Forensic Sciences Research, 2024, Vol. 9, Issue 2
https://doi.org/10.1093/fsr/owad058
Advance access publication date 5 January 2024
Research Article

Forensic efficiency and population genetic construction of Guizhou Gelao minority from Southwest China revealed by a panel of 23 autosomal STR loci

Siyu Chai[3,4,†], Shuhua Li[1,2,4,†], Ruxin Zhu[3,†], Li Luo[4,5], Kaiqin Chen[1,2,3], Yinlei Lei[1,2,3], Weihong Wan[1,2], Xijie Hu[1,2], Shiquan Liu[6,*], Pengyu Chen[1,2,3,4,*]

[1]Key Laboratory of Cell Engineering of Guizhou Province, Clinical Stem Cell Research Institute, Affiliated Hospital of Zunyi Medical University, Guizhou, China
[2]Collaborative Innovation Center of Chinese Ministry of Education, Zunyi Medical University, Guizhou, China
[3]Shanghai Key Lab of Forensic Medicine, Key Lab of Forensic Science, Ministry of Justice, China (Academy of Forensic Science), Shanghai, China
[4]Center of Forensic Expertise, Affiliated hospital of Zunyi Medical University, Guizhou, China
[5]School of forensic medicine, Shanxi Medical University, Taiyuan, China
[6]Institute of Evidence Law and Forensic Science, China University of Political Science and Law, Beijing, China

*Corresponding authors: shiquan.liu@cupl.edu.cn; pychenfs@163.com

[†]Siyu Chai, Shuhua Li and Ruxin Zhu contributed equally to this work.

Abstract

Short tandem repeats (STRs) are the most common genetic markers in forensic and human population genetics due to their high polymorphism, rapid detection, and reliable genotyping. To adapt the rapid growth of forensic DNA database and solve problems in disputed cases, a panel of 23 autosomal STR loci with high discriminating ability was constructed recently. The Tai-Kadai-speaking Gelao is the most ancient indigenous minority in Guizhou province, however, the forensic efficiency and population genetic structure remain poorly explored. Here, 490 Guizhou Gelao individuals from Southwest China were genotyped with the panel of 23 STRs using the Huaxia Platinum Kit. A total of 265 alleles were screened. The combined discrimination power and the combined probability of paternity were 0.9999 and 0.9999, respectively. This indicated the 23 loci had higher discrimination power in Guizhou Gelao and could be applied to forensic practice. Comprehensive population structures with reference populations from China and abroad using the neighbour-joining phylogenetic tree (N-J tree), multidimensional scaling, principal component analysis and heatmap demonstrated that Guizhou Gelao was genetically closer to Guizhou Han than other populations. Moreover, our results showed that a complex phylogenetic model was influenced by ethnic, geographic, and linguistic factors.

Key points

- The first batch of genetic data for 23 autosomal STRs in 490 Geolao individuals from Guizhou was provided.
- The 23 STR panel can afford high genetic polymorphisms and discrimination power and can be efficiently applied to forensic practice in Guizhou Gelao population.
- A complex phylogenetic model influenced by ethnic, geographic, and linguistic factors was uncovered.

Keywords: short tandem repeats; Guizhou Gelao; genetic polymorphisms; population structure

Introduction

Short tandem repeats (STRs), the repetitive nucleotide sequence of the human genome are composed of 2–6 base pairs, have become the most common genetic tool in forensic and human population genetics [1]. Due to their high polymorphism, rapid detection, and reliable typing results, STR markers play an indispensable role in routine forensic casework, especially in forensic examinations for individual identification and paternity testing [2, 3]. Forensic DNA databases, gathering vast individual STR profiling data from criminal cases, play a key role in the criminal justice system for forensic investigation. However, the rapid growth of DNA data in forensic DNA databases worldwide significantly increases the risk of accidental matching of unrelated individuals [4]. In addition, conventional STR multiple systems often lack sufficient discriminating power for disputed parentage testing, such as the duo parentage analysis, which may lack sampling from the father or mother, or mutation cases with one or two mismatches between parent and offspring.

To improve the identification ability of multiple STR detecting systems, the Ministry of Public Security of China has expanded the core loci of the national database of China from 13 to 20 [5]. Accordingly, many new panels with more STR loci have been developed. The Huaxia Platinum Kit consists of 23 autosomal STR markers as well as two

Received: August 13, 2023. **Accepted:** November 16, 2023

sex-determining loci, which comprises the latest core loci to satisfy the application needs of rapid growth of China's forensic databases and has shown higher discriminating power than traditional kits [6]. Presently, this kit is widely used in genetic research of Chinese populations but has been scarcely used in Guizhou Gelao, hampering its forensic applications.

Guizhou, located in the hinterland of Southwest China, is an ethnically diverse province comprising 56 ethnic groups, including 17 indigenous minorities and the Han ethnicity, which makes up most of the population. Its linguistic landscape belongs to several major language families, for instance Tai-Kadai, Hmong-Mien, Sinitic, Turkic, Mongolian, and Sino-Tibetan. Thus, due to the diversified nationalities, Guizhou is considered an important and hotspot region for forensic features and population genetic exploration. The Gelao is one of the most ancient indigenous minorities mainly residing in Guizhou, except for a few of them distributed in Yunnan, Guangxi, and Sichuan in China and the Hajiang Province in northern Vietnam. According to the China Statistical Yearbook-2021 (http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm), the population of Chinese Gelao is ~677 521, of whom over 97% are in Guizhou Province. Their unique Gelao language is subordinate to the Tai-Kadai language branch of the Sino-Tibetan language family. Scholars agree that the Gelao group originated from the ancient Liao people who lived in Southwest China during the Tang and Han dynasties and have experienced frequent and complicated historical migration, genetic exchange, and integration with the neighbouring Han ethnicity [7, 8].

