Genetic Polymorphisms of Nine X-STR Loci in Four Population Groups from Inner Mongolia, China

Qiao-Fang Hou, Bin Yu, and Sheng-Bin Li*

Department of Forensic Sciences, Medical College of Xi'an Jiaotong University/Key Laboratory of Forensic Sciences, Ministry of Health/Key Laboratory of Environment and Genes Related to Diseases, Ministry of Education, Xi'an Jiaotong University, Xi'an 710061, China.

Nine short tandem repeat (STR) markers on the X chromosome (DXS101, DXS6789, DXS6799, DXS6804, DXS7132, DXS7133, DXS7423, DXS8378, and HPRTB) were analyzed in four population groups (Mongol, Ewenki, Oroqen, and Daur) from Inner Mongolia, China, in order to learn about the genetic diversity, forensic suitability, and possible genetic affinities of the populations. Frequency estimates, Hardy-Weinberg equilibrium, and other parameters of forensic interest were computed. The results revealed that the nine markers have a moderate degree of variability in the population groups. Most heterozygosity values for the nine loci range from 0.480 to 0.891, and there are evident differences of genetic variability among the populations. A UPGMA tree constructed on the basis of the generated data shows very low genetic distance betweent Mongol and Han (Xi'an) populations. Our results based on genetic distance analysis are consistent with the results of earlier studies based on linguistics and the immigration history and origin of these populations. The minisatellite loci on the X chromosome studied here are not only useful in showing significant genetic variation between the populations, but also are suitable for human identity testing among Inner Mongolian populations.

Key words: X chromosome STR, genetic polymorphism, genetic distance, Inner Mongolian populations

Introduction

The Inner Mongolia Autonomous Region of China is a region with diverse genetic and cultural fusion, where there settled 49 ethnic groups with lots of differences in language, culture, immigration history, and traditional occupation (http://www.nmg.gov.cn/zjnmg/rwli.htm). All these implicate that there may be different genetic structures among the populations. Therefore, it is a native genetic pool to study the genetic relationship among the Inner Mongolian populations, which could have significant contributions to population genetic diversity, disease linkage analysis, and forensic casework.

The X chromosome short tandem repeats (X-STRs) have been recently recognized as useful tools in forensic kinship testing (*1–3*) and disease linkage analysis (*4–6*). The highly effective strategy of X chromosome microsatellite haplotyping requires the description of numerous STR markers. A number of genetic studies based on traditional serological and other protein markers have been carried out on Inner Mongolian populations, and some data on STR genetic markers have also become available, which were mainly reported as frequency data from individual populations, with information about Hardy-Weinberg equilibrium and other parameters of forensic interest (*7–10*). However, few of these studies were aimed to comprehensively analyze X-STRs and the level of anthropological substructure in these populations. Here we present the data for X-STR markers after extensively investigated the genetic background of different Inner Mongolian populations, in order to add new X-STRs to the panel of X chromosome markers and to the genetic pool of Inner Mongolia area, which will contribute to general forensics, anthropological genetic study, and disease genetic study.

***Corresponding author.**
E-mail: shengbinlee@mail.xjtu.edu.cn

Results

Nine STR markers on the X chromosome (DXS101, DXS6789, DXS6799, DXS6804, DXS7132, DXS7133, DXS7423, DXS8378, and HPRTB) were analyzed in four population groups (Mongol, Ewenki, Oroqen, and Daur) from Inner Mongolia, China. We examined the observed value and expected value of the genotype for these X-STR loci using the statistical method χ^2 test. The results indicate that the distribution of the genotype at the loci coordinates with the Hardy-Weinberg law (P>0.05). Allele frequency distributions and polymorphism indexes were estimated for these X-STR loci in the four populations as provided below. The allele frequencies of both male and female were calculated respectively. The loci that have differences between male and female were excluded when analyzing the allele frequencies of the whole population.

