTCTTCTTAAGTTCCTACA-3'

RESULTS

For the 5 X-STR loci, DXS-p11.3, DXS-q12, DXS-q13.3, DXS-q22.1, and DXS-q25, 6, 8, 7, 7, and 5 alleles were detected in 200 Han Chinese individuals in Xi'an, China. Additionally, 17, 21, 18, 19, and 11 genotypes, respectively, were detected in the female samples. All allele frequencies and forensic parameters of the men and women are shown in Table 2, and the genotype frequency distribution is listed in Table 3.

The frequency distribution of all loci were in line with Hardy-Weinberg equilibrium based on the results of the χ^2 test ($P > 0.05$). The heterozygosities of the 5 X-STR loci were 0.75, 0.74, 0.74, 0.72, and 0.56, respectively. The polymorphic information contents of the 5 X-STR were 0.70, 0.69, 0.69, 0.68, and 0.51, respectively. The individual discrimination values of the 5 X-STR were 0.88, 0.86, 0.88, 0.87, and 0.74, respectively. The results showed that the 5 X-chromosome STR loci have high genetic diversity and discrimination power.

DISCUSSION

Because of their unique characteristics, X-STR lociare very important in gene mapping and the genetic diagnosis of X chromosome-linked genetic diseases (Lai et al., 2006). In addition, in special paternity casesin the judicial practice of forensic science, such as cases lacking parents for paternity testing, particularly in the identification of half-sister kinship when the children being investigated are girls, X-STR analysisprovides a great advantage over constant staining and Y chromosome STR analysis (Szibor et al., 2003b; Shin et al., 2005; Nadeem et al., 2009). In cases of rape-induced pregnancy, when human chorionic tissues are used to identify the biological father of female fetuses, X-STR typing showsadvantagesover other methods of DNA typing in excluding suspects (Szibor et al., 2003a).

Table 2. Allele frequency distribution of 5 X-STRs loci in the present study (N = 200; 100 males; 100 females).

Alleles	DXS-p11.3				DXS-q12				DXS-q13.3				DXS-q22.1				DXS-q25			
	Num	Freq	Num	Freq	Num	Freq	Num	Freq	Num	Freq	Num	Freq	Num	Freq	Num	Freq	Num	Freq	Num	Freq
9															2	0.01	2	0.02	2	0.01
10	6	0.06	2	0.01									2	0.02	4	0.02	7	0.07	19	0.095
11	26	0.26	17	0.085									6	0.06	25	0.125	64	0.64	121	0.605
12	17	0.17	51	0.255					2	0.02	6	0.03	45	0.45	82	0.41	21	0.21	50	0.25
13	34	0.34	64	0.32					18	0.18	36	0.18	33	0.33	57	0.285	6	0.06	8	0.04
14	15	0.15	55	0.275					41	0.41	85	0.425	8	0.08	11	0.055				
15	2	0.02	11	0.055					29	0.29	42	0.21	6	0.06	21	0.105				
16					2	0.02			2	0.02	23	0.115				1				
17					5	0.05	12	0.06	6	0.06	8	0.04								
18																				
19																				
20					26	0.26	45	0.225	2	0.02										
21																				
22					7	0.07	12	0.06												
23					44	0.44	88	0.44												
24																				
25					2	0.02	5	0.025												
26					12	0.12	31	0.155												
29					2	0.02	7	0.035												
PIC		0.702421608				0.685482491				0.68877896				0.684987245				0.50709682		
PD		0.878225441				0.859410431				0.876736111				0.866319444				0.736979167		
PE		0.261355448				0.413681355				0.226631573				0.206389054				0.1875		
H				0.75				0.74				0.74				0.72				0.56

PIC = polymorphism information content; PD = power of discrimination; PE = power of exclusion; H = heterozygosity.

Table 3. Genotyping frequency distribution of 5 X-STR loci in the present study (N = 200; 100 males; 100 females).

	DXS-p11.3		DXS-q12		DXS-q13.3		DXS-q22.1		DXS-q25	
No	Genotype	Frequency	Genotype	Frequency	Genotype	Frequency	Genotype	Frequency	Genotype	Frequency
1	10.12	0.01	17.20	0.03	12.13	0.01	9.12	0.01	9.11	0.01
2	10.13	0.01	17.22	0.01	12.14	0.03	9.13	0.01	9.12	0.01
3	10.14	0.01	17.23	0.05	12.15	0.01	10.11	0.01	10.10	0.01
4	11.11	0.01	17.26	0.02	12.16	0.01	10.12	0.02	10.11	0.11
5	11.12	0.04	17.29	0.01	13.13	0.03	10.13	0.01	10.12	0.05
6	11.13	0.05	20.20	0.05	13.14	0.15	11.11	0.02	10.13	0.01
7	11.14	0.05	20.22	0.03	13.15	0.08	11.12	0.10	11.11	0.37
8	11.15	0.01	20.23	0.20	13.16	0.04	11.13	0.07	11.12	0.30
9	12.12	0.06	20.25	0.01	13.17	0.02	11.14	0.01	11.13	0.05
10	12.13	0.16	20.26	0.07	14.14	0.18	11.15	0.02	12.12	0.06
11	12.14	0.14	20.29	0.02	14.15	0.18	12.12	0.17	12.13	0.02
12	12.15	0.03	22.23	0.05	14.16	0.10	12.13	0.23		
13	13.13	0.10	22.26	0.02	14.17	0.03	12.14	0.05		
14	13.14	0.18	22.29	0.01	15.15	0.04	12.15	0.08		
15	13.15	0.03	23.23	0.19	15.16	0.05	13.13	0.08		
16	14.14	0.08	23.25	0.02	15.17	0.02	13.14	0.03		
17	14.15	0.03	23.26	0.14	16.16	0.01	13.15	0.06		
18			23.29	0.03	16.17	0.01	14.15	0.01		
19			25.26	0.01			15.15	0.01		
20			26.26	0.02						
21			26.29	0.01						

However, information regarding the sites of X-STRs is limited. In recent years, because of the gradual increase in STR testing and the increase in the number of special forensic cases (Szibor et al., 2007), research focused on identifying additional highly polymorphic and ethnic characteristics of X-STR locihas become increasingly important.