Currently, population genetic research on Gelao people covers a variety of genetic markers, such as autosomal Indels [9], X-STRs Indels [10], Y-STRs Indels [11], mitochondrial DNA [12], and autosomal STRs [13, 14]. However, the marker sets used in previous studies on autosomal STRs were relatively small, and the data of 15 STR loci [13, 14] are insufficient for updating the current database and forensic use. Thus, this work aimed to acquire the genetic diversity of the Guizhou Gelao population using a 23 STR panel. Based on the highly polymorphic STR genotyping data, we intended to further assess the forensic efficiency of the panel in Guizhou Gelao people and reconstruct the genetic relationships based on Chinese and overseas populations.

Materials and methods

Sample preparation and DNA extraction

Peripheral blood samples stored in FTA cards from 490 unrelated healthy Gelao individuals (235 male and 255 female) residing in the Zunyi region, Guizhou province, Southwest China, were collected after obtaining the participants' written informed consent. All participants were self-identified as indigenous Gelaos without immigration and interracial marriages for three or more generations. Genomic DNA was extracted by QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) following the manufacturer's guidelines. Then, DNA was quantified by NanoDrop 2000c (Thermo Fisher Scientific, Waltham, MA, USA) and normalized to a final concentration of 1 ng/μL. This study was approved by the Ethics Committee of the Affiliated Hospital of Zunyi Medical University (KLLY-2021-110).

Multiplex amplification and STR genotyping

PCR amplification of 23 STRs (D16S539, D8S1179, D3S1358, vWA, TPOX, CSF1PO, D21S11, D18S51, Penta E, D2S441, TH01, D19S433, D5S818, FGA, D22S1045, D7S820, D13S317, Penta D, D10S1248, D12S391, D1S1656, D2S1338, and D6S1043) was performed on the ProFlex 96-well PCR System (Thermo Fisher Scientific) by the Huaxia Platinum Kit (Thermo Fisher Scientific) following the manufacturer's protocol. A total of 25 μL of PCR reaction volume, which comprised 10 μL of the master mix, 10 μL of the primer set, 4 μL of deionized water, and 1 μL of DNA template was employed. We performed the thermal cycling as following conditions: an initial step for 1 min at 95°C, followed by 26 cycles at 94°C for 30 s, 59°C for 16 s, then 65°C for 29 s, and a final extension at 60°C for 5 min. The PCR products were separated by multi-capillary electrophoresis on the Applied Biosystems 3500XL Genetic Analyzer (Thermo Fisher Scientific). Allele nomenclature was conducted using GeneMapper ID-X software (Thermo Fisher Scientific) and the allelic ladder provided by the kit. Negative control (sdH_2O) and positive control (9947A DNA sample) were run in each Polymerase Chain Reaction (PCR) system.

Dataset composition

To delineate the genetic structure among the Guizhou Gelao and other populations at the level of China and worldwide, two sets of reference population genetic data comprising the same 23 STR loci were retrieved. The first dataset consisted of 13 296 samples from 19 previously reported Chinese populations, including 10 Sinitic-speaking populations, four Tibeto-Burman-speaking populations, four Turkic-speaking populations, and one Hmong-Mien-speaking population. The second reference dataset contained a total of 22 889 samples from 34 groups around the world, including 25 East Asian populations speaking Tibeto-Burman, Tai-Kadai, Sinitic, Korean, Turkic, Mongolic, and Hmong-Mien languages, one South Asian group speaking Indo-Aryan language, two Southeast Asia group speaking Austronesian language, one European population speaking the Germanic language, one South America population speaking the Romance language, and four populations from North America speaking Germanic language [15–41]. The detailed information of the reference datasets are shown in Supplementary Table S1.

Statistical analysis

The expected heterozygosity (He), observed heterozygosity (Ho), the linkage disequilibrium (LD), and the probability values of Hardy–Weinberg equilibrium (HWE) were measured by Arlequin v3.5.2.2 [42]. Allele frequencies and other forensic parameters such as matching probability (PM), polymorphism information content (PIC), typical paternity index (TPI), power of exclusion (PE), and power of discrimination (PD) were evaluated by STRAF [43]. The combined probability of paternity (CPE) and combined discrimination power (CDP) were computed as $CPE = 1-(1-PE1)\ (1-PE2)\ (1-PE3)\ldots(1-PEk)$ and $CDP = 1-(1-DP1)\ (1-DP2)\ (1-DP3)\ldots(1-DPk)$, where k represents the number of loci.

A comprehensive population comparison among Guizhou Gelao and other reference populations was conducted based on the allele frequency distribution dataset of the 23 STR loci. The pairwise Nei's genetic distance was assessed using Phylip v3.69 [44]. Based on the Nei's genetic distance, the neighbour-joining phylogenetic tree (N-J tree) was built with the MEGA 7.0 software [45], principal component analysis (PCA) and multidimensional scaling (MDS) were performed using IBM

SPSS Statistics v25.0. Moreover, a heatmap was built using the online tool Omicshare (https://www.omicshare.com/tools/Home/Soft/heatmap).

