


SCIENTIFIC REPORTS



OPEN

Investigation of 12 X-STR loci in Mongolian and Eastern Han populations of China with comparison to other populations

Ruiyang Tao^{1,2}, Jiashuo Zhang³, Yingnan Bian², Rixia Dong⁴, Xiling Liu², Chao Jin⁵, Ruxin Zhu², Suhua Zhang² & Chengtao Li^{1,2} 

Due to the unique inheritance pattern, X-chromosomal short tandem repeats (X-STRs) have several advantages in complex kinship cases, such as deficiency cases or grandparent-grandchild and half-sisters testing. In our study, 541 unrelated individuals gathered from Mongolian and Eastern Chinese Han populations were successfully genotyped using the Investigator Argus X-12 kit. We calculated allele/haplotype frequencies and other forensic parameters of the two populations and further explored their genetic distance with already published Chinese populations and six global populations. Our results showed that the 12 X-STR markers were highly informative in the two populations when compared with nine other Chinese populations: significant differences were found at several loci. Geographically neighboring populations or different ethnic groups within the same area appeared to have closer evolutionary relationships. We also analyzed population genetic structure by performing clustering with the STRUCTURE program and Principal Coordinate Analysis (PCoA), and we found that the Chinese and other populations enrolled in this study could be distinguished. Furthermore, Mongolian males were distinguishable from the other studied males by a moderate genetic distance. Our study also expanded the X-STR database, which could facilitate the appropriate application of the 12 X-STR markers in the forensic field in China.

Short tandem repeats (STRs) are relatively mature genetic markers that were widely used in forensic applications. Genotyping of X-chromosomal STR (X-STR) markers has developed into a valuable method used by many forensic researchers as a supplement to the information provided by autosomal STR (AS-STR), Y-chromosomal STR (Y-STR), and mitochondrial DNA, particularly in deficiency and complex kinship cases^{1,2}. These different approaches can complement each other with their characteristic genetic information. For instance, the ancestral information of populations deduced from genetic analyses could be quite different depending on the type of genetic marker used³. Meanwhile, in the identification of some certain kind of mixed stains, X-STRs can be used to exclude male individuals from the same paternal line. Moreover, because of the unique genetic inheritance of the X chromosome, the use of X-STRs is of great benefit in the identification of victims of war or mass disaster⁴.

There are several X-STR multiplexes being applied to forensic cases and genetics research. Among them, the Investigator Argus X-12 kit (Qiagen, Hilden, Germany), which includes 12 X-chromosomal STR markers as well as the amelogenin locus, has been used worldwide in a number of population sets⁵⁻⁹. To our knowledge, five Chinese Han subpopulations¹⁰⁻¹⁴ and eight ethnic minority groups from China¹⁵⁻¹⁹ have been previously studied using this kit. There are 56 ethnic groups in China, which has the largest population and third largest territory in the world. Each ethnic group has different geographical distributions, religions, customs, etc. Throughout history, some of these groups have been relatively isolated from each other, which may have resulted in relatively stable

¹Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, 610041, P.R. China. ²Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Academy of Forensic Sciences, Ministry of Justice, P.R. China, Shanghai, 200063, P.R. China. ³Department of Forensic Science, Medical School of Soochow University, Suzhou, 215123, P.R. China. ⁴The Affiliated Guangji Hospital of Soochow University, Suzhou, 215008, P.R. China. ⁵Shanghai OE Biotechnology Co, Ltd, Shanghai, 201114, P.R. China. Correspondence and requests for materials should be addressed to S.Z. (email: zhangsh@ssfj.cn) or C.L. (email: lichengtaohla@163.com)

genetic differences between populations in defined areas. Therefore, for a better forensic application of the 12 X-STRs, it is important to understand the similarities or differences in genetic distance between different Chinese populations.

The purpose of this study was to investigate the frequency distribution data and forensic parameters of 12 X-STR loci in Mongolian and Eastern Chinese Han unrelated individuals as well as to illustrate the X-chromosomal evolutionary relationships between different Chinese populations and in comparison to other global populations. Additionally, this study enriched the X-STR database in order to highlight its significance in forensic identification and kinship analysis in different ethnic populations of China.

Materials and Methods

Population samples and DNA extraction. More than 70% of China's Mongolian population lives in Inner Mongolia, which is located on the Eurasian continent²⁰. Eastern China also has a large Mongolian population and accounts for 29.3% of the total population of China. In our study, samples from 541 healthy unrelated individuals (116 males and 116 females for the Mongolian samples, 200 males and 109 females for the Eastern Chinese Han samples) were collected. Blood samples were obtained from volunteer donors had information available under informed consent, following the protocols approved by the ethics committee at the Academy of Forensic Science, Ministry of Justice, P.R. China. DNA isolation was carried out by a Chelex-100 extraction protocol²¹. All the methods were carried out in accordance with the approved guidelines of the Academy of Forensic Sciences, Ministry of Justice, P.R. China.

