



Male-Dominated Migration and Massive Assimilation of Indigenous East Asians in the Formation of Muslim Hui People in Southwest China

Qiyang Wang^{1†}, Jing Zhao^{2†}, Zheng Ren¹, Jin Sun², Guanglin He², Jianxin Guo², Hongling Zhang¹, Jingyan Ji¹, Yubo Liu¹, Meiqing Yang¹, Xiaomin Yang², Jinwen Chen², Kongyang Zhu², Rui Wang², Yingxiang Li², Gang Chen³, Jiang Huang^{1*} and Chuan-Chao Wang^{2*}

OPEN ACCESS

Edited by:

Horolma Pamjav,
Ministry of Interior, Hungary

Reviewed by:

Balazs Egyed,
Eötvös Loránd University, Hungary
Jatupol Kampuansai,
Chiang Mai University, Thailand

*Correspondence:

Jiang Huang
mmm_hj@126.com
Chuan-Chao Wang
wang@xmu.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Genetics

Received: 17 October 2020

Accepted: 10 December 2020

Published: 11 January 2021

Citation:

Wang Q, Zhao J, Ren Z, Sun J,
He G, Guo J, Zhang H, Ji J, Liu Y,
Yang M, Yang X, Chen J, Zhu K,
Wang R, Li Y, Chen G, Huang J and
Wang C-C (2021) Male-Dominated
Migration and Massive Assimilation
of Indigenous East Asians
in the Formation of Muslim Hui People
in Southwest China.
Front. Genet. 11:618614.
doi: 10.3389/fgene.2020.618614

¹ Department of Forensic Medicine, Guizhou Medical University, Guiyang, China, ² Department of Anthropology and Ethnology, Institute of Anthropology, School of Life Sciences, Xiamen University, Xiamen, China, ³ WeGene, Shenzhen, China

The origin and diversification of Muslim Hui people in China via demic or simple cultural diffusion is a long-going debate. We here generated genome-wide data at nearly 700,000 single nucleotide polymorphisms (SNPs) from 45 Hui and 14 Han Chinese individuals collected from Guizhou province in southwest China. We applied principal component analysis (PCA), ADMIXTURE, *f*-statistics, *qpWave*, and *qpAdm* analysis to infer the population genetic structure and admixture history. Our results revealed the Guizhou Hui people have a limited amount of West Eurasian related ancestry at a proportion of 6%, but show massive genetic assimilation with indigenous southern Han Chinese and Tibetan or Tungusic/Mongolic related northern East Asians. We also detected a high frequency of North Asia or Central Asia related paternal Y-chromosome but not maternal mtDNA lineages in Guizhou Hui. Our observation supports the cultural diffusion has played a vital role in the formation of Hui people and the migration of Hui people to southwest China was probably a sex-biased male-driven process.

Keywords: genetic structure, population genetics, southwest China, population admixture, population history

INTRODUCTION

The Hui people are an East Asian ethno-religious group distributing throughout China with a population of approximately 20 million predominantly composed of Chinese speaking practitioners of Islam (Mu, 1985). Although the majority of Hui people speak Han Chinese language nowadays, their culture and food habits have distinct differences with Han Chinese. The origin and diversification of Hui groups via demic diffusion involving the mass movement of people from West Eurasia and the Middle East or simple cultural diffusion with massive assimilation of indigenous East Asians is a long-going debate. The historical records suggest the origins of Hui people were mainly in two different periods: first, starting from the Tang dynasty (around seventh century AD), the Persians and Arabs came to the southeast coast of China for trade and then gradually mixed with

local East Asian populations, especially with the Han Chinese. Second, the Central Asians, Persians and Arabs came to China following the Mongol invasions and conquests during the thirteenth and fourteenth centuries in the Yuan Dynasty (Bao, 1982; Du and Yip, 1993; Gladney, 1998). The immigration was suggested to be male-dominated involving a large number of soldiers, merchants, and political emissaries (Du and Yip, 1993).

Previous studies from the genetic perspective revealed that the origin of Hui people in China had involved massive assimilation of indigenous East Asians inferred from paternal Y-chromosomal single nucleotide polymorphism (SNP) and short tandem repeat (STR) analysis (Zhang et al., 2010; Lan et al., 2018; Wang et al., 2019; Xie et al., 2019). From the maternal mitochondrial DNA (mtDNA) side, Western Eurasian related lineages were found in Hui people of Xinjiang in northwest China, but only accounting for a low frequency at about 6.7% (Yao et al., 2004). The autosomal STR analysis on Hui people in Gansu province in northwest China showed genetic homogeneity of the Muslim populations and local East Asian populations, with no evidence of substantial gene flow from the Middle East or Europe into Hui people during their Islamization (Xie and Shan, 2002; Yao et al., 2016). Previous studies have shed some light but far from giving a comprehensive and conclusive understanding of the origin of Hui people due to the limited markers used. Besides, the majority of previous genetic studies on Hui people have focused on northern China, but the Hui people in southern China are seldom investigated.