STR loci are widely distributed in the human genome. The gene coding region of the STR loci may not be suitable for forensic applications (Sziboret al., 2003a); however, the non-coding regionsof STR loci have a high degree of stability, and STR polymorphisms are not subject to selection pressure (Mitra et al., 1976). Therefore, we examined the non-coding regions of STR loci, which were far away from genes. Of the 5 X-STR loci identified by our lab, the shortest distance from an X-STR locus to a gene was 28.1 kb for the DXS-q22.1 locus and the BEXL1 gene, while the greatest distance was929 kb for the DXS-q12 locus and the EDA2R gene.

In forensic applications, X-STR loci are only valuable when the linkage of the X-STR loci is clear. The possibility of linkage should be considered when using genetic markers on the same chromosome (Tillmar et al., 2008). According to the physical map distance between loci in the human gene mapping database, 1 cm represents 1 million base pairs. Additionally, linkage disequilibrium among the different loci can be analyzed by population census to determine whether locus linkage exists.

In our study, among the 5 X-STR genetic markers identified by our lab, the length between DXS-q12 and DXS-q13.3 was 7.12 cM, which was the shortest distance between any 2 of the 5 markers, while the length between DXS-q13.3 and DXS-q22.1 was 28.52cM, which was the greatest distance between any 2 of the 5 markers. Upon examination, no linkage among the 5 loci was identified; thus, these loci can be used forforensic applications. Furthermore, the 5 new X-STR loci, DXS-p11.3, DXS-q12, DXS-q13.3, DXS-q22.1, and DXS-q25, showed high degrees of polymorphism and individual recognition. These new X-STR loci have value for applicationsin forensic individual identification, paternity identification, and population genetics research.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research partially supported by Genetic Resource of Chinese Population (#57004). The funding sources had no input intothe design of this study, the collection, analysis and interpretationof data, the writing of the report or the decision to submit the paper for publication.

REFERENCES

Ayres KL and Powley WM (2005). Calculating the exclusion probability and paternity index for X-chromosomal loci in the presence of substructure. *Forensic Sci. Int.* 149: 201-203.

Botstein D, White RI, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.

Cuc LM, Mace ES, Crouch JH, Quang VD, et al. (2008). Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). *BMC Plant Biol*. 8: 55.

Desmarais D, Zhong Y, Chakraborty R, Perreault C, et al. (1998). Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene. *J. Forensic Sci*. 43: 1046-1049.

Edwards A, Civitello A, Hammond HA and Caskey CT (1991). DNA typing and genetic-mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* 49: 746-756.

Lai KK, Lo IF, Tong TM, Cheng LY, et al. (2006). Detecting exon deletions and duplications of the DMD gene using Multiplex Ligation-dependent Probe Amplification (MLPA). *Clin. Biochem.* 39: 367-372.

Machado FB and Medina-Acosta E (2009). Genetic map of human X-linked microsatellites used in forensic practice. *Forensic Sci. Int. Genet.* 3: 202-204.

Mitra S (1976). More on Nei and Roychoudhury's sampling variances of heterozygosity and genetic distance. *Genetics* 82: 543-545.

Nadeem A, Babar ME, Hussain M and Tahir MA (2009). Development of pentaplex PCR and genetic analysis of X chromosomal STRs in Punjabi population of Pakistan. *Mol. Biol. Rep*. 36: 1671-1675.

Shin SH, Yu JS, Park SW, Min GS, et al. (2005). Genetic analysis of 18 X-linked short tandem repeat markers in Korean population. *Forensic Sci. Int*. 147: 35-41.

Silveira D, Silva FF, Jesus PR and Whittle MR (2007). Use of X-linked short tandem repeat loci in routine parentage casework. *Transfusion* 47: 1050-1053.

Szibor R (2007). X-chromosomal markers: Past, present and future. *Forensic Sci. Int. Genet*. 1: 93-99.

Szibor R, Edelmann J, Hering S, Plate I, et al. (2003a). Cell line DNA typing in forensic genetics - the necessity of reliable standards. *Forensic Sci. Int*. 138: 37-43.

Szibor R, Krawczak M, Hering S, Edelmann J, et al. (2003b). Use of X-linked markers for forensic purposes. *Int. J. Legal Med*. 117: 67-74.

Tillmar AO, Mostad P, Egeland T, Lindbolm B, et al. (2008). Analysis of linkage and linkage disequilibrium for eight X-STR markers. *Forensic Sci. Int. Genet*. 3: 37-41.

Datta P and Datta SS (2012). Role of deoxyribonucleic acid technology in forensic dentistry. *J. Forensic Dent Sci*. 4: 42-46.