PCR amplification and capillary electrophoresis. All DNA samples and positive control samples (9947 A and XX28) were amplified with the Investigator Argus X-12 kit (Qiagen, Hilden, Germany) on a GeneAmp PCR System 9700 Thermal Cycler (Thermo Fisher Scientific, MA, USA), according to the manufacturer's protocol. PCR products were detected by capillary electrophoresis on an ABI PRISM 3130XL Genetic Analyzer (Thermo Fisher Scientific, MA, USA) with the protocol guide for genotyping and the allele designation was analyzed using the Genemapper ID v3.2.1 software (Thermo Fisher Scientific, MA, USA).

Statistical analysis and population comparisons. Allelic frequencies of the 12 X-STRs were calculated with CERVUS software/PowerStats v1.2.xls²² for female samples and were counted for male samples. Arlequin v3.5.2²³ was used to calculate haplotype frequencies of the four linkage groups in male samples; Hardy-Weinberg equilibrium (HWE) in female samples; the exact test of population differentiation based on allelic frequencies of males and females; linkage disequilibrium (LD) in males and females; and pairwise F_{st} genetic distances per locus among Mongolian, Eastern Chinese Han and nine other published populations from China^{10-14,16-19}. Forensic parameters for 12 X-STRs including polymorphism information content (PIC), homozygosity (HOM), heterozygosity (HET), power of exclusion (PE), paternity index (PI), paternity exclusion chance in duos and trios (MEC_D and MEC_T)²⁴, as well as power of discrimination for males (PD_M) and females (PD_F) were calculated with pooled allelic frequencies of the two populations using the online tool of the Forensic ChrX Research database (<http://www.chrx-str.org>).

China map was generated by and Package (ggplot2) and (maptools) of R: A Language and Environment for Statistical (<https://www.R-project.org>). The evolutionary history of eleven Chinese populations was inferred using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) method²⁵, and the optimal tree with the sum of branch length = 0.00821881 was shown (next to the branches). The evolutionary distances were used for phylogenetic analyses conducted in MEGA7 software²⁶. Detailed population genetic structure analysis was performed with STRUCTURE v2.3.4^{27,28} under the condition of the admixture model and correlated allelic frequencies between populations. We analyzed the structure of Mongolian/Eastern Chinese Han populations and nine other populations from previous studies based on the same 12 X-STRs. The analysis results were uploaded to the Structure Harvester²⁹ to detect the true K value of the data. Further comparisons among different male populations based on the Nei's unbiased genetic distance^{30,31} were calculated and Principal Coordinate Analysis (PCoA, known previously as PCA)³² was performed using GENALEX 6.5 software.

Data Availability. All data generated during this study are included in this published article (and its Supplementary Information files) and from the corresponding author on reasonable request.

Results and Discussion

HWE and LD test within populations. The results of HWE and LD tests in Eastern Chinese Han were presented in our previous report³³. In the Mongolian samples, a total of 164 alleles were typed using the Argus X-12 kit and no statistically significant deviation from HWE was observed except for the DXS10134 locus in female samples ($p = 0.0027$) even after applying the Bonferroni's correction for multiple testing ($p = 0.05/12$). This deviation from HWE may be caused by the substructure of the chosen population, which could result in a bias of the samples. In addition, no statistically significant LD was found in Mongolian females after Bonferroni's correction ($p = 0.05/66$), which was different from the Eastern Chinese Han analysis. This might be due to the rigorous statistical methods. On the other hand, statistically significant LD was found between DXS10101 and DXS10103 loci in Mongolian males ($p < 0.0001$), which has also been reported in other populations^{5,34,35}.

Haplotype analysis in males. The 12 X-STR markers are clustered into 4 linkage groups based on their physical localizations: LG1 (Xp22) DXS8378-DXS10135-DXS10148, LG2 (Xq11) DXS7132-DXS10074-DXS10079, LG3 (Xq26) DXS10101-DXS10103-HPRTB, LG4 (Xq28) DXS7423-DXS10134-DXS10146, and thus each cluster of 3 markers is considered as one haplotype for the genotyping of males. The numbers of observed haplotypes in Mongolian males were 99, 71, 76 and 81 for LG1, LG2, LG3 and LG4, respectively (Supplementary Table S1); while in Eastern Chinese Han males the numbers were 159, 94, 101 and 114 for LG1, LG2, LG3 and

