POPULATION DATA



Genetic diversity and haplotype analysis of Guizhou Miao identified with 19 X-chromosomal short tandem repeats

Yanyan Han¹ • Guanglin He² • Sheng Gong³ • Junying Chen⁴ • Zhigang Jiang¹ • Pengyu Chen^{5,6}

Received: 27 March 2018 / Accepted: 24 May 2018 / Published online: 30 May 2018 © Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Keywords X-chromosomal STRs · Miao population · Population genetics · Haplotype diversity

X-chromosomal short tandem repeats (STRs) with the advantages of both autosomal and uni-parental biomarkers have been widely used in the forensic deficiency cases and other complex kinship identification [1]. Chinese Miao (also called Hmong) is the fourth largest minority ethnic group officially

Yanyan Han and Guanglin He contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00414-018-1871-z) contains supplementary material, which is available to authorized users.

Pengyu Chen pychenfs@163.com

- ¹ School of Public Health, Zunyi Medical University, Zunyi 563099, Guizhou, China
- ² Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, Sichuan, China
- ³ Criminal Investigation Detachment, Qianxinan Prefecture Public Security Bureau, Xingyi, Guizhou, China
- ⁴ Judicial Expertise Center, Zunyi City Public Security Bureau, Zunyi 563099, Guizhou, China
- ⁵ Center of Forensic Expertise, Affiliated Hospital of Zunyi Medical University, Zunyi 563099, Guizhou, China
- ⁶ School of Forensic Medicine, Zunyi Medical University, Zunyi 563099, Guizhou, China

recognized by the government of the People's Republic of China. The population size of Chinese Miao is approximately 12 million over the world and around 9 million in China. Chinese Miao live primarily in remote Southwest China's mountains, in the provinces of Guizhou, Yunnan, Sichuan, Hubei, Hunan, Guangxi, Guangdong, and Hainan. The Miao language belongs to the Hmong-Mien linguistic family. Historically, contemporary Miao is the descendants of the Jiuli tribe led by Chiyou and has experienced five largescale migrations. Archeological evidence from the Daxi Culture (5300–6000 years ago) led to the hypothesis that Miao was the first people to settle in present-day China. The language formations and changes of present-day Miao ethnicity have genetic assimilation of neighboring ethnically diverse populations [2].

In the presented study, we used the AGCU X19 kit to genotype the 19 X-STRs belonging to seven linkage groups [3–5] (DXS10074-DXS10075-DXS10079-DXS7132, DXS101-DXS7424, DXS10101-DXS10103-HPRTB, DXS10134-DXS7423, DXS10135-DXS10148-DXS8378, DXS10159-DXS10162-DXS10164, and DXS6789-DXS6809) in 268 volunteers (117 females and 151 males) who self-identified themselves as Miaos and compared the genetic relationship with 12 reference populations. All peripheral blood samples reported were obtained from Zunyi City in Guizhou Province (Southwest China) under the protocols

approved by Ethnics Committee in the Zunyi Medical University. Human DNA was isolated using the salt-outing method [6] and genotyped using the AGCU X19 STR Fluorescent Detection Kit on a Mastercycler Pro® Thermal Cycler (Eppendorf, Germany) on the basis of the manufacturer's instruction. Separation and detection of amplification products were carried out on the Applied Biosystems 3130 Genetic Analyzer (Applied Biosystems), and allele allocation and nomenclature were analyzed by GeneMapper ID V3.2 software.

We calculated allele frequency distributions using the modified PowerState V1.2 spreadsheet and forensic parameters (Polymorphism Information Content, PIC; Probability of Exclusion, PE; Paternity Index, PI; Probability of Discrimination, PD, and Mean Paternity Exclusion Change, MEC) via online tool implemented in ChrX-STR.org 2.0 (http://www.chrx-str.org). Arlequin v3.513 [7] was used to estimate the p values of Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD), as well as the observed heterozygosity (Ho) and expected heterozygosity (He) in females. The frequencies of seven linkage male haplotypes were calculated using the direct counting method based on the observed genotypes. Haplotype diversity (HD) and match probability (MP) were obtained respectively using the following formula: HD = $\frac{N}{N-1} (1 - \sum P i^2)$ and MP = $\sum P i^2$, where *n* denotes the population size and Pi means the haplotype frequency. Discrimination capability was calculated as the ratio of haplotype number to sample size. We compared the genetic differentiation between Guizhou Chinese Miao and other 12 reference Chinese populations defined by ethnic/linguistic boundaries and administrative divisions. The pairwise Nei's genetic distances were calculated with the Gendist program in the Phylip3.695 packages. Principal component analysis (PCA) based on allele frequency distributions of 19 X-STRs was carried out using the MVSP software. Multidimensional scaling plots (MDS) and neighboring-joining (N-J) tree were constructed respectively using SPSS software and Mega 7.0 [8] on the basis of Nei's genetic matrix.

