Int J Legal Med (2011) 125:33–37
DOI 10.1007/s00414-009-0414-z

ORIGINAL ARTICLE

Genetic analysis of eight population groups living in Taiwan using a 13 X-chromosomal STR loci multiplex system

Hsiao-Lin Hwa · James Chun-I Lee · Yih-Yuan Chang · Hsiang-Yi Yin · Ya-Hui Chen · Li-Hui Tseng · Yi-Ning Su · Tsang-Ming Ko

Received: 27 February 2009 / Accepted: 21 December 2009 / Published online: 29 April 2010

Abstract A 13 X-chromosomal short tandem repeat (STR) multiplex system (DXS6807, DXS8378, DSX9902, DXS7132, DXS9898, DXS6809, DXS6789, DXS7424, DXS101, GATA172D05, HPRTB, DXS8377, and DXS7423) was tested on 1,037 DNA samples from eight population groups currently living in Taiwan. Different distributions of the allelic frequencies in different populations were presented. DXS8377 and DXS101 were the two most polymorphic loci in these eight populations, whereas DXS7423 was the least informative marker in most of the populations studied. The genetic distances between the populations and the constructed phylogenetic tree revealed a long genetic distance between Asian and Caucasian populations as well as isolation of the Tao population. The phylogenetic tree grouped populations into clusters compatible with their ethnogeographic relationships. This 13 X-chromosomal short tandem repeat multiplex system offers a considerable number of polymorphic patterns in different populations. This system can be useful in forensic identification casework and ethnogeographic research.

Keywords Forensic identification · Multiplex PCR · X chromosome · Short tandem repeat · Genetic distance · Phylogenetic tree

Electronic supplementary material The online version of this article (doi:10.1007/s00414-009-0414-z) contains supplementary material, which is available to authorized users.

H.-L. Hwa · J. C.-I. Lee · H.-Y. Yin · Y.-H. Chen
Department and Graduate Institute of Forensic Medicine,
College of Medicine, National Taiwan University,
Taipei, Taiwan

H.-L. Hwa
Department of Obstetrics and Gynecology,
National Taiwan University Hospital and College of Medicine,
National Taiwan University,
Taipei, Taiwan

H.-L. Hwa · L.-H. Tseng · Y.-N. Su
Department of Medical Genetics,
National Taiwan University Hospital,
Taipei, Taiwan

J. C.-I. Lee
Institute of Forensic Medicine, Ministry of Justice,
Republic of China,
Taipei, Taiwan

Y.-Y. Chang · T.-M. Ko (✉)
Genephile Bioscience Laboratory,
No. 10, 1 F, Lin Sen South Road, Zhong Zheng District,
Taipei, Taiwan 100
e-mail: tsangming@gigigaga.com

Introduction

Analysis of the human variation of microsatellites of the X chromosome has become a useful tool in forensic practice in recent years. The X-chromosome short tandem repeat (X-STR) markers are particularly valuable in maternity testing and in detecting father-daughter relationships. Increasing numbers of X-STRs markers among different populations have now been published [1–7]. Our group has described the characterization of a new multiplex system modified from a report of Gomes et al [5] which allows the simultaneous analysis of 13 X-STRs and can be applied in forensic casework [8]. The variation of allelic polymorphisms of X-STR in different populations of the world is still unclear. Establishing an X-STR allelic frequency database and evaluating the extent of difference of X-linked markers for forensic and ethnogeographic studies are important issues. Different allelic frequency distributions of

STR between different populations can be used to reconstruct human phylogenies.

The population in Taiwan is heterogeneous and is made up of two major groups, the indigenous population (approximately 480,000 persons or 2% of the total population) and the Han population (about 22,470,000 or 98%) [9]. The indigenous population is comprised of 13 tribes living mainly in the mountainous regions of Taiwan Island and a specific tribe (Tao) living on Orchid Island. The Taiwanese Han population includes the descendants of individuals who migrated from southern China from the 1600s to the early 1900s, and a large-scale migration from areas throughout mainland China in 1949. There are also populations from other countries presently working or living in Taiwan (about 560,000) [10]. For these population groups, the available data on X chromosomal genetic diversity and population structure is still limited.

This present work is an analysis of the genotyping results of 13 X-STR loci of different populations living in Taiwan. The aim of this study was to present and compare the distribution of the allelic frequencies of these population groups and evaluate the genetic distances between them.

