

RESEARCH

Open Access



Analysis of maternal genetic structure of mitochondrial DNA control region from Tai-Kadai-speaking Buyei population in southwestern China

Yuhang Feng^{1†}, Li Chen^{1†}, Xiaoxue Wang¹, Hongling Zhang¹, Qiyang Wang¹, Yubo Liu¹, Xiaoye Jin¹, Meiqing Yang¹, Jiang Huang^{1*} and Zheng Ren^{1*}

Abstract

Background Even though the Buyei are a recognised ethnic group in southwestern China, there hasn't been much work done on forensic population genetics, notably using mitochondrial DNA. The sequences and haplogroups of mitochondrial DNA control regions of the Buyei peoples were studied to provide support for the establishment of a reference database for forensic DNA analysis in East Asia.

Methods and results The mitochondrial DNA control region sequences of 200 Buyei individuals in Guizhou were investigated. The haplotype frequencies and haplogroup distribution of the Buyei nationality in Guizhou were calculated. At the same time, the paired *Fst* values of the study population and other populations around the world were computed, to explore their genetic polymorphism and population relationship. A total of 179 haplotypes were detected in the Buyei population, with frequencies of 0.005–0.015. All haplotypes were assigned to 89 different haplogroups. The haplotype diversity and random matching probability were 0.999283 and 0.0063, respectively. The paired *Fst* genetic distances and correlation p-values among the 54 populations revealed that the Guizhou Buyei was most closely related to the Henan Han and the Guizhou Miao, and closer to the Hazara population in Pakistan and the Chiang Mai population.

Conclusions The study of mitochondrial DNA based on the maternal genetic structure of the Buyei nationality in Guizhou will benefit the establishment of an East Asian forensic DNA reference database and provide a reference for anthropological research in the future.

Keywords mtDNA, Control region, Guizhou Buyei, Haplotype, Haplogroup

[†]Yuhang Feng and Li Chen contributed equally to this work.

*Correspondence:

Jiang Huang
mmm_hj@126.com
Zheng Ren
mmm_rz@163.com

¹ Department of Forensic Medicine, Guizhou Medical University, Guiyang 550004, Guizhou, China

Introduction

Human mitochondrial DNA (mtDNA) has an elevated copy number in comparison with nuclear DNA, so it can be used as a forensic sample if nuclear DNA is not available [1]. In addition to its elevated copy number, mtDNA is advantageous in that it does not recombine, and mutations accumulate over time. Matrilineal pedigree analysis can be performed based on mtDNA sequences if sufficient autosomal DNA is not available [2]. With the



application of numerous molecular biological detection techniques in medicine, mtDNA sequence analysis has been properly verified and has become a reliable technique for detecting biological evidence in forensic criminal cases.

Buyei people are one of 17 permanent ethnic minorities in southwest China's Guizhou Province, descended from the ancient "Baiyue", mainly distributed in Guizhou, Yunnan, Sichuan, and other provinces, of which Guizhou Province has the largest population, accounting for 97% of the nationwide Buyei population. The Buyei language is part of the Tai-Kadai family of Sino-Tibetan languages. (<https://www.britannica.com/topic/Buyei>) With its lengthy history and distinctive customs, the Buyei ethnic group deserves anthropological and demographic genetic studies. Currently, there is a lack of mtDNA sequence data for the Buyei people, which is insufficient for forensic science and demographic genetic studies. In our study, the mtDNA control region sequences of 200 unrelated individuals of the Buyei people of Guizhou Province were analyzed to aid in the establishment of the database and the determination of ancestral composition from the point of view of matrilineal inheritance.

Materials and methods

Sample collection

Blood samples were collected from 200 unrelated paternity tests of Buyei individuals in Guizhou Province. All of these people are indigenous Buyei people of Guizhou who are not related by blood within three generations. All participants provided written informed consent after we explained the purpose and procedure of the study. The mitochondrial DNA control region sequences of all Tai-Kadai-speaking Buyei individuals generated in this study have been submitted to GenBank (<http://www.ncbi.nlm.nih.gov/BankIt/>), and the accession numbers are ON983171-ON983370.

DNA extraction, amplification, and sequencing

In accordance with the manufacturer's instructions, DNA was extracted using the QIAamp DNA Mini kit (Qiagen, Hagen, Germany). GeneAmp PCR System 9700 (Thermo Fisher, Waltham, MA) was used to amplify the entire mtDNA control region using primers F15975 and R637 (displayed in Table S1; in the supplementary material). Each reaction mixture contained One-Shot LA PCR Mix 25 μ l (TaKaRa Bio Inc., Dalian, China), 0.4 μ M of each primer, and 10 ng DNA. The amplification was carried out under the conditions of 95°C for 5 min, 30 amplification cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 30 s, and one full extension cycle at 72°C for 10 min. Purification of the PCR products was performed using Exonuclease I (TaKaRa Bio Inc., Dalian, China) and Shrimp

Alkaline Phosphatase (TaKaRa Bio Inc., Dalian, China), and they were sequenced with the BigDyeTM Terminator version 3.0 Ready Reaction Cycle Sequencing Kit (Thermo Fisher, Waltham, MA) using the 3730xl DNA Analyser (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's manual. Following an earlier report, sequencing primers were used [3]. A combination of forward and reverse-direction sequencing was used to enhance the data's accuracy.

Sequence nomenclature and haplogroup assignment

With the help of the DNAMAN software (<http://www.lynnon.com/>), forward and reverse sequences were aligned and compared with the revised Cambridge Reference Sequences (rCRSs) [4]. As suggested by Parson et al., insertions at nt16193, nt309, nt315, and nt573 were omitted from statistical analyses and all comparisons [5]. Haplogroup assignment was performed using PhyloTree Build 17 [6]-based Haplogrep [7] and EMMA [8].