Results

Allele frequency and forensic parameters of 23 STRs

The genotype data of 23 autosomal STR loci in the Huaxia Platinum Kit of the 490 unrelated Guizhou Gelao individuals were presented. The allelic frequencies and related forensic parameters of the Gelao population are listed in Supplementary Table S2. All 23 STR loci were in HWE after applying Bonferroni correction ($\alpha = 0.05/23$) (Supplementary Table S2). No deviations from the LD were found after the Bonferroni correction (0.05/253) (Supplementary Table S3). A total of 264 alleles were identified with allelic frequencies ranging from 0.001 (26, Penta D) to 0.5255 (8, TPOX). Penta E detected a maximum of 21 alleles, whereas TPOX only detected 5 alleles. The Ho and He values ranged from 0.6020 (TPOX) to 0.9122 (Penta E) and 0.6229 (TPOX) to 0.9195 (Penta E), respectively. The values of TPI ranged from 1.2564 (TPOX) to 5.6977 (Penta E), that of PIC from 0.5627 (TPOX) to 0.9129 (Penta E) while that of PD and PE ranged from 0.7939 (TPOX) to 0.9862 (Penta E) and 0.2933 (TPOX) to 0.8205 (Penta E), respectively. The CDP and CPE were 0.999 999 999 999 999 999 999 999 999 47 and 0.999 999 999 70, respectively.

Comparisons with the Chinese populations

We first constructed a reference dataset comprising 23 autosomal loci to compare the Gelao and Chinese populations. A total of 17 564 samples from 20 groups were included. The Nei's genetic distances are shown in Supplementary Table S4. The closest genetic distance was observed between Guizhou Geolao and the geographically neighbouring Guizhou Han (0.0029), followed by Guangdong Han (0.0055), while the maximal distance observed was with the Xinjiang population at the China Western border (0.0602).

Based on Nei's genetic distance, the population relationships between Gelao and the other 19 Chinese groups were constructed and visualized using the comprehensive means of the N-J tree, MDS plots, PCA, and heatmap. As shown in Figure 1, three major genetic cluster branches, namely, the Sinitic-speaking, Tibeto-Burman-speaking, and Turkic-speaking clusters, could be clearly identified, distinguishing them from each other in the phylogenetic relationship tree. Thus, the investigated Guizhou Gelao population was first clustered with Guizhou Han and Guangdong Han and next converged with the other Sinitic-speaking populations, including the Han populations from different provinces and the Hui minority group.

The MDS plot is shown in Figure 2. Three Tibeto-Burman-speaking populations were congregated in the fourth quadrant, except the Hubei Tujia, which was in the first quadrant. Four populations who speak Turkic language were on the second and third quadrants. Guizhou Gelao and Sinitic-speaking populations showed strong genetic affinity and clustered in the first and second quadrants. Moreover, Guizhou Miao was located in the first quadrant and was relatively far away from Guizhou Gelao compared with other Han groups.

Our PCA results showed that 90.8% of the genetic variation was extracted from the first three components

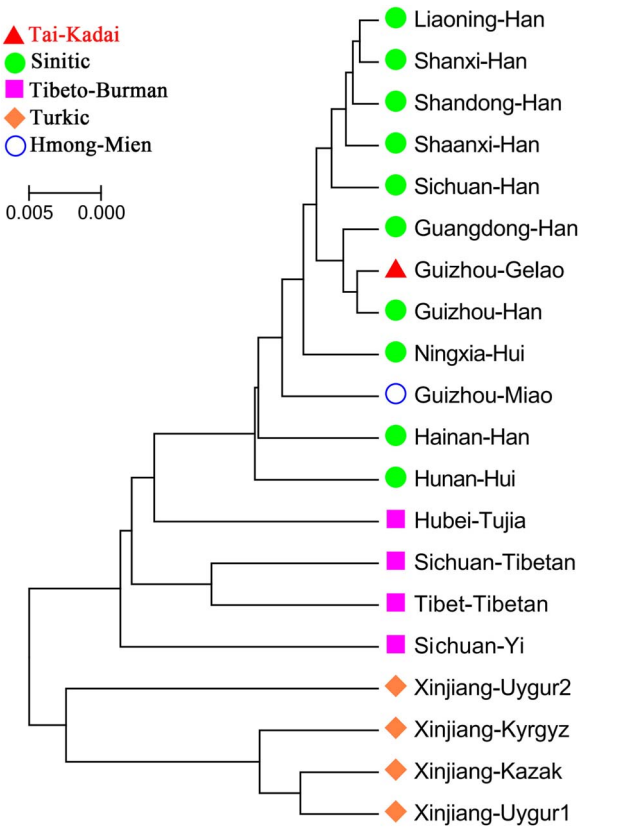

Figure 1 Neighbour-joining phylogenetic tree among 20 Chinese populations base on Nei's genetic distances.

(PC1:67.656%; PC2:14.374%; PC3:8.459%) (Figure 3). PC1 gathered the Guizhou Miao, Guizhou Gelao, and Sinitic-speaking populations. PC2 could differentiate the Turkic-speaking populations except for one Xinjiang Uyghur population. Comparatively, PC3 could distinguish the Sichuan Yi, Sichuan Tibetan, and Tibet Tibetan populations from others.

The genetic relationship revealed by heatmap analysis (Figure 4) concorded with the results of the MDS and N-J tree. The Guizhou Gelao was grouped with Sinitic-speaking populations. Geographical clustering was also observed. For example, all Xinjiang populations clustered together but were more distant from other populations.

Comparisons with populations across the world

To further illustrate the genetic heterozygosity and homogeneity among the Guizhou Gelao and other populations, group comparisons were conducted using a dataset of 34 populations around the world with 23 autosomal loci. The Nei's genetic distance results showed that the closest genetic distance was between Guizhou Geolao and Guizhou Han (0.0029), followed by two Guangdong Han groups (0.0051, 0.0055), while Guizhou Gelao displayed the farthest genetic distance with African American (0.1510) (Supplementary Table S5). The N-J tree of 35 populations is shown in Figure 5. Two main branches were clustered in the dendrogram, the Asian Branch (except India) and the European and American Branch (except Asian American). Guizhou Gelao was first clustered with Guizhou Han, then formed a clade with other Han populations and the Guizhou Miao population.

