RESEARCH



Exploring the forensic effectiveness and population genetic differentiation in Guizhou Miao and Bouyei group by the selfconstructed panel of X chromosomal multiinsertion/deletions

Check for updates

Xiaolan Huang¹, Changyun Gu¹, Qianchong Ran¹, Li Chen¹, Shunyi Tian¹, Min Zhong¹, Zheng Ren¹, Qiyan Wang¹, Meiqing Yang¹, Jingyan Ji¹, Wen Wan¹, Jiang Huang^{2*}, Hongling Zhang^{1*} and Xiaoye Jin^{1*}

Abstract

In this research, a self-developed panel comprising 22 X chromosomal multi-InDels and one X-STR was used to explore the genetic polymorphisms and forensic characteristics of these loci in Guizhou Miao and Guizhou Bouyei populations. Besides, genetic affiliations among Guizhou Miao, Guizhou Bouyei and Guizhou Han populations were investigated using principal component analysis, STRUCTURE and machine learning methods. The findings indicated that these loci in the male and female samples had comprehensive discrimination powers greater than 0.999999999. Meanwhile, the cumulative mean exclusion chance of these 23 loci for trio and duo cases were also greater than 0.99999 in Guizhou Miao and Guizhou Bouyei populations. Population genetic analyses of three Guizhou populations revealed that there were relatively low genetic divergences among these populations based on the self-constructed panel. In conclusion, this system could be utilized as the valuable tool for forensic personal identification and parentage testing in Guizhou Miao and Guizhou Bouyei populations.

Keywords X chromosome, Multi-InDel, Guizhou Miao, Guizhou Bouyei, Forensics

*Correspondence: Jiang Huang mmm_hj@126.com Hongling Zhang 229598103@qq.com Xiaoye Jin 1115259825@qq.com ¹Department of Forensic Medicine, Guizhou Medical University, Guiyang, Guizhou 550025, China

²School of Public Health, the Key Laboratory of Environmental Pollution Monitoring and Disease Control, Ministry of Education, Guizhou Medical University, Guiyang, Guizhou 550025, China

Introduction

The X chromosome has a unique inheritance pattern: genetic markers for the female can transmit to son or daughter; whereas, genetic markers for the male can only pass to the daughter, which has become a useful tool for forensic practice [1, 2]. In addition, X chromosomal genetic markers are also more suitable for solving complex kinship analysis (like sisterhood when parents are missing) that cannot be solved with autosomal markers [3, 4].

In 2013, Kidd et al. introduced a new type of forensic genetic marker, named microhaplotype [5]. It consists



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article are provided in the article's Creative Commons licence, unless indicate otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, wisit http://creativecommons.org/licenses/by-nc-nd/4.0/.

of two or more single nucleotide polymorphisms (SNPs) within the 200–300 bp range of the human genome [6]. The genetic polymorphism of microhaplotypes was significantly higher than that of a single SNP locus, which showed high application values in forensic researches. For microhaplotypes, next generation sequencing (NGS) technology can directly determine cis/trans relationships of multiple SNPs in the short distance and distinguish each parental allele combination of SNPs at one microhaplotype locus. Nowadays, NGS is widely used as the optimal method for detecting and typing microhaplotypes [7]. However, NGS is commonly involved in high detecting costs and complex experimental operation (especially for data analysis), which was hard to be put into practice in forensic grassroots units. These shortcomings further limited forensic application of microhaplotypes.

Similar to SNPs, a single insertion/deletion polymorphism (InDel) commonly exhibits two allele variations in population and provides limited information for forensic researches. Given the idea of microhaplotype, Huang et al. proposed another novel genetic marker, multi-InDel [8]. Multi-InDels consist of two or more InDels (their distance is common less than 100 bp) and belong to the generalized microhaplotypes [8, 9]. Therefore, multi-InDels have some advantageous features similar to microhaplotypes; for example, multi-InDels possess higher genetic diversities than a single SNP/InDel locus in population. More critically, multi-InDels belong to length polymorphism, which can be compatible with the capillary electrophoresis platform of forensic DNA laboratories [10, 11]. Accordingly, forensic geneticists could integrate favorable features of X chromosomes and multi-InDels to develop a high-efficient panel for forensic genetics and population genetics.