Locus	PIC	HOM	HET	PE	PI	PD _F	PD _M	MECK _{rüger}	MECK _{ishida}	MEC _{Desmarais}	MEC _{Desmarais Duo}
DXS8378	0.5558	0.3896	0.6104	0.3036	0.1948	0.7936	0.6104	0.3608	0.5558	0.5558	0.4079
DXS10135	0.9070	0.0866	0.9134	0.8228	0.0433	0.9862	0.9134	0.8256	0.9070	0.9070	0.8363
DXS10148	0.9043	0.0889	0.9111	0.8181	0.0445	0.9853	0.9111	0.8201	0.9041	0.9043	0.8317
DXS7132	0.6975	0.2590	0.7410	0.4945	0.1295	0.8895	0.7410	0.5081	0.6975	0.6975	0.5583
DXS10074	0.7444	0.2213	0.7787	0.5600	0.1107	0.9167	0.7787	0.5693	0.7442	0.7444	0.6130
DXS10079	0.8032	0.1738	0.8262	0.6486	0.0869	0.9467	0.8262	0.6538	0.8032	0.8032	0.6867
DXS10101	0.8889	0.1026	0.8974	0.7901	0.0513	0.9810	0.8974	0.7944	0.8889	0.8889	0.8083
DXS10103	0.7496	0.2180	0.7820	0.5660	0.1090	0.9201	0.7820	0.5779	0.7496	0.7496	0.6194
HPRTB	0.6594	0.2926	0.7074	0.4398	0.1463	0.8663	0.7074	0.4643	0.6594	0.6594	0.5158
DXS7423	0.4597	0.4732	0.5268	0.2121	0.2366	0.7090	0.5268	0.2728	0.4596	0.4597	0.3181
DXS10134	0.8323	0.1505	0.8495	0.6939	0.0753	0.9602	0.8495	0.7017	0.8323	0.8323	0.7268
DXS10146	0.8670	0.1209	0.8791	0.7530	0.0605	0.9733	0.8791	0.7561	0.8670	0.8670	0.7753

Table 1. Forensic parameters calculated with pooled allele frequencies of the Mongolian population. PIC: polymorphism information content; HOM: homozygosity; HET: heterozygosity; PE: power of exclusion; PI: paternity index; PD_M: power of discrimination in males; PD_F: power of discrimination in females; MECK_{rüger}: mean exclusion chance according to Krüger *et al.*; MECK_{Kishida}: mean exclusion chance according to Kishida *et al.*; MEC_{Desmarais}: mean exclusion chance in standard trios involving daughters according to Desmarais *et al.*; MEC_{Desmarais Duo}: mean exclusion chance in father/daughter or mother/son duos according to Desmarais *et al.*

LG4, respectively (Supplementary Table S2). Haplotype diversity (HD) of the four LGs varied from 0.9901 (in LG2) to 0.9972 (in LG1) in Mongolian males and varied from 0.9876 (in LG3) to 0.9973 (in LG1) in Eastern Chinese Han males. This illustrated that LG1 was the most polymorphic group with a frequency of 0.0259 for the most common haplotype in Mongolian males, 10-21-24.1 and 10-21-25.1 (for DXS8378-DXS10135-DXS10148), and a frequency of 0.0200 for the most common haplotype in Eastern Chinese Han males, 10-25-27.1. Haplotype 12-16-32 (for HPRTB-DXS10103-DXS10101) was the most common, observed 12 times in LG3 of the Eastern Chinese Han samples with a population frequency of 0.0600. The results showed that the four closely linked X-STR groups are highly informative genetic markers in the two studied populations.

Genetic variability. The allelic frequency and forensic parameters of Eastern Chinese Han were also presented in our previous report³³. Based on the calculated allelic frequencies of 12 X-STRs of Mongolian females and males, no significant difference was found between them ($p > 0.05$) by the Exact Test; therefore, males and females were pooled for calculating forensic parameters. The allelic frequencies of 12 X-STRs are shown in Supplementary Table S3. The results indicated that all 12 X-STR markers were polymorphic. The DXS10135 locus showed the highest forensic efficiency in both Mongolian (PIC = 0.9070) and Eastern Chinese Han populations, while DXS7423 was the least informative in both Mongolian (PIC = 0.4597) and Eastern Chinese Han populations. HET values varied from 0.5268 (DXS7423) to 0.9134 (DXS10135) in the Mongolian population. All the forensic parameters of 12 X-STRs for the Mongolian population are shown in Table 1. In total, the combined PD_F, PD_M, MEC_D and MEC_T were calculated as exceeding 0.999999999, 0.999999983, 0.9999987070 and 0.999999931, respectively. All these forensic parameters demonstrated that the 12 X-STR markers were highly polymorphic and will be useful in forensic applications or anthropological research among the Mongolian population of Inner Mongolia and the Eastern Chinese Han population.