Materials and methods

Sample sources and DNA extraction

This retrospective study was approved by the Institute Review Board. A total of 1,037 DNA samples composed of 258 apparently healthy unrelated Taiwanese Han (Tw; 118 males, 140 females), 165 indigenous Taiwanese living mainly in the mountainous regions (Ti; 96 males, 69 females), 99 Taos living on Orchid Island (Tao; 51 males, 48 females), 192 mainland Chinese (Cn; 56 males, 136 females), 114 Filipinos (Ph; 50 males, 64 females), 54 Thais (Th; 17 males, 37 females), 122 Vietnamese (Vn; three males, 119 females), and 33 Caucasians (Cau; 24 males, nine females) were analyzed. The blood samples and/or buccal swab samples were obtained from volunteer donors with informed consent between 1993 and 2007. Standard methods of phenol chloroform isoamyl alcohol extraction and the QIAamp blood kit (Qiagen, Hilden, Germany) were used for DNA extraction from peripheral whole blood samples, whereas the Viogene Blood and Tissue Genomic DNA extraction Miniprep system (Viogene, Taipei, Taiwan) was used for DNA extraction from buccal cells.

X chromosome marker typing

One multiplex polymerase chain reaction (PCR) was performed with the previously described primer sets [8]. The 13 X-STRs DXS6807, DXS8378, DSX9902, DXS7132, DXS9898, DXS6809, DXS6789, DXS7424, DXS101, GATA172D05, HPRTB, DXS8377, DXS7423, and amelogenin were genotyped in one PCR reaction simultaneously, based on the methodology of Hwa et al [8].

The DNA sample with a missing allele was sequenced using the ABI Big Dye Terminator version 3.1 Cycle Sequencing Ready Kit (Applied Biosystems, Foster City, CA, USA). The product was detected in the ABI PRISM 3100 Genetic Analyzer electrophoresis system and analyzed with sequencing analysis 3.7 software (Applied Biosystems).

Statistical analysis

Allelic frequencies, haplotype and gene diversities, the exact test of population differentiation, population pair-wise genetic distance F_{ST} and R_{ST} and analysis of molecular variance were calculated using the Arlequin version 3.11 software [11]. Linkage disequilibrium analyses by an exact test were performed using the GENEPOP (version 3.4) software package [12]. Forensic efficiency parameters of each locus were calculated following Desmarais et al [13] using the formula offered at http://www.chrx-str.org/. The combined power of discrimination of all loci was assessed according to Poetsch et al [3]. The phylogenetic tree was constructed using the neighbor-joining methods with the PHYLIP program [14].

Results and discussion

The 13 X-linked STR markers were successfully amplified in one single PCR multiplex reaction. Allele nomenclature used for genotyping for most loci followed the description by Szibor et al [15]. A single nucleotide polymorphism of the GATA172D05R primer region was found in one Caucasian sample. The C→T transition in the primer region was confirmed by DNA sequencing with a newly designed primer set GATA172D05Fs and GATA172D05Rs, as shown in Fig. 1. DNA sequencing of the primer region has to be performed when a missing allele is noted at a certain STR locus.

The allele frequencies of the 13 X-STR loci studied in these eight population groups were analyzed (Table in ESM 1). The data from part of the Taiwanese Han (221 subjects) has been described in our previous report [8]. In the DNA samples from the 415 male subjects in this study, no identical haplotype-like allelic combination of the 13 X-STRs was found. Only one new allele with 14 repeats at the DXS9902 locus in one Filipino has not been described previously. DXS8377 was found to be the most polymorphic locus in most of the population groups followed by

```
           1                                                  40
Consensus  CTGTATGATC  CAGCAATTCT  CCTTCTGGGT  TTATACCCCA
1110015    ..........  ..........  ..........  ..........
           41                                                 80
Consensus  AATAATTGAA  AGCCCGGATT  CAAAAAGATC  TATATCTATA
1110015    ..........  ....T.....  ..........  ..........
           81                                                120
Consensus  GATATAGGTA  TTGATATAGC  TCTATCTATC  TATCTATCTA
1110015    ..........  ..........  ..........  ..........
           121                                               160
Consensus  TCTATCTATC  TATCTATCTA  TCTATATATC  TGTGCAACCA
1110015    ..........  ..........  ..........  ..........
           161
Consensus  TCACCACTAT  CC
1110015    ..........  ..
```