China is a multi-ethnic country with 55 ethnic minorities [12]. Guizhou is an inland province in southwest China on the western edge of the Yunnan-Guizhou Plateau and one of the most diverse provinces in China [13]. Guizhou and Yunnan provinces were pivotal in establishing the initial transport and communication routes that linked China with South and Southeast Asia via the alleged Southern Silk Road [14]. At the time of China's seventh census in 2020, the national Bouyei population was about 3.57 million, while about 2.71 million Bouyei population live in Guizhou province. Meanwhile, the Miao population is about 11 million (more than 4.5 million in Guizhou) and lives mainly in southern China, including the provinces of Guizhou, Hunan, Yunnan, Guangxi, Zhejiang and Guangdong [15]. The Miao population and the Bouyei population belong to the Hmong-Mien and Tai-Kadai language families, respectively [16, 17]. However, most of genetic studies on Bouyei and Miao populations mainly focus on autosomal, Y chromosomal, X chromosomal STRs and mitochondrial DNA [12, 14, 18–24], the exploration of multi-InDel in both populations is underdeveloped.

In this study, we genotyped 201 Guizhou Bouyei (GZB) and 153 Guizhou Miao (GZM) individuals using a selfdeveloped new system that included 22 X chromosomal multi-InDels (XMI) and an X-STR (DXS7424) locus and analyzed the allele frequencies and forensic parameters of these loci in the GZM and GZB populations. In addition, we also used multiple methods to dissect genetic relationships and population stratifications of these two populations and Guizhou Han (GZH) population.

Materials and methods

Sample information

A total of 354 unrelated subjects from Guizhou provided blood samples, with the group consisting of 201 participants from GZB (113 males and 88 females) and 153 participants from GZM (74 males and 79 females). The participants, whose families had resided in Guizhou for a minimum of three generations, all gave written informed consent for this study. This research was performed according to the guideline of Guizhou Medical University Ethic Commission and approved by the Guizhou Medical University Ethic Commission (the approval number: 2021-218).

PCR amplification and data analysis

PCR amplification was executed on the 9700 thermocycle PCR system (Thermo Fisher Scientific). The PCR reaction consisted of 4 μ L PCR Master Mix, 3 μ L primers, 1 mm² blood card, and 3 μ L ddH₂O. The thermal cycling parameters were listed as below: an initial denaturation phase at 95°C lasting 2 min, 32 cycles of 94°C for 30 s, 58 °C for 1 min, 72 °C for 50 s; and terminal extension at 72°C for 10 min.

The PCR amplification products were isolated with the 3500xL Genetic Analyzer (Thermo Fisher Scientific). Each reaction consisted of 9.5 μ L of Hi-Di^{**} formamide (Thermo Fisher Scientific), 0.5 μ L AGCU Marker SIZ-500 and 1.0 μ L PCR product. The cocktail was firstly denatured at 95 °C for 3 min and followed by an ice bath for 3 min and detected by the capillary electrophoresis. In order to confirm allele profiles of each genetic marker, electrophoresis results were subjected to analysis with the GeneMapper ID-X software (Thermo Fisher Scientific).

Statistical analyses

Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) of one X-STR and 22 XMI loci were determined using the STRAF online tool [25] based on the allelic typing of these genetic markers in females. In addition, we also performed genetic differentiations of each genetic marker between males and females by the Arlequin software v3.5 [26]. Next, the allele frequencies and forensic parameters of these 23 loci in GZM and GZB populations were calculated by StatsX software [27].

Principal component analysis (PCA) of GZB, GZM, and GZH populations was conducted by the STRAF online tool at individual level. The analysis of the population genetic structure for GZB and GZM populations was conducted utilizing STRUCTURE v2.3.4 software [28], with the number of genetic clusters (K) ranging from 2 to 5. The simulation involved a burn-in period of 20,000 steps, followed by a Markov chain Monte Carlo run of 20,000 steps and each value of K was tested over 10 iterations. The CLUMPP v1.1.2 software was used to align the results of the STRUCTURE run and eliminate the impact of label switching and uncover potential multimodality within the data [29]. Finally, the Structure Harvester software [30] was utilized to calculate the optimum K value.

Based on allele typing of DXS7424 and 22 XMIs in GZB, GZM and GZH populations, four machine learning algorithms (decision tree, support vector machine, extreme gradient boosting and random forest) were employed to evaluate the performance of these genetic markers to differentiate these populations. Firstly, these individuals were randomly divided into training and testing datasets at the ratio of 9:1. Secondly, the training samples were used to build the prediction model and the testing samples were utilized to evaluate model performances. For these methods, decision tree, support vector machine (SVM), extreme gradient boosting (XGBoost), and random forest models were built by the rpart, e1071, mlr3verse and randomForest packages of R software, respectively. The caret package was utilized to tune up and determine the optimal parameters of these models.