Inter-population comparison. For a better understanding of the genetic structures of these 12 X-STRs in Chinese populations, we assembled data from five Chinese Han subpopulations and four ethnic minority groups from different areas of China (Fig. 1a, Supplementary Table S4). Detailed genotyping data and allelic frequencies of these populations assayed by Investigator Argus X-12 kit were available in previously published reports. By calculation and comparison, we noticed that all the combined PD_F of the 11 populations were greater than 0.999999999, except for the Guangdong Yao ethnic group ($cPD_F = 0.999999998$); the combined PD_M ranged from 0.999999997 to 0.999999999; and the combined MEC_D and MEC_T ranged from 0.999998065 to 0.999999993 and 0.999998732 to 0.999999998, respectively. The abovementioned forensic parameters perfectly illustrated that the 12 X-STRs had great application value in forensic identification and kinship analysis among Chinese populations.

Next, the allelic frequencies of Mongolian and Eastern Chinese Han populations were compared with nine other Chinese populations. Since the F_{st} value is usually used for analyzing genetic distances between populations^{36,37}, we calculated the pairwise F_{st} genetic distances (shown in Supplementary Table S5) among the 11 Chinese populations to study their phylogenetic relationships with each other. Based on these data, inter-population comparisons were performed: Significant differences were observed between Mongolian and Taiwanese populations at 4 loci; Mongolian and Guangdong Yao populations at 3 loci; Mongolian and Eastern Chinese Han populations at 2 loci; and for 1 locus between Mongolian and Han populations of Liaoning and Henan, Mongolian and Liaoning Korean populations, and Mongolian and Guangdong Zhuang populations. No significant difference was observed between Mongolian and Hebei Han, Guangdong Han, or Fujian She populations. The maximum F_{st} value was 0.0055 for the Guangdong Yao ethnic group and Taiwanese population.

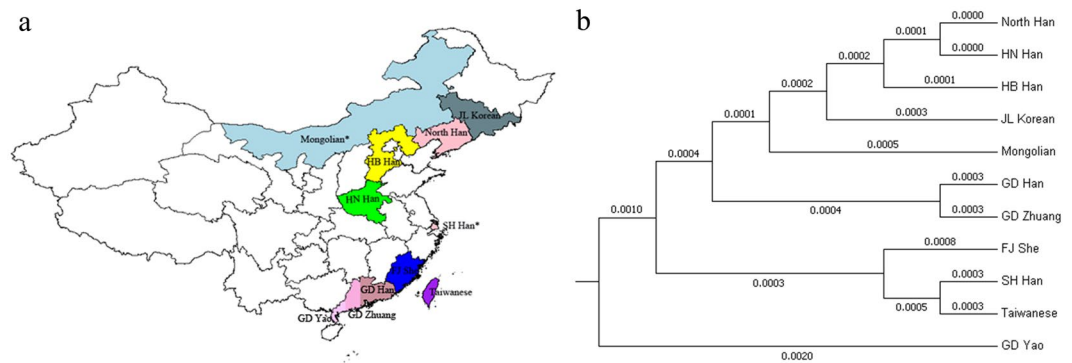


Figure 1. (a) Geographical distribution of 11 populations on the map of China, which was generated by Package (ggplot2) and (maptools) of R: A Language and Environment for Statistical (https://www.R-project.org). The areas marked by asterisks show the populations in our study. (b) Evolutionary relationships of 11 populations in different areas of China shown by a phylogenetic tree that was constructed based on 12 X-STRs.

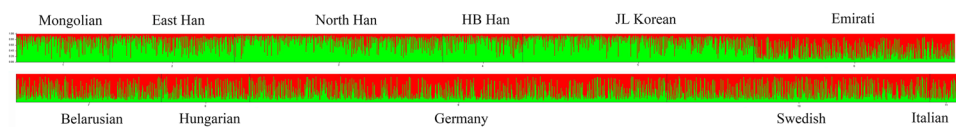


Figure 2. Estimated population genetic structure of eleven different populations at $K=2$.