Statistical analysis

The direct counting method was used to calculate haplotype and haplogroup frequencies. The diversity of haplotype and random matching probability were estimated according to Stoneking et al [9]. In addition, our data are compared with other data available from the literature, including Guizhou Miao [10], Henan Han [11], Chinese Bai [12], Xinjiang Mongolian [13], Southwest Gelao [14], Yunnan Dai [15], Liaoning Han [15], Pinghua Han in Guangxi [16], Xinjiang Kazakh [17], Beijing Han [15], Mulao in Guangxi [18], Chinese Hui [19], Kashmiri [20], Hazara people of Pakistan [21], African Americans in Orange, California [22], African Americans in Vermont, California [22], Parana of Brazil [23], Iranian [24], Alto Parana [25], Iraqi [26], Arabian [26], Kuwaiti [26], Palestinian [26], Anatolians in Turkey [26], Kurds in Iran [27], Bosnian of Roma [28], Gulagic in Africa [29], Punjab in Pakistan [30], Sierra Leone Mendes in West Africa [29], Tengenai in Sierra Leone [29], Mandinka in Sierra Leone [29], Polish Gypsies [31], Finns [32], Turks [33], Northern Tunis [34], North-central Moroccan [34], Mozabit [34], Moroccan Berber [34], Gdansk in northern Poland [35], Upper Silesia in southern Poland [35], Novgorod in northwest Russia [35], Christmas Island in Australia [36], Romanians [37], Pukhtunhwa in Pakistan [38], Mosuo [39], Bengali [40], Khattak of the Peshawar Valley [41], Khesghi of the Peshawar Valley [41], Slovak [42], Hokkaido in Japan [43], Vietnamese [44], Dutch [45] and People from Chiang Mai, Thailand [46]. The paired *Fst* values were computed with Arlequin version 3.5 software [47], and the *Fst* matrix data was imported into the "pheatmap" package of R software (<https://www.r-project.org/>) to plot the heatmap. To gain a more comprehensive understanding of the population relationships among

Table 1 Haplogroup frequencies of 200 Chinese Buyei individuals

Broad Haplogroup	Number	Frequency	Haplogroup	Number	Frequency
A	2	0.01	A15	1	0.005
-	-	-	A19	1	0.005
B	41	0.205	B4	1	0.005
-	-	-	B4 + 16261	4	0.02
-	-	-	B4a1a	1	0.005
-	-	-	B4a1c	1	0.005
-	-	-	B4a4	1	0.005
-	-	-	B4b1	1	0.005
-	-	-	B4b1a + 207	1	0.005
-	-	-	B4b1a2a	1	0.005
-	-	-	B4c1b	1	0.005
-	-	-	B4c1b2b	1	0.005
-	-	-	B4c1b2c1	1	0.005
-	-	-	B4c2	1	0.005
-	-	-	B4g	2	0.01
-	-	-	B4h1	1	0.005
-	-	-	B4k	1	0.005
-	-	-	B4m	1	0.005
-	-	-	B5a	18	0.09
-	-	-	B5a1d	1	0.005
-	-	-	B5a2	1	0.005
-	-	-	B5b	1	0.005
C	9	0.045	C	1	0.005
-	-	-	C4a1	1	0.005
-	-	-	C4c1b	6	0.03
-	-	-	C7 + 16051	1	0.005
D	13	0.065	D4a3	3	0.015
-	-	-	D4a7	1	0.005
-	-	-	D4b2b	1	0.005
-	-	-	D4b2b2a	1	0.005
-	-	-	D4e1a	3	0.015
-	-	-	D4g2a	1	0.005
-	-	-	D5b	2	0.01
-	-	-	D6c	1	0.005
F	33	0.165	F1a1	4	0.02
-	-	-	F1a1a	2	0.01
-	-	-	F1a1a1	3	0.015
-	-	-	F1a1d	4	0.02
-	-	-	F1a3a2	1	0.005
-	-	-	F1a4a	1	0.005
-	-	-	F1c1a1	1	0.005
-	-	-	F2b1	3	0.015
-	-	-	F2d	3	0.015
-	-	-	F3a + 207	4	0.02
-	-	-	F3a1	5	0.025
-	-	-	F3b1a2	1	0.005
-	-	-	F4a	1	0.005
G	5	0.025	G2a	2	0.01
-	-	-	G2a + 152	1	0.005

Table 1 (continued)

Broad Haplogroup	Number	Frequency	Haplogroup	Number	Frequency
-	-	-	G2b1a1	1	0.005
-	-	-	G3	1	0.005
M	69	0.345	M	1	0.005
-	-	-	M1	1	0.005
-	-	-	M11 + 200	1	0.005
-	-	-	M11a	1	0.005
-	-	-	M12	1	0.005
-	-	-	M1a3	1	0.005
-	-	-	M33a1b	2	0.01
-	-	-	M74a	3	0.015
-	-	-	M76a	1	0.005
-	-	-	M7b1a1	2	0.01
-	-	-	M7b1a1 + (16192)	17	0.085
-	-	-	M7b1a1a	9	0.045
-	-	-	M7b1a1a3	1	0.005
-	-	-	M7b1a1b	8	0.04
-	-	-	M7b1a1e2	1	0.005
-	-	-	M7b1a2a	3	0.015
-	-	-	M7c1c2	4	0.02
-	-	-	M7c1c2a	1	0.005
-	-	-	M8a	1	0.005
-	-	-	M8a2'3	2	0.01
-	-	-	M8a2a	2	0.01
-	-	-	M8a3a	1	0.005
-	-	-	M9a1a	1	0.005
-	-	-	M9a1a1c	1	0.005
-	-	-	M9a1a2*2	1	0.005
-	-	-	M9a1b + 150	1	0.005
-	-	-	M9a'b	1	0.005
N	4	0.02	N5	1	0.005
-	-	-	N9a1	1	0.005
-	-	-	N9a10 + 16311	2	0.01
R	24	0.12	R9	14	0.07
-	-	-	R9b1	3	0.015
-	-	-	R9b1a3	2	0.01
-	-	-	R9b1b	1	0.005
-	-	-	R9b2	1	0.005
-	-	-	R9c1	1	0.005
-	-	-	R9c1a	1	0.005
-	-	-	R9c1b1	1	0.005

All samples were assigned to 89 different haplogroups and sub-haplogroups (Figure S1 in the Supplementary Materials). The most common haplogroups were B5a (18 cases, 9%), followed by M7b1a1 + (16192) (17 cases, 8.5%), B4b1 (7 cases, 3.5%), and R9 (14 cases, 7%). At the broader haplogroup level, the most common is M (69 samples, 34.5%), followed by B (41 samples, 20.5%), and F (33 samples, 16.5%)

various populations, the principal component analysis (PCA) was conducted based on haplogroup frequencies using the Multivariate Statistical Package version 3.22 (MVSP) [48]. The neighbor-joining (NJ) phylogenetic tree based on the pairwise *Fst* value matrix was constructed with the assistance of MEGA 11 software [49].