Fig. 1 Sequencing results for locus GATA172D05 in consensus cases with an 11 repeats allele and a case (sample No. 1110015 with an 11 repeats allele) with a single nucleotide polymorphism at the GATA172D05R primer region

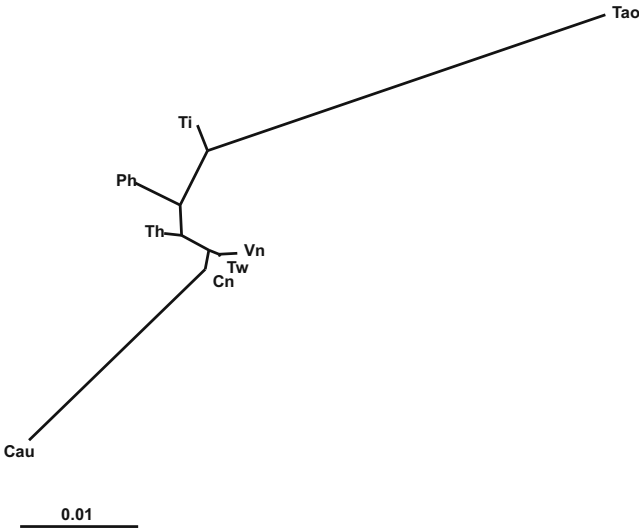

Fig. 2 Phylogenic tree based on genetic distance for the 13 X-STR system constructed by the neighbor-joining method

DXS101. These findings are consistent with previous reports [3, 5]. The least discriminating locus was DXS7423 in all population groups except the Tao and Caucasian populations. In the Tao population, DXS9898 is the least discriminating locus, with only two alleles revealed. In the Caucasian group, DXS9902 was found to be the least polymorphic locus, which is consistent with a previous report from a German population [1].

Because the number of Caucasians in this study is small, we compared our data of Caucasian group with previous reports. Compared to the data of Hispanic group described by Gomes et al., the allele distribution is similar to our data at eight (DXS8378, DXS7132, DXS9898, DXS6809, DXS6789, DXS101, GATA172D05, and DXS7423) of ten X-STR loci. At loci HPRTB, the most frequent allele was 14 repeats in their Hispanic group, whereas the most frequent allele was 13 repeats in our Caucasian group. At loci DXS8377, the most frequent allele was 49 repeats in their Hispanic group, whereas the most frequent allele was 48 repeats in our Caucasian group [5]. Compared to the data of a German population described by Edelmann, the allele distribution is similar to our data at 10 (DXS6807, DXS8378, DXS9902, DXS7132, DXS9898, DXS6789, DXS7424, DXS101, HPRTB, and DXS8377) of 12 X-STR loci. However, the most frequent alleles were nine repeats at loci GATA172D05 and 12 repeats at loci DXS7423 in the German population, whereas the most frequent alleles were ten repeats at loci GATA172D05 and 15 repeats at loci DXS7432 in our Caucasian group [1]. Compared to other Asian population, the allele distributions of Taiwanese Han are similar to that of a Korean population at ten (DXS6807, DXS9902, DXS7132, DXS 9898, DXS6789, DXS101, GATA172D05, HPRTB, DXS7423, and DXS8377) of 11 loci [15]. At loci DXS8378, the most frequent allele was nine repeats allele (57.3%) followed by ten repeats allele (28.1%) in the Korean population [16]. In our study, the most frequent allele of loci DXS8378 was ten repeats allele in Taiwanese Han (50.5%) and mainland Chinese (53.1%). The nine repeats allele was infrequent in both Taiwanese Han (2.51%) and mainland Chinese (2.13%) in this study.