Phylogenetic tree by UPGMA phylograms. Despite the limited number of loci that can provide strong information about the general level of genetic diversity, the phylogenetic analyses indicated that populations that are geographically close tend to have shorter genetic distances compared to those that are geographically far apart, which is in accordance with the conventional knowledge of population genetics³⁸. The visualized results of the observed genetic distances revealed that the Mongolian population in this study was genetically close to the populations from northern China, including the northern Han population and Korean ethnic population in the Liaoning province. On the other hand, the eastern/southern populations of China were not as related to the Mongolian population based on X-STRs analyses, regardless of ethnic identity. Of all the populations presented, the Guangdong Yao ethnic group was relatively far from the other populations according to the phylogenetic tree we constructed (shown in Fig. 1b). These observed results were consistent with the cultural and geographical background and development of the abovementioned populations.

Clustering by STRUCTURE analysis. To further investigate the population genetic structure at an individual level on a worldwide scale, the STRUCTURE program was used to process the genetic data of five representative Chinese populations (Mongolian, Han from Eastern China, Han from Liaoning province, Han from Hebei province, and Korean from Jilin province)^{10,11,16,33}, as well as six extra published populations from European and Asia^{39–44}. The range of possible K s was tested from 2 to 11. According to the Evanno method⁴⁵, when the real K is indeterminate under the interference of K s plateaus in the distribution of $L(K)$, it often locates at the modal value of the distribution of ΔK ; therefore, the most appropriate K was 2 (Supplementary Figure S1). The clustering of the populations presented an easily distinguishable geographic pattern (Fig. 2). Mongolian, Eastern Chinese Han and the other Chinese populations belonged to cluster 1, while the other six global populations belonged to cluster 2. When $K=3-11$, there were no recognizable boundaries within the Chinese populations by STRUCTURE analysis. The same situation was also observed in cluster 2 (the Emirati, Belarusian, Hungarian, Germany, Swedish and Italian populations): the five geographically close Western European populations and the Emirati from United Arab Emirates (UAE), who live in the tri-continental crossroads connecting Africa, Europe and Asia, shared a relatively consistent genetic structure pattern no matter how the K value was assumed (Supplementary Figure S2). These results conformed to the findings of Brissenden *et al.* based on an autosomal SNP study⁴⁶ and indicated that the Mongolian and Eastern Chinese Han were inseparable from other Chinese at the genetic level from STRUCTURE analysis.

Principal Coordinate Analysis (PCoA). The genetic pattern of the male X chromosome inherited from the mother, normally referred to as the haplotype data set, makes the genetic distribution of X-STRs in males worth exploring. PCoA provides a way of visualizing the essential patterns of genetic relationships contained in a special matrix (e.g., distance matrix) and allows us to find and plot the major patterns within a multivariate data set. PCoA gives a reasonable assumption at a higher taxonomic level, compared to the algorithms that always assume a hierarchical genetic structure in Tree building methods. In this respect, PCoA is an effective complement to the evolutionary tree. Therefore, PCoA was carried out to test for significant variation in the genetic distribution of 12 X-STR markers among the males of abovementioned eleven populations in the STRUCTURE analyses. As shown

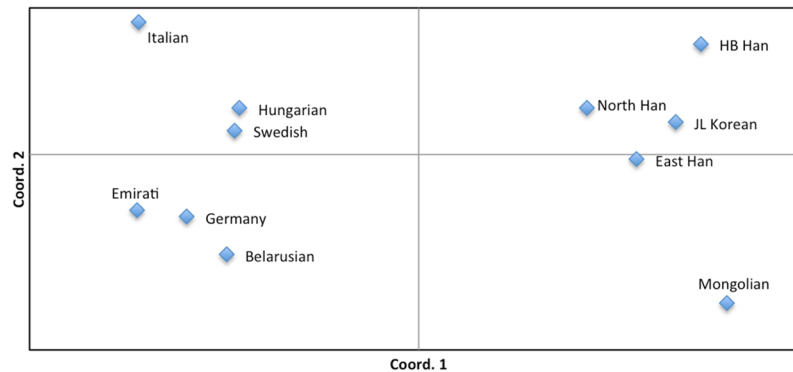


Figure 3. Principal Coordinate Analysis (PCoA) plot displaying the genetic relationships among five Chinese male populations and six male populations from Europe and Asia based on Nei's unbiased genetic distance matrix of 12 X-STRs.

in Fig. 3, the first two principal components explained 94.62% of the total variance observed within these populations (the first and second component accounted for 88.02% and 6.60%, respectively). In the PCoA diagram, Mongolian, Eastern Chinese Han and other Chinese populations were clustered together on the right, while other global populations were clustered on the left. Among the European male populations, Hungarian, Swedish and Italian were located in the upper left quadrant, while Germany and Belarusian were clustered together in the lower left quadrant, which was in accordance with the national stratification within Europe in other reports^{41,43}. Meanwhile, the Emirati males from the Arabian Peninsula of Western Asia also located in the lower left quadrant and could not be distinguished from the Europeans by the 12 X-STRs, which was consistent with the result of the STRUCTURE analysis in our study. This may due to the complex biogeography and history background of UAE. Among the cluster of Chinese populations, Hebei Han, Liaoning Han and Korean from Jilin shared a closer genetic distance, which supported the results observed from the phylogenetic tree. Additionally, the Mongolian population was located in the lower right corner, slightly far from other Chinese groups in terms of genetic relationship. Similar results were found in a study of Mongolian Y-STRs, and this probably reflects customs and territory restriction to some extent⁴⁷. Our analysis indicated that the 12 X-STR loci had some advantages in distinguishing Mongolian males from others by the PCoA method.