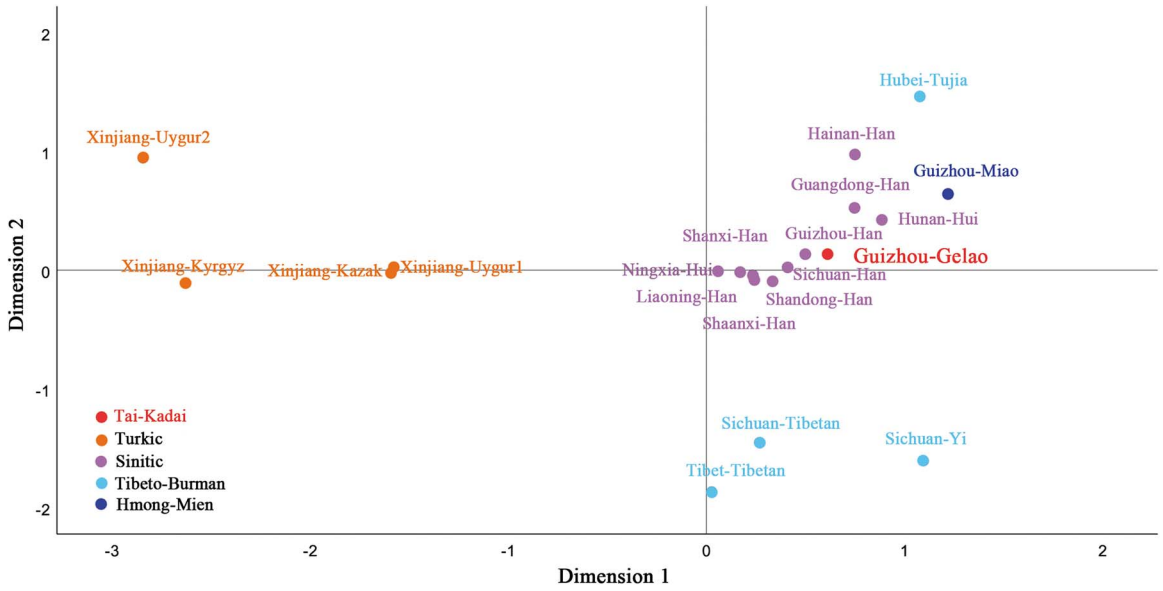

Figure 2 Multidimensional scaling plot among 20 Chinese populations base on Nei's genetic distances. The two-dimensional plot of (A) PCA1 and PCA2, (B) PCA1 and PCA3, and (C) PCA2 and PCA3. (D) The three-dimensional plot of principal component analysis results.

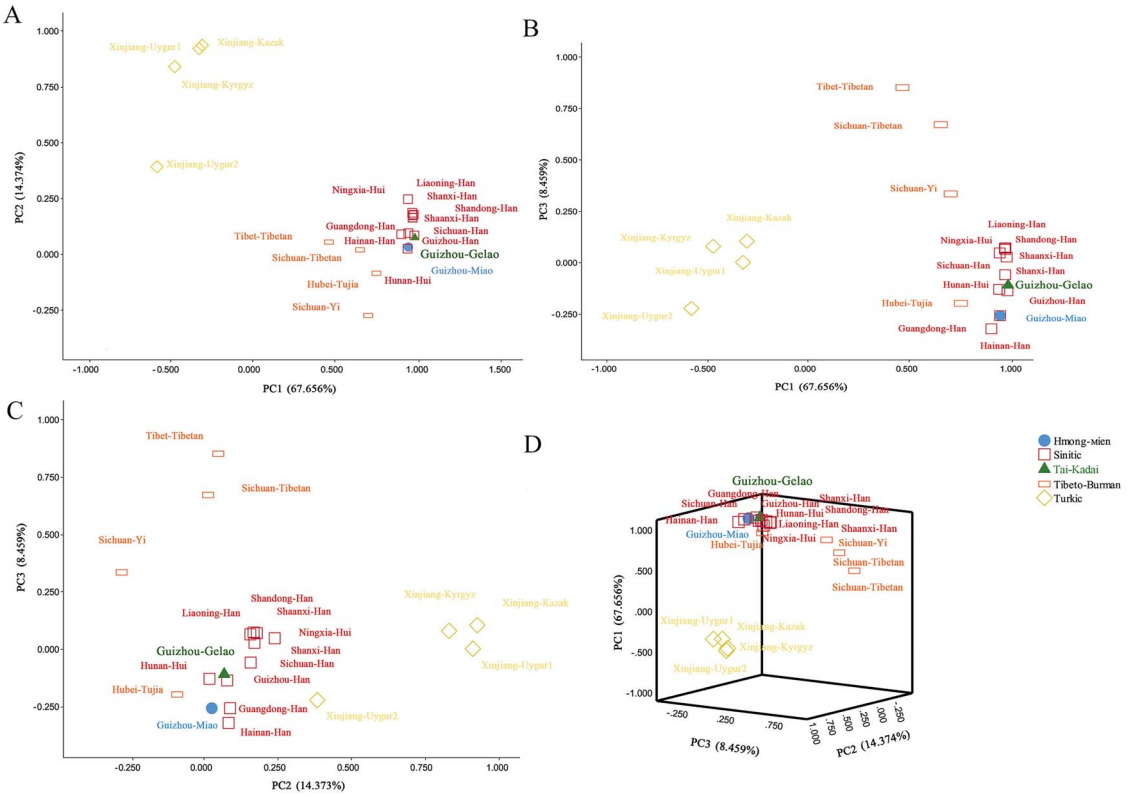

Figure 3 Principal component analysis among 20 Chinese populations base on Nei's genetic distances. The two-dimensional plot of (A) PCA1 and PCA2, (B) PCA1 and PCA3, and (C) PCA2 and PCA3. (D) The three-dimensional plot of principal component analysis results.