Allele frequency and polymorphism valuation for Mongol population

From 100 unrelated individuals of Mongol population, we detected 61 alleles with frequencies between 0.0069–0.5724 (Table 1). The polymorphism indexes of these loci are shown in Table 2, which reveals that DXS7133 and DXS7423 have lower polymorphism in Mongol population and are not suitable for forensic identity.

Table 1 Allele frequencies of nine X-STR loci in Mongol population (n=100)*

Allele	DXS101	DXS6789	DXS6799	DXS6804	DXS7132	DXS7133	DXS8378	HPRTB
7	–	–	–	–	–	–	–	–
8	–	–	0.0207	–	–	–	–	–
9	–	–	0.0345	0.0069	–	0.6828	0.0276	0.0070
10	–	–	0.1517	0.0138	–	0.2069	0.5724	0.0775
11	–	–	0.5862	0.1793	–	0.0897	0.2483	0.1197
12	–	–	0.1793	0.1931	0.0544	0.0207	0.1310	0.2606
13	–	–	0.0276	0.3793	0.2313	–	0.0207	0.2746
14	0.0069	–	–	0.1310	0.3810	–	–	0.2113
15	–	–	–	0.0897	0.2449	–	–	0.0423
16	–	0.1049	–	0.0069	0.0816	–	–	0.0070
17	–	0.2098	–	–	0.0068	–	–	–
18	0.0069	0.0210	–	–	–	–	–	–
19	–	–	–	–	–	–	–	–
20	–	0.0769	–	–	–	–	–	–
21	0.0625	0.2308	–	–	–	–	–	–
22	0.1250	0.1818	–	–	–	–	–	–
23	0.2847	0.1608	–	–	–	–	–	–
24	0.1389	0.0140	–	–	–	–	–	–
25	0.2083	–	–	–	–	–	–	–
26	0.1042	–	–	–	–	–	–	–
27	0.0347	–	–	–	–	–	–	–
28	0.0139	–	–	–	–	–	–	–
29	0.0139	–	–	–	–	–	–	–

*The locus DXS7423 that has differences between male and female was excluded.

Table 2 Polymorphism indexes of nine X-STR loci in Mongol population (n=100)

Parameter*	DXS101	DXS6789	DXS6799	DXS6804	DXS7132	**DXS7133**	**DXS7423**	DXS8378	HPRTB
H	0.674	0.844	0.543	0.891	0.468	**0.348**	**0.370**	0.565	0.609
PIC	0.794	0.815	0.604	0.733	0.690	**0.425**	**0.389**	0.559	0.745
P_e	0.389	0.815	0.228	0.733	0.690	**0.085**	**0.096**	0.251	0.745
PD (f)	0.813	0.787	0.511	0.749	0.726	**0.489**	**0.499**	0.548	0.794
PD (m)	0.938	0.929	0.836	0.866	0.866	**0.662**	**0.653**	0.771	0.909

*H, heterozygosity; PIC, polymorphism information content; P_e, probability of paternity exclusion; PD (f)/PD (m), average power of discrimination in female/male.

Allele frequency and polymorphism valuation for Ewenki population

From 99 unrelated Ewenki individuals, we detected 51 alleles with frequencies between 0.0052–0.6407 (Table 3). The nine loci in Ewenki population all showed moderate degree of polymorphism and forensic application value (Table 4).

Allele frequency and polymorphism valuation for Oroqen population

From 108 unrelated individuals of Oroqen, we detected 55 alleles with frequencies between 0.0096–0.6258 (Table 5). Except DXS7423, the other loci in Oroqen population showed moderate degree of polymorphism and forensic application value (Table 6).

Allele frequency and polymorphism valuation for Daur population

From 87 unrelated individuals of Daur, we detected 51 alleles with frequencies between 0.0161–0.6935 (Table 7). Except DXS7133, the other loci in Daur population showed moderate degree of polymorphism and forensic application value (Table 8).