The Mongolian is a traditional nomadic nationality mainly distributed in Eastern Asia, it is one of the largest minority nationalities in China, and also the main nation of Mongolia. In addition, the Mongolians are also distributed in Asian and European countries such as Russia. The Mongolian nationality first originated from the eastern bank of the ancient Wangjian River. In the early thirteenth century, the Mongolian ministry, headed by Genghis Khan, unified the Mongolia region, and gradually formed a new national community. In the history of China and even in the world, the Mongolian people have had a profound influence and significant role in the economy, politics and culture. Besides, Eastern China is a densely populated area, the population there accounts for 29.3% of the total population of China, which has the most common characteristics of Chinese. Therefore, they are both very important and particular population groups in China, that's why our study focused on them. Its particular genetic inheritance that makes the X chromosome valuable for human evolutionary and historical research^{48,49}. On the one hand, there is a single copy of the X chromosome in males, it is possible to detect their haplotypes which are needed to infer the phylogeny of a region. The X/Y chromosome and mitochondrion are candidates for researches that use phylogenies, an advantage that explicates their dominant role in human historical studies⁵⁰⁻⁵². On the other hand, recombination on women X chromosome also creates vast quantities of information that is vital for providing a complete view of the history of human populations. Although the bulk of human historical researches have been done with Y chromosome or mitochondrion, X chromosome has the advantage of existing in both genders and much lower rates of genetic drift⁵⁰. The X-STR markers included in Argus X-12 kit are the most developed and widely used, there have been a great amount of 12 X-STRs data of different populations, by which we can conduct the analysis of population characteristics or geographical studies. Therefore, we conducted genetic analysis of Mongolian and Eastern Chinese Han populations based on 12 X-STR markers and achieved significant and prominent results that might be used in more precise substructure detection of Eastern Asian population and the estimation of individual biogeographical ancestry in further study.

Conclusions

Our current study indicated that the Investigator Argus X-12 kit used in Mongolian and Eastern Chinese Han sample sets provided highly polymorphic data for discriminating individuals and testing kinship, which enriched the Chinese ethnic genetic information. The significant deviations from HWE observed at the DXS10134 locus in the Mongolian population could possibly be eliminated by increasing the sample size from geographically closed regions in Inner Mongolia. The genetic comparison among eleven Chinese populations demonstrated that the statistically significant genetic distances were consistent with the known situation of these national mixed, small-settled and staggered populations living in China. In addition, neighboring populations and different ethnic groups within the same area appeared to have closer evolutionary relationships in the phylogenetic tree. The STRUCTURE analyses revealed that the Mongolian and Eastern Chinese Han populations belonged to the same cluster as the other Chinese populations, but were apparently distinct from other global populations. The PCoA

analysis indicated that the 12 X-STRs could not only be applied in distinguishing Chinese males and the males from the abovementioned countries, but also able to show a moderate genetic distance between Mongolian and the other Chinese male populations. To our best knowledge, this is the first report to demonstrate that the 12 X-STR markers could be useful in dealing with the genetic relationship between Mongolian males and others.

However, for some populations, such as the Manchu and Miao with their large populations, or the Kazak with significant differences from other Chinese populations in other genetic markers, there are still no published data concerning the 12 X-STRs panel. Therefore, further studies should be focus on increasing the sample size of current ethnic groups and including more nationalities. Hopefully, our endeavor will help to establish an appropriate and integrated X-STR database that is suitable for the large population of China.