The MDS plot is shown in Supplementary Figure S1. The results showed that Asian groups were concentrated in the first and fourth quadrants, except India in the third quadrant. Among these Asian populations, Guizhou Gelao and Sinitic-speaking populations showed strong aggregation. Besides, the Turkic-speaking populations in Northwest China were between Europeans and East Asians.

The PCA results indicated that the first two components (PC1: 78.289%; PC2: 13.416%) could extract 91.705% variations from the total variations of 35 world populations (Supplementary Figure S2). From the PCA plots, we observed that most Chinese groups, two Southeast Asian groups and Asian American gathered at the bottom right, Turkic-speaking populations were scattered on the upper right, and the American

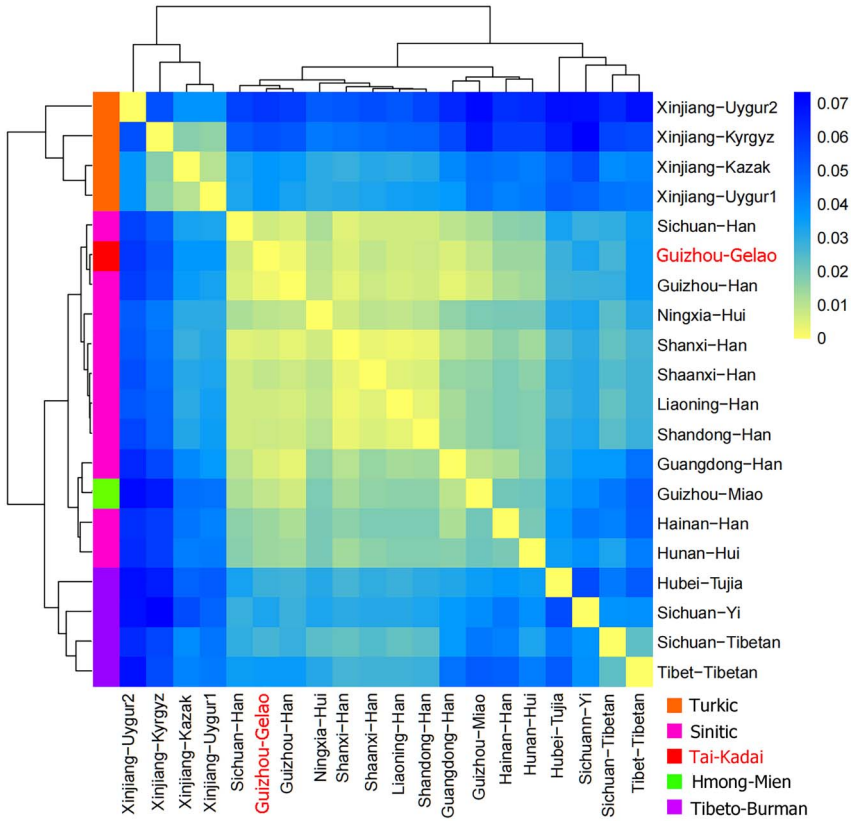

Figure 4 Heatmap among 20 Chinese populations base on Nei's genetic distances.