No evident population-specific allele was found in this 13 X-STR system. However, some differences of distribu-

Table 1 Genetic distance matrix among eight population groups on 13 X-STR loci

	Tw	Ti	Tao	Cn	Ph	Th	Vn	Cau
Tw	-							
Ti	0.01224*	-						
Tao	0.04588*	0.03795*	-					
Cn	0.00239*	0.01643*	0.04679*	-				
Ph	0.00938*	0.01156*	0.04510*	0.01284*	-			
Th	0.00413	0.00724*	0.04044*	0.00466	0.00664	-		
Vn	0.00178	0.01570*	0.04995*	0.00344*	0.00924*	0.00387*	-	
Cau	0.02359*	0.03548*	0.07049*	0.02096*	0.03257*	0.01978*	0.02643*	-

*p value of coancestry coefficient (Fst) <0.05

tion of allelic frequencies between population groups were found. A higher frequency of a specific allele at a certain locus in some populations, such as 14 repeats of DXS6807, 14 repeats of DXS7424, 25 repeats of DXS101, nine repeats of GATA172D05, and 16 repeats of DXS7423 in the Tao, has been noted in this study. For the Caucasians, a 20 repeats allele was the most frequent allele for locus DXS6789, whereas for the other Asian population groups, a 16 repeats allele was the most frequent allele for the same locus. This finding is consistent with the high frequency of 16 repeats of DXS6789 in a Korean population and the high frequency of 20 repeats in a German population described previously [1, 16]. We found that the allelic distribution of DXS6789 of Asians was different from that of Caucasian populations. According to our observation and previous reports, the frequency of allele 8.3 at loci DXS9898 in Caucasian is higher than that in Asian populations [1, 3, 5, 16].

The forensic statistical evaluation parameters were calculated (Table in ESM 2). There is high genetic diversity in these eight populations for the 13 X-STRs analyzed. High values of mean exclusion chances in trios involving daughters (MEC_T) in all eight populations support the potential of this 13 X-STR multiplex system to be a specific kinship analysis context when father-daughter relationships are being investigated. As high discrimination power values were obtained for this X-STR system, it can be used in forensic identity and paternity testing.

The exact test for linkage disequilibrium for all pairs of loci did not show consistent evidence of an association between these 13 X-STR markers in these population groups. A haplotype cluster group with DXS6801, DXS6809, and DXS6789 and a linkage between DXS7424 and DXS101 has been described previously [17, 18]. However, linkage analysis did not confirm the existence of an association between these haplotype groups in this study. The lack of association between these X-STR loci increases the power of discrimination of this multiple X system. Because the results of linkage disequilibrium analysis depend on sample size, a larger sample in each population group is necessary for further linkage disequilibrium studies.

Genetic distance analysis was used to evaluate the population differentiation of these eight population groups. (Table 1) The value of the genetic distance represented the evolutionary distance or the migration route between two groups. The results showed that the Taiwanese Hans had a relatively shorter genetic distance from the Vietnamese, Mainland Chinese, and Thais. The Taiwanese Han population is close to the Mainland Chinese population, which is compatible with the migration history of the Han population from mainland China to Taiwan.

Figure 2 illustrates the phylogenetic tree constructed by the neighbor-joining method of the 13 X-STR loci of these eight populations. The grouping together of the Taiwanese Han, Mainland Chinese, and Vietnamese showed the ethno-geographic relationship between these populations. The significant difference of the Tao of Orchid Island from other populations revealed the geographical isolation of this population. The clustering of indigenous Taiwanese and Filipinos together is consistent with a previous report of Asian population migration [19]. The unique branch for the Caucasians represented the racial difference between Caucasian and Asian populations.

The neighbor-joining tree in this study is similar to the dendrogram of the populations established from the genetic distance matrices of the autosomal forensic STRs and FUT2 gene reported previously [19, 20]. However, this is different from the neighbor-joining trees for populations based on Alu nuclear insertion polymorphisms, or based on the D-loop or the MTCYB gene of mitochondrial DNA [21–23]. The patterns of neighbor-joining trees established from different genetic markers on autosomal chromosome, sex chromosomes, or mitochondrial DNA may vary [24].

In conclusion, this 13 X-chromosomal STR multiplex system offers a considerable number of polymorphic patterns among different populations. The distributions of the allelic frequencies at certain loci differ between different population groups. The genetic distance between populations can be estimated and the phylogenetic tree can be constructed according to this system. It may be useful in forensic identification analysis and ethno-geographic research.

Acknowledgments The authors thank Ms Hong-Ciao Wan, Ms Pi-Mei Hsu, Ms Shwu-Fang Li, Ms Yi-Li Liu, and Ms I-Chin Huang for technical support on DNA extraction and DNA sequencing. This work was partially supported by a grant from the Department of Medical Research at National Taiwan University Hospital.