Results

In 200 individuals, a total of 179 haplotypes (89.5%) were observed (Table S1, in the supplementary material), of which 163 haplotypes were unique (91.1%). The most common haplotypes were 16140C, 16183C, 16188.1C, 16189C, 16266A, 16519C, 73G, 210G, 263G,

309.1C, 315.1C, 522DEL, 523DEL (haplogroup B5a); 16108 T, 16129A, 16162G, 16172C, 16304C, 73G, 150 T, 195C, 248DEL, 263G, 315.1C, 522DEL, 523DEL (haplogroup F1a1a1); 16129A, 16192 T, 16223 T, 16297C, 73G, 150 T, 182 T, 199C, 263G, 315.1C, 489C (haplogroup M7b1a1+(16192)); 16129A, 16192 T, 16223 T, 16297C, 73G, 150 T, 199C, 263G, 309.1C, 315.1C, 489C (haplogroup M7b1a1+(16192)); 16086C, 16297C, 16324C, 16399G, 73G, 199C, 263G, 315.1C, 489C (haplogroup M7b1a2a), that were all shared by three individuals (1.5%). The haplotype diversity and random matching probability of the mtDNA control region of the Buyei nationality in Guizhou were 0.999283 and 0.0063, respectively.

All samples were assigned to 89 different haplogroups and sub-haplogroups (Table 1, Figure S1 in the Supplementary Materials). The most common haplogroups were B5a (18 cases, 9%), followed by M7b1a1+(16192) (17 cases, 8.5%), B4b1 (7 cases, 3.5%), and R9 (14 cases, 7%). At the broader haplogroup level, the most common is M (69 samples, 34.5%), followed by B (41 samples, 20.5%), and F (33 samples, 16.5%).

In order to further reveal the genetic similarities and divergences among the Guizhou Buyei population and 53 reference worldwide populations, the paired *Fst* genetic distance between the studied population and reference populations was calculated. The results were displayed in the form of a heatmap (Fig. 1). As shown in Table S2

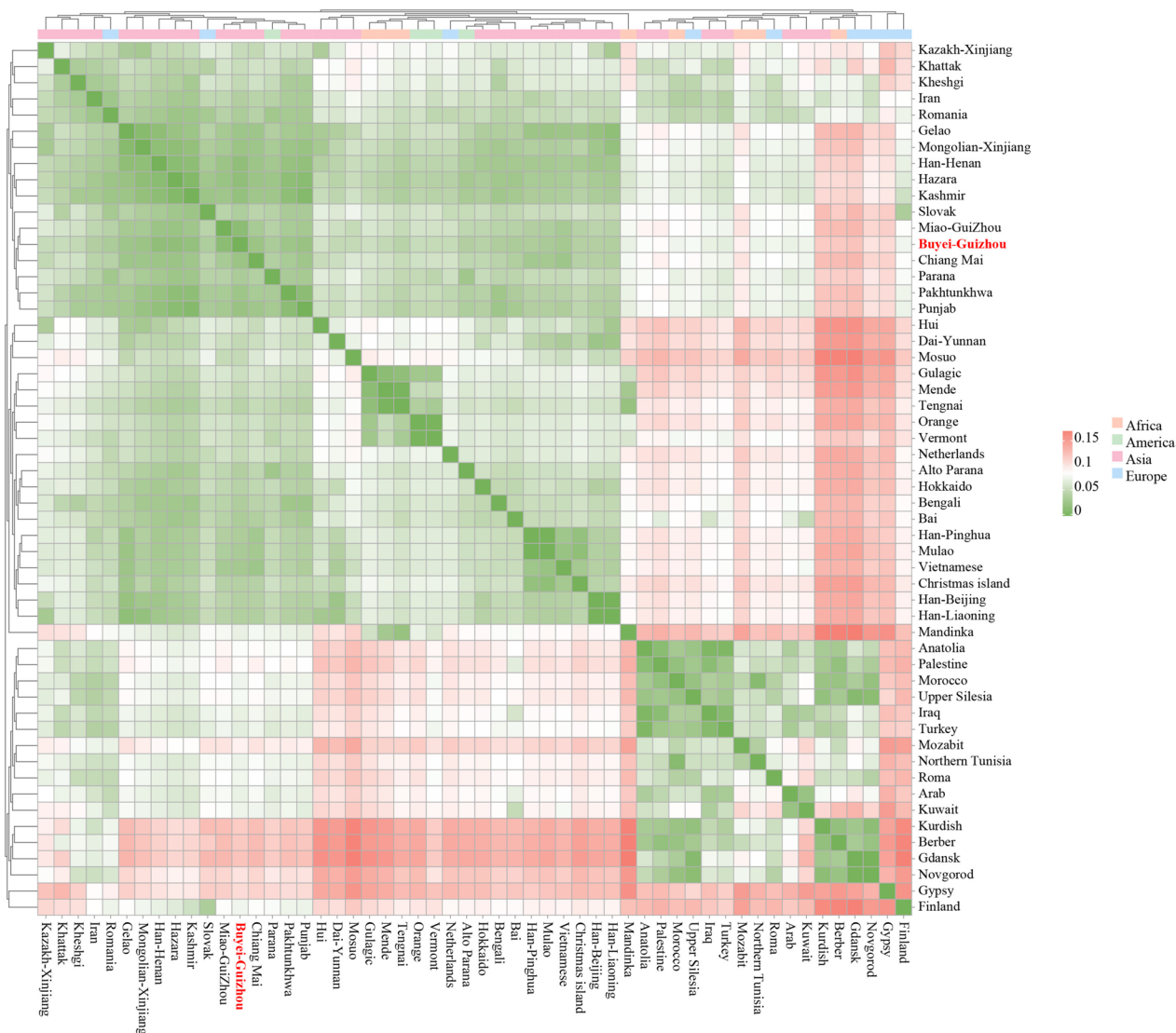


Fig. 1 Heatmap for genetic distances between the studied Guizhou Buyei population and 53 worldwide reference populations. Visualising the *Fst* values with different colours. The powder orange represents the high *Fst* values, and green represents the low *Fst* values. Highlight the Guizhou Buyei population in red font

of the supplementary material, the paired *Fst* and p-values of the Guizhou Buyei population and 53 other published populations worldwide were calculated. Among the 13 reference populations in China, the genetic differentiation between the Guizhou Buyei and Guizhou Miao was the smallest (with the closest genetic affinity, *Fst*=0.01508), followed by the Henan Han population (*Fst*=0.01799). The genetic distance between the north-west Hui and Guizhou Buyei was the largest (with the farthest genetic affinity, *Fst*=0.05908). The showed that among 40 global reference populations (except China), the Guizhou Buyei population and Pakistan Hazara population had the smallest genetic distance (with the closest genetic affinity, *Fst*=0.01783), followed by the Kashmiri

(*Fst* = 0.02084), and had the largest genetic differentiation (with the farthest genetic affinity, *Fst* = 0.12165) with the Gdansk people in Poland.