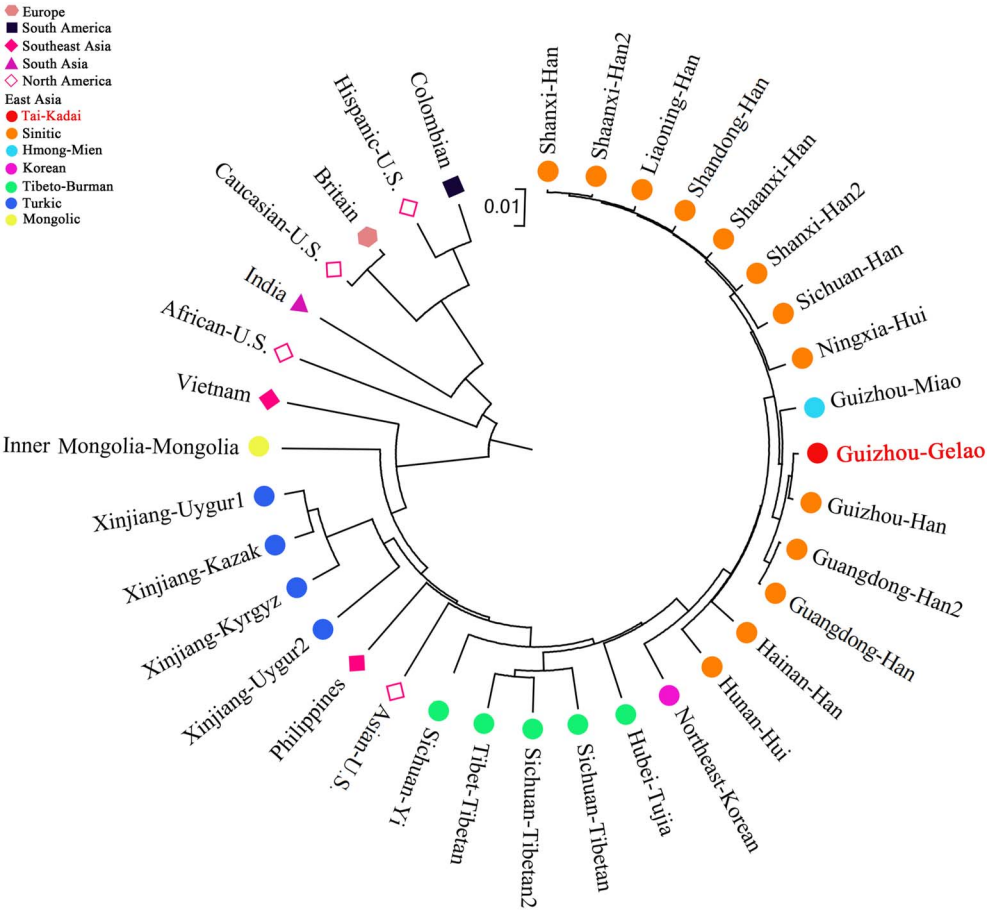

Figure 5 Neighbour-joining phylogenetic tree among 35 world populations base on Nei's distances.

and European groups were distributed on the left side of the PCA plot.

Heatmap analysis indicated that the African American had the furthest genetic distance from the Guizhou Gelao, while some southern Chinese groups, such as Guizhou Han and Guangdong Han, showed a close genetic relationship with Guizhou Gelao (Supplementary Figure S3). Besides, we found that the Sinitic-speaking, Tibeto-Burman-speaking, and Turkic-speaking populations demonstrated obvious linguistic aggregation. Asian groups, including Chinese groups, Vietnam, and the Philippines, aggregated as a large group, showing significant geographic agglomeration.

Discussion

China, populated by over 1.3 billion people, consists of 56 officially recognized ethnicities. Because of China's large population size, diversified geographic distributed, complex ethic origination and admixture, it is of great significance for human genetics to study the genetic characteristics of different Chinese populations. Guizhou, a mountainous province in Southwest China and geographically isolated from other places, harbours the most ethnically and linguistically diversified minority groups. Recently, an increasing number of genetic studies in Guizhou populations were performed by forensic and human genetic researchers using a variety of genetic markers such as STR, single nucleotide polymorphisms (SNPs), InDels, and mitochondrial DNA. Newly developed STR multiple detection kits, which include the 20 updated core loci of the Ministry of Public Security of China, can improve discriminating ability and have been widely applied in forensic paternity testing, personal identification, and the construction of DNA databases. Although the forensic characteristics and genetic relations of the Guizhou Gelao population were also widely investigated using several markers, they were not explored using highly discriminating STR marker sets, such as the 23 STR panel.

In this study, we provide the first genetic data using a 23 STR panel based on 490 Gelao individuals from the Guizhou province of Southwest China. Compared with the previous report on 15 STRs on the Guizhou Gelao people [13], except for the additional eight loci, consistent results were observed in POX and D18S51, which was the lowest and highest polymorphic locus, respectively. Moreover, based on the additional eight STR, Pent E and D6S1043 displayed higher discrimination ability than D18S51. The comprehensive analysis of allele frequencies and forensic parameters proved that the 23 STR panel had much higher genetic polymorphism and discrimination power than the 15 STR set, proving that the kit could be efficiently used for forensic investigations on the Guizhou Gelao population, especially for individual identification, paternity testing, and DNA database construction.

Moreover, population genetic construction of the Gelao and other populations was performed at both domestic and international levels with two reference population datasets based on genetic data from the same 23 STRs. A series of visual analyses, such as the N-J tree, MDS, heatmap, and PCA, based on Nei's genetic distance demonstrated a complex phylogenetic model influenced by ethnic, geographic, and linguistic factors.

Comparisons between Guizhou Gelao and Chinese populations indicated that the Guizhou Gelao people were closely related to the Guizhou populations and the Chinese Han populations. This result was also concordant with findings of other genetic markers, for instance autosomal Indel and X-STR [9, 10]. The genetic maps obtained by different analytic methods confirmed that Guizhou Gelao has a strong genetic affinity with Guizhou Han but was far away from populations speaking Turkic and Tibeto-Burman languages. The N-J tree showed that groups speaking Sinitic, Turkic, and Tibeto-Burman languages were closely clustered, respectively, and significant genetic homogenization was found in populations belonging to the same language families.

Additionally, a closer genetic affinity between the Guizhou Gelao who speak Tai-Kadai language and the Guizhou Miao who speak Hmong-Mien language was demonstrated than the Turkic and Tibeto-Burman speakers. As shown in the MDS plot, the two Guizhou ethnic groups shared the same first quadrant. Whereas compared with the Sinitic-speaking populations, the genetic distance between Guizhou Miao and Guizhou Gelao was relatively far, which was similar as reported in previous studies and might be attributed to geographical location and ethnic origin [9, 46].

Based on the analysis between the Guizhou Gelao and different populations across the world, similar results were also obtained, Guizhou Gelao and populations speaking the Sinitic language gather together. Additionally, except for India, populations across Asian regions were geographically clustered, and populations from the same language family also showed genetic affinities. These observed genetic affinity reflect the common pattern that populations with geographic proximity or linguistic affinity share more ancestral information [47]. Besides, the genetic map also showed that the Turkic-speaking populations were between Europeans and East Asians and are considered descendants of ancient mixed populations of East Asians and Europeans. Due to geographical isolation, group migration, customs, and many other factors, the genetic relationship between ethnic groups is more complex. Further elucidation of the origin and migration of the Gelao group requires large scale studies on populations in different nations and regions with higher density genetic marker sets.