Genetic distance and hierarchical cluster analysis

With the allele frequencies of the nine X-STR loci and other data from our laboratory, we calculated the genetic distance between the four populations and the Han population in Xi'an, China (Table 9). Based on the genetic distance, a UPGMA (Unweighted Pair-

Table 3 Allele frequencies of nine X-STR loci in Ewenki population (n=99)

Allele	DXS101	DXS6789	DXS6799	DXS6804	DXS7132	DXS7133	DXS7423	DXS8378	HPRTB
7	–	–	–	0.0102	–	0.0101	–	–	–
8	–	–	–	–	–	–	–	0.0051	–
9	–	–	0.0510	–	–	0.5404	–	0.0303	–
10	–	–	0.2602	–	–	0.3030	–	0.4242	–
11	–	–	0.4439	0.3316	–	0.1465	–	0.3232	0.0365
12	–	–	0.1582	0.1276	0.1122	–	–	0.1818	0.2448
13	–	–	0.0867	0.3112	0.1531	–	0.0208	0.0354	0.4063
14	–	–	–	0.1684	0.4082	–	0.2188	–	0.2083
15	–	0.0585	–	0.0510	0.2551	–	0.6407	–	0.1042
16	–	0.2021	–	–	0.0612	–	0.0990	–	–
17	–	0.1277	–	–	0.0102	–	0.0208	–	–
18	0.0052	–	–	–	–	–	–	–	–
19	–	–	–	–	–	–	–	–	–
20	–	–	–	–	–	–	–	–	–
21	0.0208	0.0426	–	–	–	–	–	–	–
22	0.1094	0.3085	–	–	–	–	–	–	–
23	0.2396	0.1968	–	–	–	–	–	–	–
24	0.5417	0.0638	–	–	–	–	–	–	–
25	–	–	–	–	–	–	–	–	–
26	0.0729	–	–	–	–	–	–	–	–
27	0.0104	–	–	–	–	–	–	–	–

Table 4 Polymorphism indexes of nine X-STR loci in Ewenki population (n=99)

Parameter	DXS101	DXS6789	DXS6799	DXS6804	DXS7132	DXS7133	DXS7423	DXS8378	HPRTB
H	0.533	0.465	0.756	0.733	0.578	0.500	0.556	0.652	0.455
PIC	0.609	0.757	0.634	0.687	0.638	0.517	0.523	0.612	0.657
P_e	0.218	0.159	0.519	0.482	0.265	0.188	0.241	0.358	0.151
PD (f)	0.832	0.900	0.842	0.860	0.850	0.755	0.744	0.819	0.860
PD (m)	0.607	0.787	0.710	0.754	0.750	0.600	0.484	0.684	0.720

Table 5 Allele frequencies of nine X-STR loci in Oroqen population (n=108)*

Allele	DXS6789	DXS6799	DXS6804	DXS7132	DXS7133	DXS7423	DXS8378	HPRTB
6	–	–	–	–	0.0235	–	–	–
7	–	–	–	–	0.0094	–	–	–
8	–	0.0187	0.0096	–	0.0141	–	0.0246	–
9	–	0.0187	–	–	0.6258	–	0.2826	–
10	–	0.2520	–	–	0.1964	–	0.3655	–
11	–	0.5609	0.3541	0.0098	0.1308	–	0.2085	0.0350
12	–	0.1124	0.1591	0.0534	–	–	0.1039	0.2250
13	–	0.0374	0.2718	0.2085	–	0.0148	0.0148	0.3750
14	–	–	0.1863	0.4107	–	0.2176	–	0.2900
15	–	–	0.0192	0.2604	–	0.7134	–	0.0650
16	0.1583	–	–	0.0524	–	0.0543	–	0.0100
17	0.2326	–	–	0.0049	–	–	–	–
18	0.0393	–	–	–	–	–	–	–
19	0.0791	–	–	–	–	–	–	–
20	0.2134	–	–	–	–	–	–	–
21	0.1531	–	–	–	–	–	–	–
22	0.0896	–	–	–	–	–	–	–
23	0.0346	–	–	–	–	–	–	–

*The locus DXS101 that has differences between male and female was excluded.