Result

Allelic frequencies and forensic parameters of DXS7424 and 22 XMI loci in Guizhou Miao and Guizhou Bouyei populations

Firstly, we conducted HWE tests of these 23 loci in GZM and GZB populations, as listed in Supplementary Table 1. Following Bonferroni adjustment (the significant level=0.05/23), 19 loci (excluding XMI10, XMI19, XMI23, XMI26) were in HWE status in the GZM population and 18 loci (excluding XMI1, XMI10, XMI19, XMI23, XMI26) were in HWE status in the GZB population. Next, LD analyses of pairwise loci were performed in GZM and GZB populations, as listed in Supplementary Tables 2 and Supplementary Table 3. Obtained results revealed that there was no allelic association between any of the paired loci (the significant level=0.05/253), indicating that no linkage disequilibrium for these loci was observed in the two populations. Thirdly, we also evaluated gender differences of 22 XMI and DXS7424 loci in GZB and GZM populations, as shown in Supplementary Table 4. We found that these 23 loci didn't display gender differences in GZB and GZM populations. Therefore, allele frequencies of these 23 loci were estimated in all individuals, as shown in Fig. 1a and Supplementary Table 5. For these 23 loci, more than two alleles were detected, especially for XMI 26 (15 alleles in GZB and GZM). In addition, in the GZM population, allele frequencies of these 23 loci ranged from 0.0033 to 0.8987; in the GZB population, they distributed from 0.0025 to 0.9129.



Fig. 1 (a) Allele frequencies of 23 loci in GZM and GZB populations. (b). The cloud rain plot of forensic parameters of 23 loci in GZM and GZB populations

Forensic parameters of 23 loci in GZB and GZM populations were shown in Fig. 1b and Supplementary Table 6. Within the GZM group, the 23 loci displayed a range of expected heterozygosity (He) from 0.1866 to 0.8574, polymorphic information content (PIC) from 0.1758 to 0.8403, power of discrimination for male (PD-Male) from 0.1860 to 0.8546, power of discrimination for female (PD-Female) from 0.3272 to 0.9646, mean exclusion chance for deficiency cases (MEC-Kruger) from 0.0927 to 0.7171, mean exclusion chance for trios (MEC-Desmarais) from 0.1758 to 0.8403, and mean exclusion chance for duo cases (MEC-Desmarais duo) from 0.0983 to 0.7382. Within the GZB group, the 23 loci displayed a range of He from 0.1632 to 0.8902, PIC from 0.1564 to 0.8774, PD-Male from 0.1628 to 0.8880, PD-Female from 0.2927 to 0.9769, MEC-Kruger from 0.0829 to 0.7732, MEC-Desmarais from 0.1564 to 0.8774, and MEC-Desmarais duo from 0.0866 to 0.7904. Moreover, we observed that the XMI7 locus displayed the lowest He, PIC, PD-male, PD-female and MEC values in both populations, implying that the locus possessed relatively poor performance for forensic applications. Nevertheless, most loci in two Guizhou populations showed relatively high PIC (>0.5), PD (>0.5), and MEC (>0.3) values. The combined PD-Male, PD-Female, MEC-Kruger, MEC-Desmarais and MEC-Desmarais duos of 23 loci in GZM group were 0.999999999479, 0.9999999999999999, 0.99995, 0.99999999 and 0.99999246. The combined PD-Male, PD-Female, MEC-Kruger, MEC-Desmarais and MEC-Desmarais duo from 23 loci in the GZB group were 0.999999999, 0.9999999999999999, 0.999962, 0.999999991 and 0.999994672.

Population structure and genetic relationships explorations of Guizhou three populations

To elaborate the genetic background and relatedness among different populations in Guizhou, we performed a genotype-based PCA at the individual level in combination with the experimental data of the GZH population (Fig. 2a). In this figure, we could see that the GZB population almost completely overlapped with the GZH population, the GZM only partially overlapped with the former two populations, and some GZM individuals were scattered independently.

To proceed with evaluation of genetic components of the GZH, GZM and GZB populations, we also conducted genetic structure analyses of GZM, GZB and GZH populations and set the quantity of ancestral populations (K) to 2–5. The outcomes from the analysis of population genetic structure were depicted in Fig. 2b. At the same time, we uploaded the outcomes to online platform STRUCTURE HARVESTER which could help us to determine the optimal K value. As shown in Supplementary Fig. 1, the largest Delta K was observed at K=2, indicating that K=2 was the best value. At K=2, two principal ancestral components were identified from Fig. 2b. In addition, the ancestral composition of the GZB and GZH populations were similar. Nonetheless, the red represented ancestral component was more abundant in some GZM individuals at K=3. As the K increasing, no further genetic structure could be seen from these populations.