Conclusion

In conclusion, the Huaxia Platinum PCR amplification kit demonstrated high genetic polymorphisms and discrimination power and could be efficiently used for forensic applications on the Guizhou Gelao population. The comprehensive genetic relationships at the domestic and worldwide levels using various methods demonstrated a complex phylogenetic model influenced by ethnic, geographic, and linguistic factors. The Guizhou Gelao is genetically closer to Sinitic-speaking populations, especially the geographically neighbouring Guizhou Han. In addition, a closer genetic affinity was observed between the Tai-Kadai-speaking Guizhou Gelao and the Hmong-Mien-speaking Guizhou Miao than the Tibeto-Burman and Turkic speakers.

Authors' contributions

Pengyu Chen contributed to the study's concept and design. Material preparation was conducted by Siyu Chai and Shuhua Li. Data collection and analysis were conducted by Siyu Chai, Shuhua Li, Ruxin Zhu, Li Luo, Kaiqin Chen, Yinlei Lei, Weihong Wan and Xijie Hu.The first draft of the manuscript was written by Shiquan Liu, Siyu Chai, Shuhua Li and Ruxin Zhu. Pengyu Chen and Shiquan Liu revised it critically for

important intellectual content. All authors have read and approved the final manuscript.

Compliance with ethical standards

All volunteers were adequately informed and signed written informed consent before sample collection. This study was approved by the Ethics Committee of the Affiliated Hospital of Zunyi Medical University (KLLY-2021-110). The procedures used in this study adhered to the Declaration of Helsinki and later amendments.

Disclosure statement

None declared.

Funding

This work was supported by Shanghai Key Lab of Forensic Medicine, Key Lab of Forensic Science, Ministry of Justice, China (Academy of Forensic Science), Open Project (No. KF202109); National Natural Science Foundation (No. 81401562); Science and Technology Project of Guizhou Provincial Health Commission (No. gzwkj2021-533); Science and Technology Cooperation Project of Zunyi Science and Technology and big data Bureau (No. HZ-2022-237); Science and Technology Foundation of Guizhou Province of China (No. QKHJC-ZK[2023]582).

References

1. Gymrek, Willems T, Reich D, et al. Interpreting short tandem repeat variations in humans using mutational constraint. Nat Genet. 2017;49:1495–1501.
2. Butler JM, Hill CR. Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. Forensic Sci Rev. 2012;24: 15–26.
3. Butler JM. Genetics and genomics of core short tandem repeat loci used in human identity testing. J Forensic Sci. 2006;51:253–265.
4. Ge J, Sun H, Li H, et al. Future directions of forensic DNA databases. Croat Med J. 2014;55:163–166.
5. Wang Z, Zhou D, Jia Z, et al. Developmental validation of the Huaxia Platinum System and application in 3 main ethnic groups of China. Sci Rep. 2016;6:31075.
6. Li W, Wang X, Wang X, et al. Forensic characteristics and phylogenetic analyses of one branch of Tai-Kadai language-speaking Hainan Hlai (Ha Hlai) *via* 23 autosomal STRs included in the Huaxia™ Platinum System. Mol Genet Genomic Med. 2020;8:e1462.
7. Shi H, Dong YL, Wen B, et al. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. Am J Hum Genet. 2005;77:408–419.
8. Gou S, Cang M. [Between "Diversity" and "Integration": study on the changes and transformation of Gelao Ethnic Group in Guizhou]. Guizhou Ethn Stud. 2020;41:75–83. Chinese.
9. He G, Wang Z, Zou X, et al. Tai-Kadai-speaking Gelao population: forensic features, genetic diversity and population structure. Forensic Sci Int Genet. 2019;40:e231–e239.
10. Chen P, He G, Zou X, et al. Genetic structure and polymorphisms of Gelao ethnicity residing in Southwest China revealed by X-chromosomal genetic markers. Sci Rep. 2018;8:14585.
11. Wang X, Jiang L, Qian E, et al. Genetic polymorphisms and haplotypic structure analysis of the Guizhou Gelao ethnic group based on 35 Y-STR loci. Leg Med (Tokyo). 2020;43:101666.
12. Liu C, Wang SY, Zhao M, et al. Mitochondrial DNA polymorphisms in Gelao ethnic group residing in Southwest China. Forensic Sci Int Genet. 2011;5:e4–e10.
13. Sun H, Xu S, Long F, et al. Forensic and population genetic analysis of Han, Miao, Tujia and Gelao populations from Zunyi (Southwest China) on 15 autosomal short tandem repeat loci. Forensic Sci Int Genet. 2016;25:e20–e21.
14. Yang L, Zhao Y, Liu C, et al. Allele frequencies of 15 STRs in five ethnic groups (Han, Gelao, Jing, Shui and Zhuang) in South China. Forensic Sci Int Genet. 2013;7:e9–e14.
15. Zhang H, Hu X, Liu Z, et al. [Study on OL alleles of 23 STR loci in Guizhou Han population]. J Zunyi Med Univ. 2020;43:686–691. Chinese.
16. Wang M, Wang Z, He G, et al. Genetic characteristics and phylogenetic analysis of three Chinese ethnic groups using the Huaxia Platinum System. Sci Rep. 2018;8:2429.
17. Shen H, Guo F, Jin P, et al. [Genetic polymorphism of 23 STR loci in Liaoning Han population]. Chinese J Forensic Med. 2014;29: 259–261. Chinese.
18. Sheng X, Wang Y, Zhang J, et al. Forensic investigation of 23 autosomal STRs and application in Han and Mongolia ethnic groups. Forensic Sci Res. 2018;3:138–144.
19. Liu J, Wang Z, He G, et al. Genetic polymorphism and phylogenetic differentiation of the Huaxia Platinum System in three Chinese minority ethnicities. Sci Rep. 2019;9:3371.
20. He G, Wang M, Liu J, et al. Forensic features and phylogenetic analyses of Sichuan Han population *via* 23 autosomal STR loci included in the Huaxia Platinum System. Int J Leg Med. 2018;132: 1079–1082.
21. Liu Y, Li X, Guo L, et al. [Genetic polymorphism of 23 autosomal STR loci in Xinjiang Kazak population and its genetic correlation with other nationalities]. Basic Clin Med. 2019;39:157–164. Chinese.
22. Jin X, Wei Y, Chen J, et al. Phylogenic analysis and forensic genetic characterization of Chinese Uyghur group *via* autosomal multi STR markers. Oncotarget. 2017;8:73837–73845.
23. Liu L, Song H, Jiang X. [Genetic polymorphism of 23 STR loci of Korean nationality in Northeast China]. Chinese J Forensic Med. 2015;30:612–614. Chinese.
24. Liu Y, Yue J, Li J, et al. [Genetic polymorphism of 24 autosomal short tandem repeat loci in Tujia population in Hubei]. Lab Med Clin. 2020;17:1800–1804+1810. Chinese.
25. Chen P, Wu J, Luo L, et al. Population genetic analysis of modern and ancient DNA variations yields new insights into the formation, genetic structure, and phylogenetic relationship of northern Han Chinese. Front Genet. 2019;10:1045.
26. Chen P, Zou X, Wang B, et al. Genetic admixture history and forensic characteristics of Turkic-speaking Kyrgyz population *via* 23 autosomal STRs. Ann Hum Biol. 2019;46:498–501.
27. Yang J, Hu R. [Forensic genetic polymorphism of Goldeneye 25A identification system in Hui people]. Chinese J Forensic Sci. 2018; 72–75. Chinese.
28. Li L, Zou X, Zhang G, et al. Population genetic analysis of Shaanxi male Han Chinese population reveals genetic differentiation and homogenization of East Asians. Mol Genet Genomic Med. 2020;8:e1209.
29. Rodriguez JJ, Salvador JM, Calacal GC, et al. Allele frequencies of 23 autosomal short tandem repeat loci in the Philippine population. Leg Med (Tokyo). 2015;17:295–297.
30. Qu N, Zhang X, Liang H, et al. Analysis of genetic polymorphisms and mutations at 23 autosomal STR loci in Guangdong Han population. Forensic Sci Int Genet. 2019;38:e16–e17.
31. Liu Y, Li J, Yue J, et al. [Genetic polymorphism and genetic relationship analysis of 24 autosomal STR loci in Gelao and Miao populations in Guizhou]. J PLA Med. 2019;44:682–689. Chinese.
32. Shrivastava P, Dixit S, Kumawat RK, et al. Efficiency analysis of VersaPlex™ 27PY system in central Indian population: first report from Indian population. Leg Med (Tokyo). 2022;54:101983.
33. Krzeminska-Ahmadzai U, Buckley B, Loake T, et al. Population data for 23 autosomal STR loci in White British population. Leg Med (Tokyo). 2021;50:101863.