Table 6 Polymorphism indexes of nine X-STR loci in Oroqen population (n=108)

Parameter	DXS101	DXS6789	DXS6799	DXS6804	DXS7132	DXS7133	**DXS7423**	DXS8378	HPRTB
H	0.820	0.878	0.560	0.796	0.800	0.680	**0.360**	0.500	0.580
PIC	0.730	0.790	0.560	0.710	0.660	0.580	**0.330**	0.670	0.620
P_e	0.637	0.750	0.246	0.591	0.599	0.398	**0.091**	0.188	0.268
PD (f)	0.890	0.915	0.790	0.880	0.850	0.791	**0.534**	0.862	0.830
PD (m)	0.776	0.835	0.598	0.725	0.701	0.481	**0.492**	0.709	0.744

Group Method using Arithmetic averages) tree and a hierarchical cluster were constructed to depict the genetic affinities (Figures 1 and 2). The tree shows three distinct branches, demonstrating the low genetic distance between Mongol and Han (Xi'an) as well as between Oroqen and Daur.

Discussion

The different distributions of allele frequency and genotype frequency of the four populations indicate that each population has its own characteristics of genetic structure. As to the total 218 alleles of the nine X-STRs with allele frequencies between 0.0052–0.6935, the low-frequency alleles, however, are different from group to group. One possible explanation of such allelic distribution could be the evolutionary antiquity of the alleles, which suggests that the most common alleles are the oldest, while the wider distribution of the low-frequency alleles is a reflection of rates and types of mutations (*11*).

According to the polymorphism index results, we sorted the loci DXS6789, DXS7132, DXS6804, and HPRTB as "highest diversity genetic markers", and the loci DXS101, DXS8378, and DXS6799 as "higher diversity genetic markers". The high level of genetic heterogeneity observed in these STR markers across the populations indicates their utility in human identification for forensic purposes as well as in population genetics (*12*).

The genetic distance, hierarchical cluster analysis, and pattern of the UPGMA tree reflect the ethnic background of the Inner Mongolian populations. Considering the linguistics affinity, immigration history and origin of these populations, the close genetic

Table 7 Allele frequencies of nine X-STR loci in Daur population (n=87)*

Allele	DXS101	DXS6789	DXS6799	DXS6804	DXS7132	DXS7133	DXS7423	DXS8378
8	–	–	0.0173	–	–	0.0115	–	0.0231
9	–	–	0.0347	–	–	0.6840	–	0.5973
10	–	–	0.1778	0.0115	–	0.1954	–	0.2417
11	–	–	0.5805	0.3046	–	0.1092	–	0.1034
12	–	–	0.1725	0.1958	0.0581	–	–	0.0289
13	–	–	0.0173	0.2583	0.1453	–	0.0058	0.0058
14	–	–	–	0.2012	0.4012	–	0.2928	–
15	–	–	–	0.0286	0.3140	–	0.6154	–
16	–	0.0232	–	–	0.0813	–	0.0860	–
17	–	0.0697	–	–	–	–	–	–
18	–	0.0174	–	–	–	–	–	–
19	–	0.1222	–	–	–	–	–	–
20	–	0.3231	–	–	–	–	–	–
21	0.0173	0.2923	–	–	–	–	–	–
21.5	0.0286	–	–	–	–	–	–	–
22	0.2035	0.1276	–	–	–	–	–	–
23	0.3055	0.0243	–	–	–	–	–	–
24	0.2262	–	–	–	–	–	–	–
25	0.2188	–	–	–	–	–	–	–

*The locus HPRTB that has differences between male and female was excluded.