References

1. Szibor, R. *et al.* Use of X-linked markers for forensic purposes. *Int J Legal Med.* **117**, 67–74 (2003).
2. Diegoli, T. M. Forensic typing of short tandem repeat markers on the X and Y chromosomes. *Forensic Sci Int Genet.* **18**, 140–151 (2015).
3. Wen, B. *et al.* Genetic evidence supports demic diffusion of Han culture. *Nature.* **431**, 302–305 (2004).
4. Szibor, R. X-chromosomal markers: past, present and future. *Forensic Sci Int Genet.* **1**, 93–99 (2007).
5. Poulsen, L. *et al.* NGMSelect™ and Investigator® Argus X-12 analysis in population samples from Albania, Iraq, Lithuania, Slovenia, and Turkey. *Forensic Sci Int Genet.* **22**, 110–112 (2016).
6. Sufian, A. *et al.* Genetic diversity study on 12 X-STR loci of investigator® Argus X STR kit in Bangladeshi population. *Int J Legal Med.* **131**, 963–965 (2017).
7. Ferragut, J. F., Bentayebi, K., Castro, J. A., Ramon, C. & Picornell, A. Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations. *Int J Legal Med.* **129**, 253–255 (2015).
8. Cortes-Trujillo, I. *et al.* Forensic efficiency parameters of the Investigator Argus X-12 kit in women from two Mestizo and seven Amerindian populations from Mexico. *Leg Med (Tokyo).* **26**, 62–64 (2017).
9. Mrcsic, G. *et al.* Analysis of 12 X-STR loci in the population of south Croatia. *Mol Biol Rep.* **44**, 183–189 (2017).
10. Guo, F. Population genetic data for 12 X-STR loci in the Northern Han Chinese and StatsX package as tools for population statistics on X-STR. *Forensic Sci Int Genet.* **26**, e1–e8 (2017).
11. Xu, J. *et al.* Population data of 12 X-chromosomal STR loci in Chinese Han samples from Hebei Province. *Forensic Sci Int Genet.* **7**, e43–e44 (2013).
12. Liu, Y. J., Su, W. M., Yue, J. T., Shi, S. X. & Guo, L. H. Genetic polymorphisms of twelve X chromosomal STR loci in Han population from Henan Province. *J. Forensic Med.* **29**, 297–299 (2013).
13. Zeng, X. P. *et al.* Genetic polymorphisms of twelve X-chromosomal STR loci in Chinese Han population from Guangdong Province. *Forensic Sci Int Genet.* **5**, e114–e116 (2011).
14. Chen, M. Y., Ho, C. W., Pu, C. E. & Wu, F. C. Genetic polymorphisms of 12 X-chromosomal STR loci in Taiwanese individuals and likelihood ratio calculations applied to case studies of blood relationships. *Electrophoresis.* **35**, 1912–1920 (2014).
15. Ren, Z. *et al.* The polymorphisms of 12 X-STR loci in six ethnic populations in China. *Forensic Science International. Genetics Supplement Series.* **4**, e49–e50 (2013).
16. Li, X. N. *et al.* Population data of 12 X-chromosome STR loci in Yanbian Korean samples from China. *Forensic Sci Int Genet.* **13**, 143–144 (2014).
17. Wu, X. J. *et al.* Genetic polymorphisms of 12 X-STR loci in Guangdong Zhuang population. *Forensic Science and Technology.* **4**, 23–26 (2014).
18. Wu, X. J. *et al.* Genetic polymorphisms of 12 X-STR loci in Guangdong Yao population. *Chin. J. Forensic Med.* **4**, 365–367 (2014).
19. Lin, M. *et al.* Population genetics investigation of 12 X chromosome short tandem repeats in the She ethnic minority in Fujian. *China J. Fam. Plann.* **10**, 628–631 (2011).
20. *China Statistical Yearbook* (National Bureau of Statistics of China, 2016).
21. Walsh, P. S., Metzger, D. A. & Higuchi, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques.* **10**, 506–513 (1991).
22. Promega Biotech Co. Ltd. *Powerstates V12.xls*. [EB/OL] <https://www.promega.com/geneticidtools/>.
23. Lischer, H. & Excoffier, L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* **3**, 564–567 (2010).
24. Desmarais, D., Zhong, Y., Chakraborty, R., Perreault, C. & Busque, L. Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). *J Forensic Sci.* **43**, 1046–1049 (1998).
25. Sneath, P. H. A. & Sokal, R. R. Numerical taxonomy. *W. H. Freeman and Co., San Francisco, Calif.* (1973).
26. Sudhir, K., Stecher, G. & Koichiro, T. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol.* **33**, 1870–1874 (2016).
27. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics.* **164**, 1567–1587 (2003).
28. Hubisz, M. J., Falush, D., Stephens, M. & Pritchard, J. K. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour.* **9**, 1322–1332 (2009).
29. Earl, D. A. & VonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* **4**, 359–361 (2012).
30. Nei, M. Genetic Distance between Populations. *The American Naturalist.* **106**, 283 (1972).
31. Nei, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics.* **89**, 583–590 (1978).
32. Orloci, L. & Austin, M. P. Multivariate Analysis in Vegetation Research. *J. Ecol.* **68**, 331 (1976).
33. Zhang, S., Zhao, S., Zhu, R. & Li, C. Genetic polymorphisms of 12 X-STR for forensic purposes in Shanghai Han population from China. *Mol Biol Rep.* **39**, 5705–5707 (2012).
34. Poulsen, L. *et al.* Population and forensic data for three sets of forensic genetic markers in four ethnic groups from Iran: Persians, Lurs, Kurds and Azeris. *Forensic Sci Int Genet.* **17**, 43–46 (2015).
35. Tomas, C. *et al.* Results for five sets of forensic genetic markers studied in a Greek population sample. *Forensic Sci Int Genet.* **16**, 132–137 (2015).
36. Petrosian, V. G., Tokarskaia, O. N., Kashentseva, T. A., Korochkin, L. I. & Ryskov, A. P. Assessment of population differentiation using DNA fingerprinting and modified Wright's Fst-statistics. *Genetika.* **39**, 229–235 (2003).
37. Boca, S. M. & Rosenberg, N. A. Mathematical properties of Fst between admixed populations and their parental source populations. *Theor Popul Biol.* **80**, 208–216 (2011).
38. Prieto-Fernandez, E. *et al.* A genetic overview of Atlantic coastal populations from Europe and North-West Africa based on a 17 X-STR panel. *Forensic Sci Int Genet.* **27**, 167–171 (2017).
39. Edelmann, J., Lutz-Bonengel, S., Naue, J. & Hering, S. X-chromosomal haplotype frequencies of four linkage groups using the Investigator Argus X-12 Kit. *Forensic Sci Int Genet.* **6**, e24–e34 (2012).
40. Tillmar, A. O. Population genetic analysis of 12 X-STRs in Swedish population. *Forensic Sci Int Genet.* **6**, e80–e81 (2012).