References

1. Edelmann J, Hering S, Michael M et al (2001) 16 X-chromosome STR loci frequency data from a German population. Forensic Sci Int 124:215–218
2. Bini C, Ceccardi S, Ferri G et al (2005) Development of a heptaplex PCR system to analyze X-chromosome STR loci from five Italian population samples. A collaborative study. Forensic Sci Int 153:231–236
3. Poetsch M, Petersmann H, Repenning A, Lignitz E (2005) Development of two pentaplex systems with X-chromosomal STR loci and their allele frequencies in a northeast German population. Forensic Sci Int 155:71–76
4. Asamura H, Sakai H, Kobayashi K, Ota M, Fukushima H (2006) MiniX-STR multiplex system population study in Japan and application to degraded DNA analysis. Int J Legal Med 120:174–181
5. Gomes I, Prinz M, Pereira R et al (2007) Genetic analysis of three US population groups using an X-chromosomal STR decaplex. Int J Legal Med 121:198–203

6. Rodrigues EMR, Leite FPN, Hutz MH, de JBF PT, Santos AKCR, Santos SEB (2008) A multiplex PCR for 11 X chromosome STR markers and population data from a Brazilian Amazon region. Forensic Sci Int Genet 2:154–158
7. Zarrabeitia MT, Alonso A, Martin J et al (2006) Study of six X-linked tetranucleotide microsatellites: population data from five Spanish regions. Int J Legal Med 120:147–150
8. Hwa HL, Chang YY, Lee JCI et al (2009) Thirteen X-chromosomal short tandem repeat loci multiplex data from Taiwanese. Int J Legal Med 123:263–269
9. Council of Indigenous Peoples, Executive Yuan, Taiwan. http://www.apc.gov.tw/chinese/.Accessed 11 Nov, 2008
10. Department of Statistics, Ministry of the Interior, Taiwan. http://www.moi.gov.tw/stat/index.aspx. Accessed 11 Nov 2008
11. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50
12. Raymond M, Rousset F (1995) GENEPOP (version 1.2): a population genetics software for exact tests and eucumenicism. J Hered 86: 248-249 (http://www.cefe.cnrs-mop.fr/)
13. Desmarais D, Zhong Y, Chakraborty R, Perreault C, Busque L (1998) Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). J Forensic Sci 43:1046–1049
14. Felsenstein J (1986) PHYLIP (phylogeny inference package) Version 3.5c, University of Washington
15. Szibor R, Edelmann J, Hering S et al (2003) Cell line DNA typing in forensic genetics—the necessity of reliable standards. Forensic Sci Int 138:37–43
16. 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
17. Szibor R, Hering S, Kuhlisch E et al (2005) Haplotyping of STR cluster DXS6801-DXS6809-DXS6789 on Xq21 provides a powerful tool for kinship testing. Int J Legal Med 119:363–369
18. Edelmann J, Hering S, Kuhlisch E, Szibor R (2002) Validation of the STR DXS7424 and the linkage situation on the X-chromosome. Forensic Sci Int 125:217–222
19. Chang JG, Ko YC, Lee JC et al (2002) Molecular analysis of mutations and polymorphisms of the Lewis secretor type alpha (1, 2)-fucosyltransferase gene reveals that Taiwan aborigines are of Austronesian derivation. J Hum Genet 47:60–65
20. Lee HY, Park MJ, Jeong CK et al (2004) Genetic characteristics and population study of 4 X-chromosomal STRs in Koreans: evidence for a null allele at DXS9898. Int J Legal Med 118:355–360
21. Melton T, Clifford S, Martinson J, Batzer M, Stoneking M (1998) Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. Am J Hum Genet 63:1807–1823
22. Ingman M, Gyllensten U (2003) Mitochondrial genome variation and evolutionary history of Australian and New Guinean aborigines. Genome Res 13:1600–1666
23. Hwa HL, Ko TM, Chen YC, Chang YY, Tseng LH, Su YN, Lee JCI (2008) Study of the cytochrome b gene sequence in populations of Taiwan. J Forensic Sci (in press)
24. Kayser M, Lao O, Saar K et al (2008) Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. Am J Hum Genet 82:194–198