Table 8 Polymorphism indexes of nine X-STR loci in Daur population (n=87)

Parameter	DXS101	DXS6789	DXS6799	DXS6804	DXS7132	**DXS7133**	DXS7423	DXS8378	HPRTB
H	0.677	0.613	0.742	0.645	0.613	**0.323**	0.516	0.548	0.613
PIC	0.670	0.710	0.780	0.550	0.670	**0.420**	0.400	0.580	0.680
P_e	0.394	0.307	0.496	0.349	0.307	**0.073**	0.202	0.233	0.307
PD (f)	0.870	0.895	0.920	0.783	0.868	**0.635**	0.639	0.803	0.874
PD (m)	0.789	0.760	0.820	0.601	0.700	**0.489**	0.559	0.529	0.575

Table 9 Genetic distance between the four populations and the Han (Xi'an) population

Population	Mongol	Ewenki	Oroqen	Daur
Mongol	–	–	–	–
Ewenki	0.0797	–	–	–
Oroqen	0.0816	0.0880	–	–
Daur	0.1061	0.1376	0.0352	–
Han (Xi'an)	0.0220	0.0438	0.0781	0.1183

affinity between Mongol and Han (Xi'an) as well as between Oroqen and Daur populations suggest that the two groups in each pair could have the same origin with recent separation from common stock, or they are two different groups with extensive gene flow between them (*13*, *14*). The Ewenki population, however, constitutes a separate branch with the other population groups. On the other hand, the close genetic distance between Mongol and Han (Xi'an), despite occupying far-off geographical areas, suggests that ethnic affiliation plays a greater role in genetic distance and gene flow among population stocks rather than geographical location (*13*, *15*). However, the genetic analysis of Oroqen and Daur populations

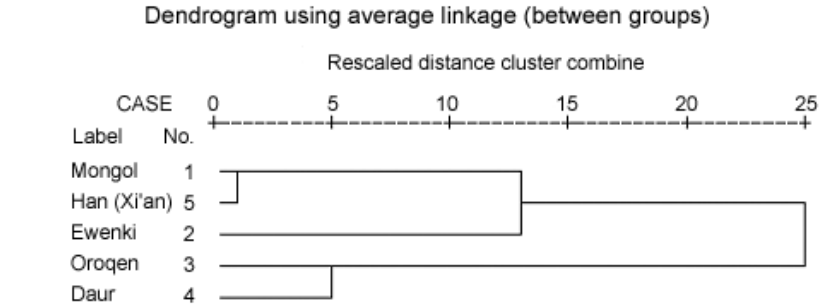

Fig. 1 Hierarchical cluster analysis based on the genetic distance between the four populations and the Han (Xi'an) population.

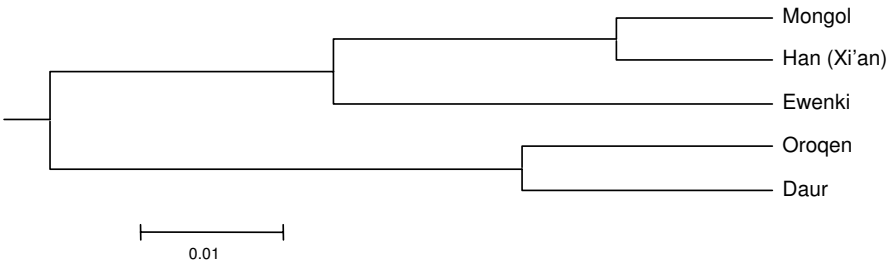

Fig. 2 A UPGMA tree depicting genetic affinities between the four populations and the Han (Xi'an) population.

shows roughly concordance with social cultural and geographical factors, which may be the results of gene flow and immigration operating between these two population stocks. Our results suggest that the microsatellite loci studied here are not only useful in showing significant genetic variation between the populations, but also are suitable for human identity testing among Inner Mongolian populations.

Materials and Methods

Sample collection, DNA isolation, and allele typing

Whole blood samples in EDTA-coated Vacutainer tubes (Becton Dickinson, Franklin Lakes, USA) were obtained with informed consent from 394 unrelated donors, including 100 Mongol, 99 Ewenki, 108 Oroqen, and 87 Daur individuals. Community, health status, and family disease history were recorded on blood donor cards of the DNA typing unit. Genomic DNA was extracted by the Chelex-100 method (*16*). The amplification was performed using a 13 μL final reaction volume containing 1 ng of sample DNA and 0.5 unit of Taq DNA polymerase with STR buffer and specific primers. Denaturing polyacrylamide gel electrophoresis and silver staining were used to detect the allele fragment. The allele ladder used was developed by cloning technique.