To elucidate the genetic relationship between the Buyei population of Guizhou and global populations, the PCA based on haplogroup frequencies was also conducted. The results of the PCA based on haplogroup frequencies indicated that the first three principal components account for 43.488% of the variation. Specifically, PC1 explains 29.172% of the variation, PC2 accounts for 8.129%, and PC3 for 6.187%. The PCA visualisation (Fig. 2) showed that geographic clustering reveals two distinct clusters: one made up of African populations and the other of mixed East and Southeast Asian populations.

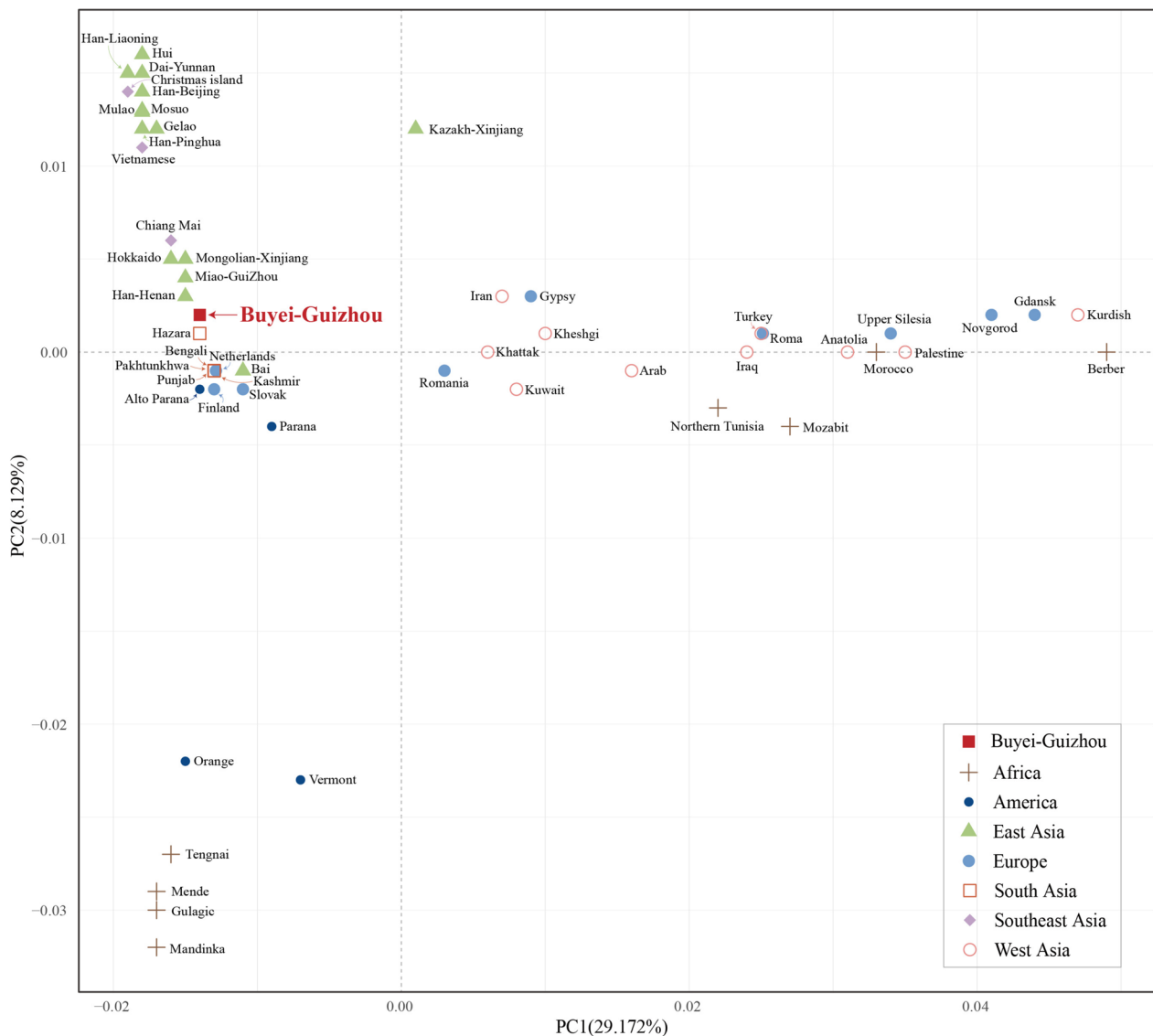


Fig. 2 The PCA plot for the Guizhou Buyei population and 53 worldwide populations according to geographic distributions. The research population belongs to the East Asia cluster (indicated in the green triangle). The Buyei-Guizhou was highlighted in bold red and indicated by arrows

The plot revealed that West Asian populations and certain European populations cluster together, with instances of partial overlap observed. Through PC1 and PC2, most populations can be differentiated; however, these components have virtually no effect on European populations. The detailed findings indicated that the point representing the Buyei population of Guizhou is situated within the East Asian cluster, in close proximity to the points representing the Han Chinese from Henan (Han-Henan) as well as the Hazaras of Pakistan (Hazara).

To elucidate the phylogenetic relationships between the Guizhou Buyei and global reference populations, an NJ phylogenetic tree was constructed based on *Fst* values (Figs. 3 and 4). The NJ tree, based on the pairwise *Fst* genetic distances among the research population and 13 reference populations within China, as depicted in Fig. 3, indicated that the Guizhou Buyei clusters on the same branch as the Miao population, which is also located in Guizhou. Additionally, the research population shared a relatively close phylogenetic relationship with the Han population from Henan. This showed that the genetic distance differentiation among the three populations is comparatively small, which is consistent with what is shown in Table S2. Based on the NJ tree, which has the Buyei population of Guizhou and 40 reference populations from around the world (excluding China,

Fig. 4), the results showed that the research population is grouped with other Asian populations in a main branch, with the Tai-Kadai-speaking population in Chiang Mai, Thailand (Chiang Mai), having the closest phylogenetic relationship.

Through the analysis of the *Fst* genetic matrix, PCA, and NJ tree, it could be concluded that geographic location, historical factors, and language families have a certain influence on gene intergenerational transmission. Based on the characteristics of matrilineal inheritance, the genetic distance of mtDNA haplogroups might be positively correlated with the geographical distance and language families among populations.