All 19 X-STRs are consistent with the HWE in our investigated Guizhou Miao ethnicity (**Table S1**). Among 171 pairwise comparisons, ten pairs are identified with linkage inheritance after Bonferroni correction (**Table S2**). A total of 217 alleles are discovered with the allele frequencies ranging from 0.0026 to 0.5870 (**Table S3**). As shown in **Table S1**, the Ho and He respectively vary from 0.5471 (DXS7423) to 0.9145 (DXS10135), and 0.5171 (DXS7423) to 0.9288 (DXS10135). The PIC varies from 0.4283 to 0.9151 and PI varies from 0.0397 to 0.2394. The combined PE is 0.9999999922. The combined PDs in male and female are 0.999999999999999994 and 0.999999999988 respectively. Besides, the MECs based on previously published algorithms of Krüger, Kishida, Desmarais, and Desmarais (in Duos) are 0.999999963638, 0.999999999999997, 0.99999999999997, and 0.999999993459, respectively. Seven linkage groups can be clustered among 19 markers based on physical localizations and previous studies [3–5]. The haplotype data and corresponding forensic parameters in males are presented in **Table S4**. There are 124, 37, 86, 27, 119, 46, and 71 distinct haplotypes in LG1-LG7, respectively. The MP values range from 0.0089 (in the linkage group of DXS10135-DXS10148-DXS8378) to 0.0738 (in the linkage group of DXS10134-DXS7423). The DC and HD vary from 0.1788 (DXS10134-DXS7423) to 0.8212 (DXS10074-DXS10075-DXS10079-DXS7132), and from 0.9324 (DXS10134-DXS7423) to 0.9968 (DXS10135-DXS10148-DXS8378). Our results show that the 19 X-STR loci and seven linkage groups are highly informative and polymorphic in Chinese Miao population.

To better understand the genetic background of Chinese Miao, we merged our allele frequency distribution data of 19 X-STRs with previously published data and constructed a new dataset consisting of 13 populations from eight Chinese ethnic groups belonging to three language families (Sino-Tibetan, Turkic, and Hmong-Mien language families). The pairwise Nei's genetic distances between Chinese Miao and the other 12 reference populations are listed in Table S5. A distant genetic relationship is identified between Guizhou Chinese Miao and Xinjiang Uyghur population (0.0460), and the intimate genetic relationship is observed between Chinese Miao and Guanzhong Han Chinese (0.0108). A total of 49.067% genetic variance can be explained by the first two principal components. PCA1 (29.889%) shows a clear separation between three Turkic-speaking populations and others, and PCA2 (19.178%) can distinguish four Tibeto-Burmanspeaking populations from the rests, especially for Tibetan populations (Fig. S1A). Furthermore, MDS constructed according to the Nei's genetic distance matrix is used to explore the genetic homogeneity and heterogeneity among the 13 populations. As shown in Fig. S2B, three Tibetan populations and the Liangshan Yi Chinese population are clustered together and located in the upper right quadrant. Two Xinjiang Uyghur populations and Xinjiang Kazakh populations are allocated in the left (second and third) quadrant. Our focus, Guizhou Miao population, keeps a genetic affinity with Xinjiang Xibo population and is located in the lower left quadrant and cannot be distinguished from Sinitic-speaking populations. Finally, an N-J tree is constructed to further dissect the genetic differentiation among 13 populations along ethnic and linguistic boundaries (Fig. S1C). Three genetic clusters are observed: an Altaic-speaking cluster consisting of three Turkic-speaking populations and one Tungusic-speaking population of Xinjiang Kazakh; a Tibeto-Burman-speaking cluster comprising three Tibetan groups and Liangshan Yi; a Sinitic-speaking and Hmong-Mien speaking admixture cluster consisting of three Han populations, one Hui population

and our focus population, the Guizhou Chinese Miao population. The Guizhou Chinese Miao is identified to be most close to the Sichuan Han population and then to Guanzhong Han population.

Our study provided the first batch of X-chromosomal STR data from Miao ethnicity, thus enriches the Chinese ethnic genetic information. The 19 X-STRs and seven linkage groups are informative and powerful markers to distinguish individuals and complex testing kinship. Comprehensive population comparisons on the genetic variations of 19X-STRs through the MDS, PCA and N-J tree indicate that as a representative Hmong-Miao-speaking population the Guizhou Miao has a close genetic relationship with surrounding Sinitic-speaking populations, especially with Sichuan and Guanzhong Han populations. In a summary, genetic differentiations exist among linguistic diverse populations and genetic affinity exists among populations within the same language family.

Funding information This study received the support of grants from the National Natural Science Foundation of China (No. 81401562), the Fundamental Research Funds for the Central University (2012017yjsy187), the Science and Technology Foundation of Guizhou Province (No. 2013-2313), and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (No. 2014-1685).

Compliance with ethical standards

This study had been approved by the institutional review boards at Zunyi Medical University. Informed consent was obtained from volunteers before sample collection.

Competing interests The authors declare that they have no competing interests.

References

- Szibor R (2007) X-chromosomal markers: past, present and future. Forensic Sci Int Genet 1:93–99. https://doi.org/10.1016/j.fsigen. 2007.03.003
- Wen B, Li H, Gao S, Mao X, Gao Y, Li F, Zhang F, He Y, Dong Y, Zhang Y, Huang W, Jin J, Xiao C, Lu D, Chakraborty R, Su B, Deka R, Jin L (2005) Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. Mol Biol Evol 22: 725–734. https://doi.org/10.1093/molbev/msi055
- He G, Li Y, Zou X, Li P, Chen P, Song F, Gao T, Liao M, Yan J, Wu J (2017) Forensic characteristics and phylogenetic analyses of the Chinese Yi population via 19 X-chromosomal STR loci. Int J Legal Med 131:1243–1246. https://doi.org/10.1007/s00414-017-1563-0
- He G, Li Y, Zou X, Wang M, Chen P, Liao M, Wu J (2017) Genetic polymorphisms for 19 X-STR loci of Sichuan Han ethnicity and its comparison with Chinese populations. Leg Med (Tokyo) 29:6–12. https://doi.org/10.1016/j.legalmed.2017.09.001
- He G, Li Y, Zou X, Zhang Y, Li H, Wang M, Wu J (2018) Xchromosomal STR-based genetic structure of Sichuan Tibetan minority ethnicity group and its relationships to various groups. Int J Legal Med 132:409–413. https://doi.org/10.1007/s00414-017-1672-9
- Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 16:1215
- Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567. https://doi.org/10. 1111/j.1755-0998.2010.02847.x
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054