41. Horvath, G., Zalan, A., Kis, Z. & Pamjav, H. A genetic study of 12 X-STR loci in the Hungarian population. *Forensic Sci Int Genet.* **6**, e46–e47 (2012).
42. Inturri, S., Menegon, S., Amoroso, A., Torre, C. & Robino, C. Linkage and linkage disequilibrium analysis of X-STRs in Italian families. *Forensic Sci Int Genet.* **5**, 152–154 (2011).
43. Rebala, K. *et al.* Variation of X-chromosomal microsatellites in Belarus within the context of their genetic diversity in Europe. *Forensic Sci Int Genet.* **16**, 105–111 (2015).
44. Almarri, M. A. & Lootah, R. A. Allelic and haplotype diversity of 12 X-STRs in the United Arab Emirates. *Forensic Sci Int Genet.* (2017).
45. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* **14**, 2611–2620 (2005).
46. Brissenden, J. E. *et al.* Mongolians in the Genetic Landscape of Central Asia: Exploring the Genetic Relations among Mongolians and Other World Populations. *Hum Biol.* **87**, 73–91 (2015).
47. Guo, F. Population genetics for 17 Y-STR loci in Mongolian ethnic minority from Liaoning Province, Northeast China. *Forensic Sci Int Genet.* **17**, 153–154 (2015).
48. Meisel, R. P. & Connallon, T. The faster-X effect: integrating theory and data. *Trends Genet.* **29**, 537–544 (2013).
49. Fan, G., Ye, Y. & Hou, Y. Detecting a hierarchical genetic population structure via Multi-InDel markers on the X chromosome. *Sci Rep.* **6**, 32178 (2016).
50. Schaffner, S. F. The X chromosome in population genetics. *Nat Rev Genet.* **5**, 43–51 (2004).
51. Kivisild, T. The study of human Y chromosome variation through ancient DNA. *Hum Genet.* **136**, 529–546 (2017).
52. Mitchell, S. L. *et al.* Characterization of mitochondrial haplogroups in a large population-based sample from the United States. *Hum Genet.* **133**, 861–868 (2014).

Acknowledgements

This study was supported by grants from the National Key R&D Program of China (No. 2016YFC0800703), the National Natural Science Fund for Distinguished Young Scholars (No. 81625013), Standard Program of Shanghai Municipality (No. 16DZ0501600), and Public Interest Research Grant Program of National Research Institutes (No. GY2017D-2). The funders had no role in study design, data analysis, publishing decisions, or manuscript preparation.

Author Contributions

R.T. and J.Z. wrote the main manuscript text, R.T. and X.L. analyzed the results, Y.B. and R.D. conducted the casework, C.J. and R.Z. conducted the experiment, S.Z. and C.L. designed the study. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-22665-3>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018