Performance evaluation of four machine learning methods for differentiating three Guizhou populations

We assessed the efficiency of the 23 loci in allocating individuals to their respective ethnic origins using four machine learning methods (decision tree, SVM, XGBoost and random forest). The confusion matrices of predicted and actual results in testing dataset were shown in Fig. 3. For these four models, we found that the random forest possessed the best performance with accuracy and Kappa values of 0.5510 and 0.2789; whereas, the decision tree displayed the worst performance with accuracy and Kappa values of 0.4286 and 0.1015. In addition, we found that most individuals from the GZB population could be assigned to the correct ethnic origins. However, these four methods possessed relatively poor performance for inferring ethnic origins of GZH and GZM groups.

Discussion

In this study, we evaluated forensic efficiency of a selfdeveloped panel comprising 22 XMI and DXS7424 loci in GZB and GZM populations. Furthermore, we also assessed genetic structure of these two populations based on these 23 loci. For these 23 loci, we discovered that four loci (XMI10, XMI19, XMI23, and XMI26) deviated from HWE in the GZM population and five loci (XMI1, XMI10, XMI19, XMI23 and XMI26) deviated from HWE in the GZB population. The reasons for deviation from HWE might be related to genetic drift, no-random mating, natural selection, and migration [14, 31].

The majority of the 23 loci in GZM and GZB populations had high He and PIC values, implying that these loci possessed comparatively high genetic polymorphisms in both groups. Besides, the cumulative PD for males and females of 23 loci were greater than 0.9999999999. At the same time, the cumulative MEC for trio and duo of 23 loci were also greater than 0.9999 in GZB and GZM populations. The results showed that the panel could provide enough genetic information and was suitable for individual identification and paternity testing of Miao and Bouyei populations in Guizhou. Furthermore, we found that the PIC value of the XMI7 locus was less than 0.2, which suggested that the locus had a relatively low level of polymorphism information content in the Guizhou populations. In order to improve system performance,



Fig. 2 (a) Principal component analysis of Guizhou three populations based on individual-level genotype data of 23 loci. (b) STRUCTURE analysis of Guizhou three populations from K=2 to 5

more X multi-InDel loci with higher polymorphism should be screened to replace it in subsequent studies.

Based on 23 loci on the X chromosome, population genetic analyses of three Guizhou populations were

carried out. According to the PCA results, the majority of individuals in the GZB and GZH populations were overlapped together, while there was some overlapping between GZM and the other two populations (GZH and



Page 6 of 8

Fig. 3 (a) The confusion matrix of predicted and actual results in testing samples based on the decision tree model; (b) the confusion matrix of predicted and actual results in testing samples based on the SVM model; (c) the confusion matrix of predicted and actual results in testing samples based on the XGBoost; (d) the confusion matrix of predicted and actual results in testing samples based on the string samples bas

GZB). This indicated that GZM, GZB and GZH were closely related, which was consistent with the results of genetic analyses of GZB and Guizhou other populations based on 19 X-STRs [14]. Nonetheless, a small percentage of the GZM individuals was not overlapped with GZB and GZH populations. Given that the samples we screened were from various Guizhou regions, we reasoned that the GZM population might have genetic substructures in different regions, which was also found in a previous study conducted by He et al. [32]. Their study showed that Miao populations in different areas of Guizhou were affected by gene flow related to Han and indigenous people in surrounding areas to different degrees. Meanwhile, our results from genetic structure analysis also showed genetic homogeneity of these three populations. We proposed that geographic context and gene flow of different populations could account for the genetic patterns of the GZM, GZB, and GZH populations. Specifically, geographically adjacent populations have more opportunities for gene flow, which gradually shapes genetic structure we observe from the genetic data. Similar results of population genetic analyses among Guizhou populations could be observed based on X-chromosomal and Y-chromosomal STRs [14, 33].