Discussion

In this research, the haplotypes and haplogroups of mtDNA of the Buyei population in Guizhou were studied and analysed, and it could be found that many individuals had the same haplotype. MtDNA is matrilineal, so the selected samples may come from the same matrilineal line, although this point could not be confirmed by their ID cards or oral descriptions. In addition, it might also be because the polymorphism of the mtDNA control region is not sufficient to distinguish these individuals.

The results of the population genetics analysis indicated that the Guizhou Buyei has a very close genetic

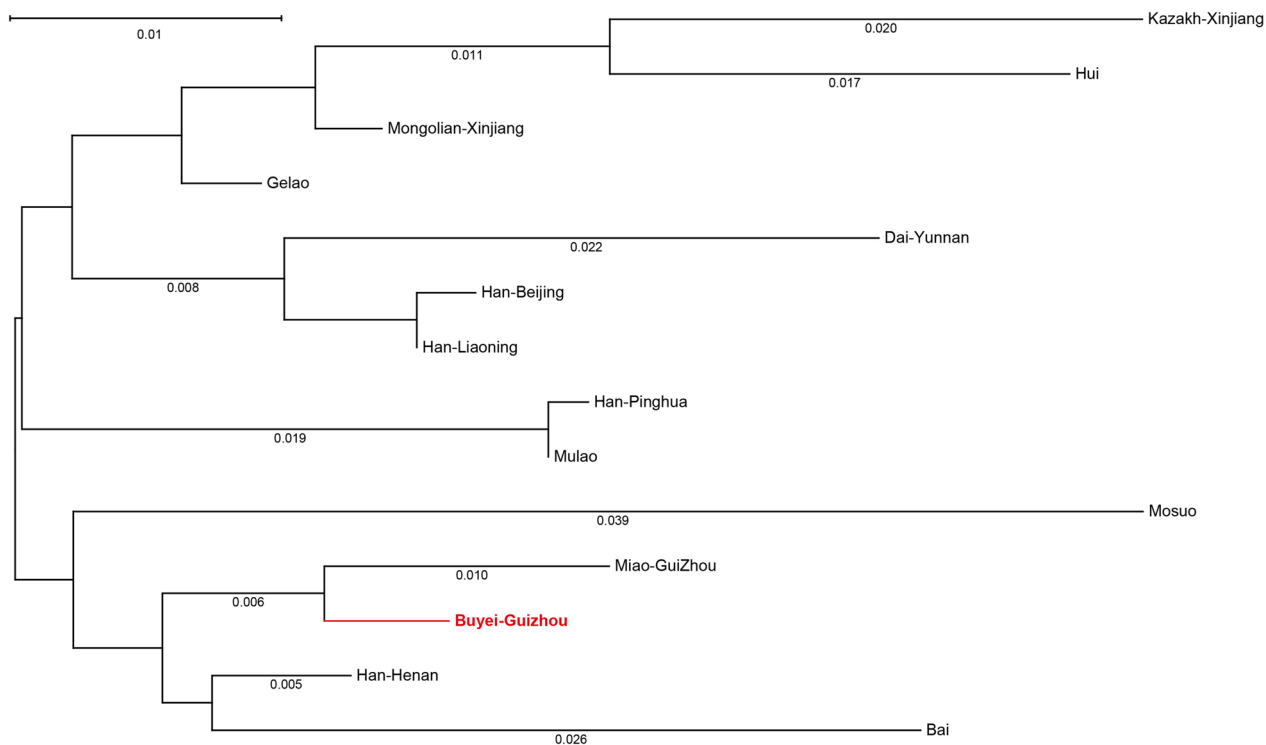


Fig. 3 The NJ phylogenetic tree based on the paired *Fst* distance matrix between Guizhou Buyei and 13 Chinese populations. Highlight the research population of Buyei-Guizhou with bold red text

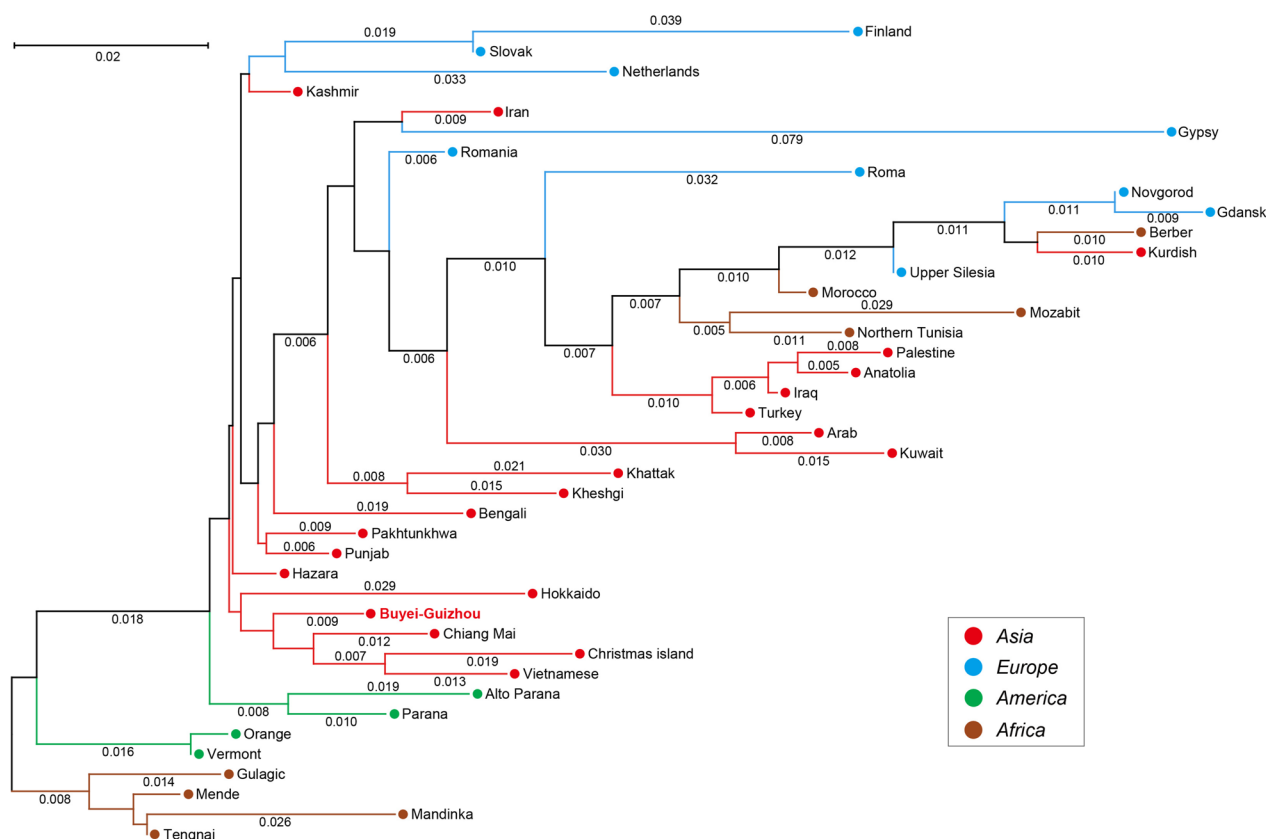


Fig. 4 The NJ phylogenetic tree based on the paired *Fst* distance matrix between Guizhou Buyei and 40 populations in the world. Group populations according to the different continents and mark each population with distinct colours