Statistical analysis

SPSS12.0 software (www.spss.com) was used to calculate the allele frequency, genotype frequency, genetic distance, forensic index, and so on. Hardy-Weinberg equilibrium was performed by Genepop software (*17*). MEGA3.1 software (*18*) was used to construct the phyletic cladogram tree.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 39970401) and the Ministry of Education of China (Grant No. 505015).

Authors' contributions

QFH and BY carried out laboratory experiments and drafted the manuscript. SBL conceived of the study, participated in its design and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors have declared that no competing interests exist.

References

1. Barbaro, A., *et al.* 2006. X-STR typing for an identification casework. *Int. Congr. Ser.* 1288: 513-515.
2. Szibor, R., *et al.* 2003. Use of X-linked markers for forensic purposes. *Int. J. Legal Med.* 117: 67-74.
3. Tie, J., *et al.* 2006. Genetic polymorphisms of 15 STR loci in a Japanese population. *J. Forensic Sci.* 51: 188-189.
4. Riley, D.E. and Krieger, J.N. 2002. X Chromosomal short tandem repeat polymorphisms near the phosphoglycerate kinase gene in men with chronic prostatitis. *Biochim. Biophys. Acta* 1586: 99-107.
5. Crawford, D.C., *et al.* 2000. Survey of the fragile X syndrome CGG repeat and the short-tandem-repeat and single-nucleotide-polymorphism haplotypes in an African American population. *Am. J. Hum. Genet.* 66: 480-493.
6. Vauhkonen, H., *et al.* 2004. Correlation between the allelic distribution of STRs in a Finnish population and phenotypically different gastrointestinal tumours: a study using four X-chromosomal markers (DXS7423, DXS8377, ARA, DXS101). *Ann. Hum. Genet.* 68: 555-562.
7. Liu, S.P., *et al.* 2004. Genetic polymorphisms of short tandem repeat loci D3S1358, D13S317, D5S818 from Nongqu Mongolia. *Chin. J. Med. Genet.* 21: 297-298.
8. Li, S.B., *et al.* 2000. STR polymorphisms in five Chinese ethnic groups (2). *Acta Genet. Sin.* 27: 1035-1041.
9. Liu, S., *et al.* 2005. Study of genetic polymorphism of 6 short tandem repeat loci in Nongqu Mongolian of Inner Mongolia Autonomous Region in China. *Chin. J. Med. Genet.* 22: 222-223.
10. Chen, L., *et al.* 2006. Genetic analysis of 15 STR loci of Chinese Ewenki ethnic population. *J. Forensic Sci.* 51: 1408-1409.
11. Watterson, G.A. and Guess, H.A. 1977. Is the most frequent allele the oldest? *Theor. Popul. Biol.* 11: 141-160.
12. Li, S.B. 2000. DNA fragment length polymorphism. In *Human DNA genetic marker* (ed. Li, S.B.), pp. 87-94. People's Medical Publishing House, Beijing, China.
13. Dutta, R. and Kashyap, V.K. 2001. Genetic variation at minisatellite loci D1S7, D4S139, D5S110 and D17S79 among three population groups of eastern India. *J. Genet.* 80: 23-30.
14. Xiao, F.X., *et al.* 2004. Human X-chromosomal lineages in Europe reveal Middle Eastern and Asiatic contacts. *Eur. J. Hum. Genet.* 12: 301-311.
15. Iriondo, M., *et al.* 2003. DNA polymorphisms detect ancient barriers to gene flow in Basques. *Am. J. Phys. Anthropol.* 122: 73-84.
16. Walsh, P.S., *et al.* 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10: 506-513.
17. Raymond, M. and Rousset, F. 1995. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *J. Heredity* 86: 248-249.
18. Kumar, S., *et al.* 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5: 150-163.