Recently, machine learning algorithms showed great potential in forensics and medical researches. Previous studies also explored the efficiency of machine learning methods for forensic ancestry analysis and found that the machine learning methods could be viewed as the highefficient tools for inferring forensic ancestry origins of unknown individuals, especially for those populations within the same continents or regions [34, 35]. In the current study, we also assessed the performance of decision tree, SVM, XGBoost and random forest methods for differentiating three Guizhou populations (GZH, GZB and GZM) based on 22 XMI and DXS7424 loci. Obtained results showed that these four machine learning algorithms exhibited relatively poor performance for forensic ancestry analyses of GZH, GZB and GZM populations, especially for GZH and GZM populations, which might be related to the following factors: (1) 23 loci that were used to build machine learning models were primarily used for forensic individual identification and paternity analysis; (2) relatively low genetic differentiation among GZH, GZB and GZM populations was also unfavorable to forensic ancestry analysis of these three populations.

To better differentiate these three populations, genome-wide data analysis for these populations should be performed to further select highly genetic differentiated markers for ethnic origin predictions of these populations. In addition, other artificial intelligence methods, like deep learning methods, could also be utilized to develop the prediction models for inferring ethnic origins of these populations given that deep learning possess better performance for feature selection and model construction than machine learning methods [36–38].

Conclusion

In the Hmong-Mein speaking GZM and Tai-Kadai speaking GZB populations, we firstly acquired the forensic reference genotype databases, forensic parameters, and allele frequencies of one X-STR and 22 XMI loci in the research. Based on the results of the forensic characteristics study, the self-developed panel has been demonstrated to be an effective tool for forensic identification and paternity analyses of GZM and GZB populations. Population genetic analyses of GZM, GZB, and GZH populations revealed that these three populations showed relatively low genetic differentiations. In order to achieve ethnic origin inferences of these three populations better, we should conduct genome-wide studies on these populations and further screen those loci with highly genetic divergences among these populations in the following research.

Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s12864-024-11088-2.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We are grateful to all volunteers for providing their blood samples.

Author contributions

X.H. write original draft. H.Z., X.J. and J.H designed the experiments and wrote review and editing. X.H., C.G., Q.R. and M.Z. were major contributors in conducting experiments. L.C. and S.T. collected samples. Z.R., Q.W. and M.Y.

performed data analysis. W.W. and J.J. prepared Figs. 1, 2 and 3. All authors reviewed and approved the submitted version of the manuscript.

Funding

The work was supported of by the Guizhou Provincial Science and Technology Projects (ZK [2024] General 162 and ZK [2022] General 355); Guizhou Innovation training program for college students (S202210660030, S202210660028 and S202310660094); National Natural Science Foundation [No. 82160324]; Guizhou Provincial Science and Technology Project ([2024] Young 240).

Data availability

Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of Guizhou Medical University, and the ethical license was issued by the Ethics Committee of Guizhou Medical University (approval number: 2021 - 218). All participants provided written informed consent to take blood samples. All experimental procedures were performed by the standards of the Declaration of Helsinki 1964.

Competing interests

The authors declare no competing interests.

Received: 18 September 2024 / Accepted: 25 November 2024 Published online: 09 December 2024

References

- Butler JM. Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019–2022. Forensic Sci Int Synerg. 2022;6:100311. https://d oi.org/10.1016/j.fsisyn.2022.100311
- Gomes I, et al. Twenty years later: a comprehensive review of the X chromosome use in forensic genetics. Front Genet. 2020;11:926.
- Garcia FM, et al. Forensic applications of markers present on the X chromosome. Genes (Basel). 2022;13(9):1597.
- Perera N, Galhena G, Ranawaka G. X-chromosomal STR based genetic polymorphisms and demographic history of Sri Lankan ethnicities and their relationship with global populations. Sci Rep. 2021;11(1):12748.
- Kidd KK, et al. Microhaplotype loci are a powerful new type of forensic marker. Forensic Sci International: Genet Supplement Ser. 2013;4(1):e123–4.
- Kidd KK. Proposed nomenclature for microhaplotypes. Hum Genomics. 2016;10(1):16.
- Oldoni F, Kidd KK, Podini D. Microhaplotypes in forensic genetics. Forensic Sci Int Genet. 2019;38:54–69.
- Huang J, et al. A novel method for the analysis of 20 multi-indel polymorphisms and its forensic application. Electrophoresis. 2014;35(4):487–93.
- 9. Jian H, et al. A novel SNP-STR system based on a Capillary Electrophoresis platform. Front Genet. 2021;12:636821.
- Qu S, et al. Multi-indel: a microhaplotype marker can be typed using Capillary Electrophoresis platforms. Front Genet. 2020;11:567082.
- 11. Liu J, et al. A new set of 20 Multi-InDel markers for forensic application. Electrophoresis. 2022;43(11):1193–202.
- Luo Y, et al. Population genetic analysis of 36 Y-chromosomal STRs yields comprehensive insights into the forensic features and phylogenetic relationship of Chinese Tai-Kadai-Speaking Bouyei. PLoS ONE. 2019;14(11):e0224601.
- Zhang H, et al. Forensic features and phylogenetic structure survey of four populations from southwest China via the autosomal insertion/deletion markers. Forensic Sci Res. 2024;9(2):owad052.
- Ren Z, et al. Forensic genetic polymorphisms and population structure of the Guizhou Bouyei people based on 19 X-STR loci. Ann Hum Biol. 2019;46(7–8):574–80.
- Fan GY, et al. Phylogenic analysis and forensic genetic characterization of Guizhou Miao tribes from 58 microareas via autosomal STR. Leg Med (Tokyo). 2020;47:101737.