34. Hill CR, Duewer DL, Kline MC, et al. U.S. population data for 29 autosomal STR loci. Forensic Sci Int Genet. 2013;7:e82–e83.
35. Castillo A, Pico A, Gil A, et al. Genetic variation of 23 STR loci in a Northeast Colombian population (Department of Santander). Forensic Sci Int Genet Suppl Ser. 2019;7:33–35.
36. Dung Pham P, Luc Hoang T, Tra Le K, et al. The first data of allele frequencies for 23 autosomal STRs in the Ede ethnic group in Vietnam. Leg Med (Tokyo). 2022;57:102072.
37. He G, Wang Z, Wang M, et al. Genetic variations and forensic characteristics of Han Chinese population residing in the Pearl River Delta revealed by 23 autosomal STRs. Mol Biol Rep. 2018;45: 1125–1133. https://doi.org/10.1007/s11033-018-4264-y.
38. Zhang J, Zhao Z, Li L, et al. [Forensic verification of Huaxia™ Platinum PCR amplification kit and investigation of STR genetic polymorphism]. Chinese J Forensic Sci. 2019;41–48. Chinese.
39. Ye Y, Liang Y, Luo H, et al. Genetic diversity of 23 autosomal STR loci in a Tibetan population. Forensic Sci Int Genet Suppl Ser. 2017;6:e101–e103.
40. Gu Z, He Y, Song Z, et al. [Genetic polymorphism of 23 STR loci in Weinan Han population in Shaanxi Province]. Chinese J Forensic Med. 2017;32:309–311. Chinese.
41. Chen Y, Wang X, Gao H, et al. [Study on genetic polymorphism of 23 STR loci in the Han population of Yuncheng, Shanxi]. Cent South Univ (Med Sci). 2018;46–54. Chinese.
42. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10:564–567.
43. Gouy A, Zieger M. STRAF—A convenient online tool for STR data evaluation in forensic genetics. Forensic Sci Int Genet. 2017;30: 148–151.
44. Retief JD. Phylogenetic analysis using PHYLIP. Methods Mol Biol. 2000;132:243–258.
45. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–1874.
46. Tang J, Yang M, Wang X, et al. Genetic structure and forensic characterisation of 36 Y-chromosomal STR loci in Hmong-Mien-speaking Miao population. Ann Hum Biol. 2020;47:541–548.
47. Haber M, Doumet-Serhal C, Scheib C, et al. Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences. Am J Hum Genet. 2017;101:274–282.