relationship with the Miao population in the same region, which might be related to their geographical location and common genetic pattern. Moreover, due to the differences in language and culture, there was little gene communication between ethnic minorities and Han nationalities, which led to a great differentiation of genetic distance among them. Nevertheless, in our study, there was little genetic differentiation between Henan Han and Guizhou Buyei. This might be related to the origin and history of the Buyei population and the historical status of Henan. According to historical documents, the ancestors of the Buyei people were one of the main ethnic groups in the ancient Yelang Kingdom, but after the development of this country, its control area was far beyond the areas inhabited by the Buyei ancestors, and many tribes might live in these areas [50]. During this period, gene exchanges between the Buyei nationality and the Han nationality might take place. In addition, research has recorded that in the Ming and Qing dynasties, the court dispatched a large number of Han troops into Guizhou to hoard the military, and many soldiers married local Buyei people and merged into the Buyei. It showed that the Buyei ethnic group may have absorbed

a small part of the Han ethnicity in the Ming and Qing dynasties [51]. And Henan belongs to the Yellow River basin, one of the birthplaces of the ancient Han population [52]. Although the gene exchange between the modern Han nationality and other ethnic groups was extremely rare, under the above historical background, the gene exchange between the ancient Han nationality and ethnic minorities was possible. Due to the matrilineal inheritance of mitochondria, it has been accumulated so far, which is consistent with the results of our study. The above discussion was only supported by historical data and wasn't supported by scientific data. Consequently, we still need to rely on modern science and technology for further research and verification, such as mitochondrial whole genome sequencing and so on.

On the other hand, compared with other populations in the world (except China), the Guizhou Buyei exhibited a closer genetic affinity with populations in Asia, likely due to geographical proximity. Notably, the studied population shared close genetic relationships with South Asian populations, particularly the Hazara people of Pakistan, which may stem from Pakistan's historical role in ancient trade and commerce as well as the presence of the ancient

“Silk Road” facilitating genetic exchanges [53]. This was very consistent with the effect of geographic location and historical background factors on genetic differentiation. The close relationship between the Guizhou Buyei and the people of Chiang Mai, Thailand, is likely influenced by their geographical location, and both populations belong to the Tai-Kadai linguistic family.

Conclusion

To sum up, in our study, haplotypes, haplogroups, and the population structure of the Buyei population in Guizhou were analysed based on mtDNA. The sequences of the mtDNA control region of the Buyei nationality in Guizhou had high polymorphism and a large amount of information. It could provide detailed information about the degree of mtDNA variation and haplogroup distribution. Additionally, it could be widely used in forensic case studies, anthropological analysis, and population genetics research. In addition, through the analysis of genetic polymorphism and population structure, it was found that the genetic distance of the Guizhou Miao, Henan Han, Hazara, and Chiang Mai populations were close to that of the Guizhou Buyei population. It was found that the genetic relationship between the Guizhou Buyei and other reference populations was predominantly consistent with how they have spread geographically and linguistically. This means that the maternal lineages determined by mtDNA exhibit a close correlation with geographical factors and linguistic families.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09941-x>.

Additional file 1: Table S1. Haplotypes observed in 200 Buyei individuals. In 200 individuals, a total of 179 haplotypes (89.5%) were observed, of which 163 haplotypes were unique (91.1%).

Additional file 2: Table S2. The pairwise *F_{st}* values (below diagonal lines) and *p*-values (above diagonal lines) of Guizhou Buyei population and 53 other published populations worldwide. The paired *F_{st}* and *p*-values of the Guizhou Buyei population and 13 other published populations in China were calculated. The genetic differentiation between the Guizhou Buyei and Guizhou Miao was the smallest (with the closest genetic affinity, *F_{st}*= 0.01508), followed by the Henan Han nationality (*F_{st}*= 0.01799). The genetic distance between the northwest Hui and Guizhou Buyei was the largest (with the farthest genetic affinity, *F_{st}*= 0.05908). The paired *F_{st}* genetic distance and correlation coefficient *p*-values between Guizhou Buyei nationality and 40 other reference populations in the world (except China) showed that Guizhou Buyei nationality and Pakistan Hazara nationality had the smallest genetic distance (with the closest genetic affinity, *F_{st}*= 0.01783), followed by Kashmiri (*F_{st}*= 0.02084), and had the largest genetic differentiation (with the farthest genetic affinity, *F_{st}*= 0.12165) with the Gdansk people in Poland.

Additional file 3: Figure S1. Guizhou Buyei were assigned to 89 different haplogroups and sub-haplogroups. The most common haplogroups were B5a (18 cases, 9%), followed by M7b1a1+(16192) (17 cases, 8.5%), B4b1 (7 cases, 3.5%), and R9 (14 cases, 7%).

Acknowledgements

Sincere thanks to the volunteers who provided the blood samples.

Authors' contributions

Conceptualization (lead): ZR and JH; Writing—review and editing (equal): ZR and JH; Software and Methodology: YHF, HLZ, and QYW; Methodology: XXW, YBL, XYJ and MQY. YHF and LC analyzed the data, and were major contributors in writing the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the foundations as follows: National Natural Science Foundation (No. 82160324), Guizhou Scientific Support Project, Qian Science Support [2021] General 448, Shanghai Key Lab of Forensic Medicine, Key Lab of Forensic Science, Ministry of Justice, China (Academy of Forensic Science), Open Project, KF202009, Guizhou Province Education Department, Characteristic Region Project, Qian Education KY No. [2021]065, Guizhou “Hundred” High-level Innovative Talent Project, Qian Science Platform Talents [2020]6012, Guizhou Scientific Support Project, Qian Science Support [2020] 4Y057, Guizhou Science Project, Qian Science Foundation [2020] 1Y353, Guizhou Scientific Cultivation Project, Qian Science Platform Talent [2018] 5779-X, Guizhou Engineering Technology Research Center Project, Qian High-Tech of Development and Reform Commission NO. [2016] 1345.