- Duan S, et al. Malaria resistance-related biological adaptation and complex evolutionary footprints inferred from one integrative Tai-Kadai-related genomic resource. Heliyon. 2024;10(8):e29235.
- Wang Y, et al. The genomic history of southwestern Chinese populations demonstrated massive population migration and admixture among proto-Hmong-Mien speakers and incoming migrants. Mol Genet Genomics. 2022;297(1):241–62.
- Ren Z, et al. Population genetic data of 22 autosomal STRs in the Guizhou Miao population, southwestern China. Forensic Sci International: Genet. 2018;32:e7–8.
- Zhang L. Population data for 15 autosomal STR loci in the Bouyei ethnic minority from Guizhou Province, Southwest China. Volume 17. Forensic Science International: Genetics,; 2015. pp. 108–9.
- 20. Tran LH, et al. Genetic structure and population connection of two Bouyei populations in northern Vietnam based on short tandem repeat analysis. Am J Hum Biol. 2022;34(5):e23702.
- 21. Zhang X, et al. Forensic genetic polymorphisms of 16 X-STR loci in the Yunnan Miao population and their relationship to other Chinese groups. Leg Med (Tokyo). 2021;53:101961.
- Le C, et al. The mitochondrial DNA control region sequences from the Chinese Miao population of southeastern China. Ann Hum Biol. 2019;46(7–8):606–9.
- 23. Feng Y, et al. Analysis of maternal genetic structure of mitochondrial DNA control region from Tai-Kadai-speaking Buyei population in southwestern China. BMC Genomics. 2024;25(1):50.
- 24. Feng R, et al. Genetic analysis of 50 Y-STR loci in Dong, Miao, Tujia, and Yao populations from Hunan. Int J Legal Med. 2020;134(3):981–3.
- 25. Gouy A, Zieger M. STRAF-A convenient online tool for STR data evaluation in forensic genetics. Forensic Sci Int Genet. 2017;Early View:p148–151.
- 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(3):564–7.
- 27. Lang Y, Guo F, Niu Q. StatsX v2.0: the interactive graphical software for population statistics on X-STR. Int J Legal Med. 2018;133(1):39–44.
- Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003;164(4):1567–87.

- Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics. 2007;23(14):1801–6.
- Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour. 2012;4(2):359–61.
- 31. Chen L, et al. Development and validation of a forensic Multiplex System with 38 X-InDel loci. Front Genet. 2021;12:670482.
- 32. He G, et al. Genome-wide allele and haplotype-sharing patterns suggested one unique Hmong-Mein-related lineage and biological adaptation history in Southwest China. Hum Genomics. 2023;17(1):3.
- Chen P, et al. Genetic diversities and phylogenetic analyses of three Chinese main ethnic groups in southwest China: a Y-Chromosomal STR study. Sci Rep. 2018;8(1):15339.
- 34. Sun K, et al. Application of machine learning for ancestry inference using multi-InDel markers. Forensic Sci Int Genet. 2022;59:102702.
- 35. Wang X, et al. Investigating the effectiveness of forensic genetics and population genetic diversity using a multi-InDel system in Chinese Hezhou and Southern Shaanxi Han populations. Ann Hum Genet; 2024.
- Wei G, Zhou R. Comparison of machine learning and deep learning models for evaluating suitable areas for premium teas in Yunnan, China. PLoS ONE. 2023;18(2):e0282105.
- Günen MA. Performance comparison of deep learning and machine learning methods in determining wetland water areas using EuroSAT dataset. Environ Sci Pollut Res Int. 2022;29(14):21092–106.
- Park YR, et al. Comparison of machine and deep learning for the classification of cervical cancer based on cervicography images. Sci Rep. 2021;11(1):16143.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.