Availability of data and materials

The mitochondrial DNA control region sequences of all Tai-Kadai-speaking Buyei individuals generated in this study have been submitted to GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), and the accession number was ON983171-ON983370.

Declarations

Ethics approval and consent to participate

All participants provided written informed consent after we explained the purpose and procedure of the study. Ethical permission was issued by the Ethics Committee of Guizhou Medical University (2021 Ethics Approval Document NO.65), and in accordance with the recommendations of the revised Helsinki Declaration [54].

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 17 October 2023 Accepted: 26 December 2023
Published online: 11 January 2024

References

- Bandelt HJ, van Oven M, Salas A. Haplogrouping mitochondrial DNA sequences in legal medicine/forensic genetics. *Int J Legal Med*. 2012;126(6):901–16.
- Brujijns B, Tiggelaar R, Gardeniers H. Massively parallel sequencing techniques for forensics: a review. *Electrophoresis*. 2018;39(21):2642–54.
- Feng Y, Zhang H, Wang Q, Yang M, Liu Y, Wang J, Huang J, Ren Z. The mitochondrial DNA control region sequences from the Chinese Sui population of southwestern China. *Ann Hum Biol*. 2021;48(7–8):635–40.
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999;23(2):147.
- Parson W, Gusmão L, Hares DR, Irwin JA, Mayr WR, Morling N, Pokorak E, Prinz M, Salas A, Schneider PM, et al. DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing. *Forensic Sci Int Genet*. 2014;13:134–42.
- van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 2009;30(2):E386–394.

7. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt HJ, Kronenberg F, Salas A, Schönherr S. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 2016;44(W1):W58–63.
8. Röck AW, Dür A, van Oven M, Parson W. Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). *Forensic Sci Int Genet.* 2013;7(6):601–9.
9. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet.* 1991;48(2):370–82.
10. Le C, Ren Z, Zhang H, Wang Q, Yang M, Liu Y, Huang J, Wang J. The mitochondrial DNA control region sequences from the Chinese Miao population of southeastern China. *Ann Hum Biol.* 2019;46(7–8):606–9.
11. Xu K, Hu S. Population data of mitochondrial DNA HVS-I and HVS-II sequences for 208 Henan Han Chinese. *Leg Med (Tokyo).* 2015;17(4):287–94.
12. Chen F, Yin CY, Qian XQ, Fan HT, Deng YJ, Zhang YD, Meng HT, Shen CM, Yang CH, Jin R, et al. Single nucleotide polymorphisms of mitochondrial DNA HVS-I and HVS-II in Chinese Bai ethnic group. *Electrophoresis.* 2015;36(6):930–6.
13. Lan Q, Xie T, Jin X, Fang Y, Mei S, Yang G, Zhu B. MtDNA polymorphism analyses in the Chinese Mongolian group: efficiency evaluation and further matrilineal genetic structure exploration. *Mol Genet Genomic Med.* 2019;7(10):e00934.
14. Liu C, Wang SY, Zhao M, Xu ZY, Hu YH, Chen F, Zhang RZ, Gao GF, Yu YS, Kong QP. Mitochondrial DNA polymorphisms in Gelao ethnic group residing in Southwest China. *Forensic Sci Int Genet.* 2011;5(1):e4–10.
15. Yao L, Xu Z, Wan L. Whole mitochondrial DNA sequencing analysis in 47 Han populations in Southwest China. *Med Sci Monit.* 2019;25:6482–90.
16. Gan RJ, Pan SL, Mustavich LF, Qin ZD, Cai XY, Qian J, Liu CW, Peng JH, Li SL, Xu JS, et al. Pinghua population as an exception of Han Chinese's coherent genetic structure. *J Hum Genet.* 2008;53(4):303–13.
17. Xie T, Hu L, Guo YX, Li YC, Chen F, Zhu BF. Genetic polymorphism analysis of mitochondrial DNA from Chinese Xinjiang Kazak ethnic group by a novel mitochondrial DNA genotyping panel. *Mol Biol Rep.* 2019;46(1):17–25.
18. Wang XQ, Wang CC, Deng QY, Li H. [Genetic analysis of Y chromosome and mitochondrial DNA poly-morphism of Mulam ethnic group in Guangxi, China]. *Yi Chuan.* 2013;35(2):168–74.
19. Chen C, Li Y, Tao R, Jin X, Guo Y, Cui W, Chen A, Yang Y, Zhang X, Zhang J, et al. The genetic structure of Chinese Hui ethnic group revealed by complete mitochondrial genome analyses using massively parallel sequencing. *Genes (Basel).* 2020;11(11):1352.
20. Rakha A, Peng MS, Bi R, Song JJ, Salahudin Z, Adan A, Israr M, Yao YG. EMPOP-quality mtDNA control region sequences from Kashmiri of Azad Jammu & Kashmir, Pakistan. *Forensic Sci Int Genet.* 2016;25:125–31.
21. Rakha A, Fatima, Peng MS, Adan A, Bi R, Yasmin M, Yao YG. mtDNA sequence diversity of Hazara ethnic group from Pakistan. *Forensic Sci Int Genet.* 2017;30:e1–5.
22. Scheible M, Just R, Sturk-Andreaggi K, Saunier J, Parson W, Parsons T, Coble M, Irwin J. The mitochondrial landscape of African Americans: an examination of more than 2500 control region haplotypes from 22 U.S. locations. *Forensic Sci Int Genet.* 2016;22:139–48.
23. Poletto MM, Malaghini M, Silva JS, Bicalho MG, Braun-Prado K. Mitochondrial DNA control region diversity in a population from Parana state-increasing the Brazilian forensic database. *Int J Legal Med.* 2019;133(2):347–51.
24. Terreros MC, Rowold DJ, Mirabal S, Herrera RJ. Mitochondrial DNA and Y-chromosomal stratification in Iran: relationship between Iran and the Arabian Peninsula. *J Hum Genet.* 2011;56(3):235–46.
25. Simão F, Strobl C, Vullo C, Catelli L, Machado P, Huber N, Schnaller L, Huber G, Xavier C, Carvalho EF, et al. The maternal inheritance of Alto Paraná revealed by full mitogenome sequences. *Forensic Sci Int Genet.* 2019;39:66–72.
26. Theyab JB, Al-Bustan S, Crawford MH. The genetic structure of the Kuwaiti population: mtDNA Inter- and intra-population variation. *Hum Biol.* 2012;84(4):379–403.
27. Zarei F, Rajabi-Maham H. Phylogeography, genetic diversity and demographic history of the Iranian Kurdish groups based on mtDNA sequences. *J Genet.* 2016;95(4):767–76.
28. Ahmic A, Kalamujic B, Ismailovic A, Lasic L, Halilovic E, Mujkic I, Pojskic N. Mitochondrial DNA diversity of the Roma from northeastern Bosnia, Bosnia and Herzegovina. *Homo.* 2018;69(6):347–56.
29. Jackson BA, Wilson JL, Kirbah S, Sidney SS, Rosenberger J, Bassie L, Alie JA, McLean DC, Garvey WT, Ely B. Mitochondrial DNA genetic diversity among four ethnic groups in Sierra Leone. *Am J Phys Anthropol.* 2005;128(1):156–63.
30. Bhatti S, Abbas S, Aslamkhan M, Attimonelli M, Trinidad MS, Aydin HH, de Souza EMS, Gonzalez GR. Genetic perspective of uniparental mitochondrial DNA landscape on the Punjabi population, Pakistan. *Mitochondrial DNA A DNA Mapp Seq Anal.* 2018;29(5):714–26.
31. Malyarchuk BA, Grzybowski T, Derenko MV, Czarny J, Miścicka-Sliwka D. Mitochondrial DNA diversity in the Polish Roma. *Ann Hum Genet.* 2006;70(Pt 2):195–206.
32. Hedman M, Brandstätter A, Pimenoff V, Sistonen P, Palo JU, Parson W, Sajantila A. Finnish mitochondrial DNA HVS-I and HVS-II population data. *Forensic Sci Int.* 2007;172(2–3):171–8.
33. Serin A, Canan H, Alper B, Korkut Gulmen M, Zimmermann B, Parson W. Mitochondrial DNA control region haplotype and haplogroup diversity in South Eastern Turkey. *Forensic Sci Int Genet.* 2016;24:176–9.
34. Turchi C, Buscemi L, Giacchino E, Onofri V, Fendt L, Parson W, Tagliabracchi A. Polymorphisms of mtDNA control region in Tunisian and Moroccan populations: an enrichment of forensic mtDNA databases with Northern Africa data. *Forensic Sci Int Genet.* 2009;3(3):166–72.
35. Grzybowski T, Malyarchuk BA, Derenko MV, Perkova MA, Bednarek J, Woźniak M. Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis. *Forensic Sci Int Genet.* 2007;1(2):141–7.
36. Wise CA, Sullivan SG, Black ML, Erber WN, Bittles AH. Y-chromosome and mitochondrial DNA studies on the population structure of the Christmas Island community. *Am J Phys Anthropol.* 2005;128(3):670–7.
37. Turchi C, Stanciu F, Paselli G, Buscemi L, Parson W, Tagliabracchi A. The mitochondrial DNA makeup of Romanians: a forensic mtDNA control region database and phylogenetic characterization. *Forensic Sci Int Genet.* 2016;24:136–42.
38. Bhatti S, Aslamkhan M, Abbas S, Attimonelli M, Aydin HH, de Souza EMS. Genetic analysis of mitochondrial DNA control region variations in four tribes of Khyber Pakhtunkhwa, Pakistan. *Mitochondrial DNA A DNA Mapp Seq Anal.* 2017;28(5):687–97.
39. Lu Y, Wang C, Qin Z, Wen B, Farina SE, Jin L, Li H. Mitochondrial origin of the matrilineal Mosuo people in China. *Mitochondrial DNA.* 2012;23(1):13–9.
40. Li YC, Wang HW, Tian JY, Li RL, Rahman ZU, Kong QP. Cultural diffusion of Indo-aryan languages into Bangladesh: a perspective from mitochondrial DNA. *Mitochondrion.* 2018;38:23–30.
41. Zubair M, Hemphill BE, Schurr TG, Tariq M, Ilyas M, Ahmad H. Mitochondrial DNA diversity in the Khattak and Khesghi of the Peshawar Valley, Pakistan. *Genetica.* 2020;148(3–4):195–206.
42. Malyarchuk BA, Perkova MA, Derenko MV, Vanecsek T, Lazur J, Gomolcak P. Mitochondrial DNA variability in Slovaks, with application to the Roma origin. *Ann Hum Genet.* 2008;72(Pt 2):228–40.
43. Asari M, Umetsu K, Adachi N, Azumi J, Shimizu K, Shiono H. Utility of haplogroup determination for forensic mtDNA analysis in the Japanese population. *Leg Med (Tokyo).* 2007;9(5):237–40.
44. Tran TTH, Nguyen DH, Tran VK, Nguyen QL, Trinh HA, Luong LH, Tran VA, Pham LAT, Nguyen TT, Nguyen VB, et al. Variation of mitochondrial DNA HV1 AND HV2 of the Vietnamese Population. *Adv Exp Med Biol.* 2020;1292:37–63.
45. Chaitanya L, van Oven M, Brauer S, Zimmermann B, Huber G, Xavier C, Parson W, de Knijff P, Kayser M. High-quality mtDNA control region sequences from 680 individuals sampled across the Netherlands to establish a national forensic mtDNA reference database. *Forensic Sci Int Genet.* 2016;21:158–67.
46. Zimmermann B, Bodner M, Amory S, Fendt L, Röck A, Horst D, Horst B, Sanguansersri T, Parson W, Brandstätter A. Forensic and phylogeographic characterization of mtDNA lineages from northern Thailand (Chiang Mai). *Int J Legal Med.* 2009;123(6):495–501.
47. 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.

48. Kovach W. MVSP-A multivariate statistical package for windows, ver. 3.1. Pentraeth, Wales: Kovach Computing Services; 1999. p. 137.
49. Tamura K, Stecher G, Kumar S. MEGA1 1: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol.* 2021;38(7):3022–7.
50. Chen C. The Bouyei Ethnic Group and the State of Yelang. Guizhou, China: Bouyei studies; 1998.
51. Wei Q. Research on Bouyei Cultural studies. Guizhou, China: Guizhou people's publishing house; 1999.
52. Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, Gao Y, Mao X, Zhang L, et al. Genetic evidence supports demic diffusion of Han culture. *Nature.* 2004;431(7006):302–5.
53. Mezzavilla M, Vozzi D, Pirastu N, Girotto G, d'Adamo P, Gasparini P, Colonna V. Genetic landscape of populations along the Silk Road: admixture and migration patterns. *BMC Genet.* 2014;15:131.
54. Nathanson V. Revising the declaration of Helsinki. *BMJ.* 2013;346: f2837.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