Guizhou province in southwest China is a multi-ethnic province with abundant genetic and cultural diversities. The Han Chinese have the largest proportion of the population in Guizhou, accounting for 62.2% of the total population of the province. The Muslim Hui is about 90,000 people mainly distributing in Weining County, accounting for 0.5% of the total population in Guizhou province. Weining County was the earliest place for the Hui people moving to Guizhou Province (Bao, 1982). According to historical records, the ancestors of the Hui people had already arrived in Guizhou in the Yuan Dynasty (Bao, 1982). The origin of the Chinese Hui is always controversial though considerable studies have been performed. For exploring the origin and genetic structure of Guizhou Hui, in this study, we generated genome-wide data including 699,537 paternal, maternal phylogenetic relevant SNPs and autosomal SNPs on Guizhou Hui samples using array genotyping. In addition, we also sampled Guizhou Han as a reference population to facilitate the analysis on the assimilation of indigenous ethnic groups in Guizhou Hui people. Here we aimed to explore the origin and genetic admixture of Guizhou Hui people and shed light on the understanding of early Chinese Hui migration patterns from a genetic perspective.

MATERIALS AND METHODS

We collected 59 saliva samples in Guizhou province with informed consent, including 45 Hui people from Weining County and 14 Han people from Guiyang city. We collected

the samples following genetic population criteria. These samples enrolled in the present study were collected randomly from unrelated participants whose parents and grandparents are indigenous people and have the non-consanguineous marriage within the same ethnical group for at least three generations. We listed the detailed sample information in **Supplementary Table S1** and the geographic locations of sampling in **Figure 1**. Of all samples, 23 were male and used for Y-haplogroup paternal analysis. All individuals were performed for maternal mtDNA analysis.

Genomic DNA of 59 samples were extracted using DP-318 Kit (Tiangen Biotechnology, Beijing). The DNA quality control was carried out at the experimental center of WeGene-Shenzhen. Genotyping was performed on the Illumina WeGene Arrays at the WeGene genotyping center, Shenzhen. We merged the data of Guizhou Hui and Han individuals with published populations from Human Origin Dataset (Patterson et al., 2012; Lazaridis et al., 2014), Simons Genome Diversity Project (SGDP) (Mallick et al., 2016), 1000 Genomes Project (1000 Genomes Project Consortium et al., 2015), and also ancient East Eurasian samples from 1240K capture dataset curated by David Reich lab¹.

We carried out Principal Component Analysis (PCA) using smartpca, part of the EIGENSOFT package (Patterson et al., 2006). We used default parameters with the numoutlieriter: 0 settings and assessed statistical significance with a Tracy-Widom test by the twstats program of EIGENSOFT. We discussed and plotted principal components in what follows were highly statistically significant ($P < 10^{-12}$). We carried out ADMIXTURE (Alexander et al., 2009) analysis by the following steps: we firstly pruned SNPs in strong linkage disequilibrium with each other using PLINK tools with the parameters “-indep-pairwise 200 25 0.4”; we then run ADMIXTURE with the K values (number of assumed ancestral components) ranging from 2 to 10 applying 100 bootstraps iterations with different random seeds. The best run was chosen according to the highest likelihood. An optimal K value was selected using 10-fold cross-validation plotted in **Figure 2C**.

We used ADMIXTOOLS (Patterson et al., 2012) to compute f -statistics and determined standard errors with a block jackknife and default parameters. We computed outgroup f_3 -statistics of the form f_3 (Mbuti; X, Y), which measured the shared genetic drift between the two populations X and Y since their separation from an African outgroup Mbuti. We computed admixture f_3 -statistics of the form f_3 (Target; Source 1, Source 2) to explore possible sources for Hui_Guizhou and Han_Guizhou people. We also calculated the f_4 -statistics of the form f_4 (X, Y; Test, Outgroup) to show if population Test is symmetrically related to X and Y or shares an excess of alleles with either of the two. Moreover, we used *qpAdm* (Haak et al., 2015) as implemented in ADMIXTOOLS with the option “allsnps: YES” to test the number of sources of ancestry that is needed to estimate the admixture proportions of Guizhou Hui population with the proposed sources. The outgroups selected are differentially related to the ancestral sources of Guizhou Hui people.

¹<https://reich.hms.harvard.edu/downloadable-genotypes-present-day-and-ancient-dna-data-compiled-published-papers>



FIGURE 1 | Geographic locations of sampling marked with black solid circles, including 45 Hui and 14 Han individuals in Guizhou province of southwest China. Detail information for each samples was listed in **Supplementary Table S1**.

The Y chromosomal haplogroups were assigned by identifying the most derived upstream allele and the most ancestral downstream allele in the phylogenetic tree in the ISOGG version 11.89². The mtDNA haplogroup assignment was determined with mtDNA tree Build 16 (van Oven and Kayser, 2009)³.

RESULTS

Genetic Structure of Hui and Han Ethnic Groups in Guizhou

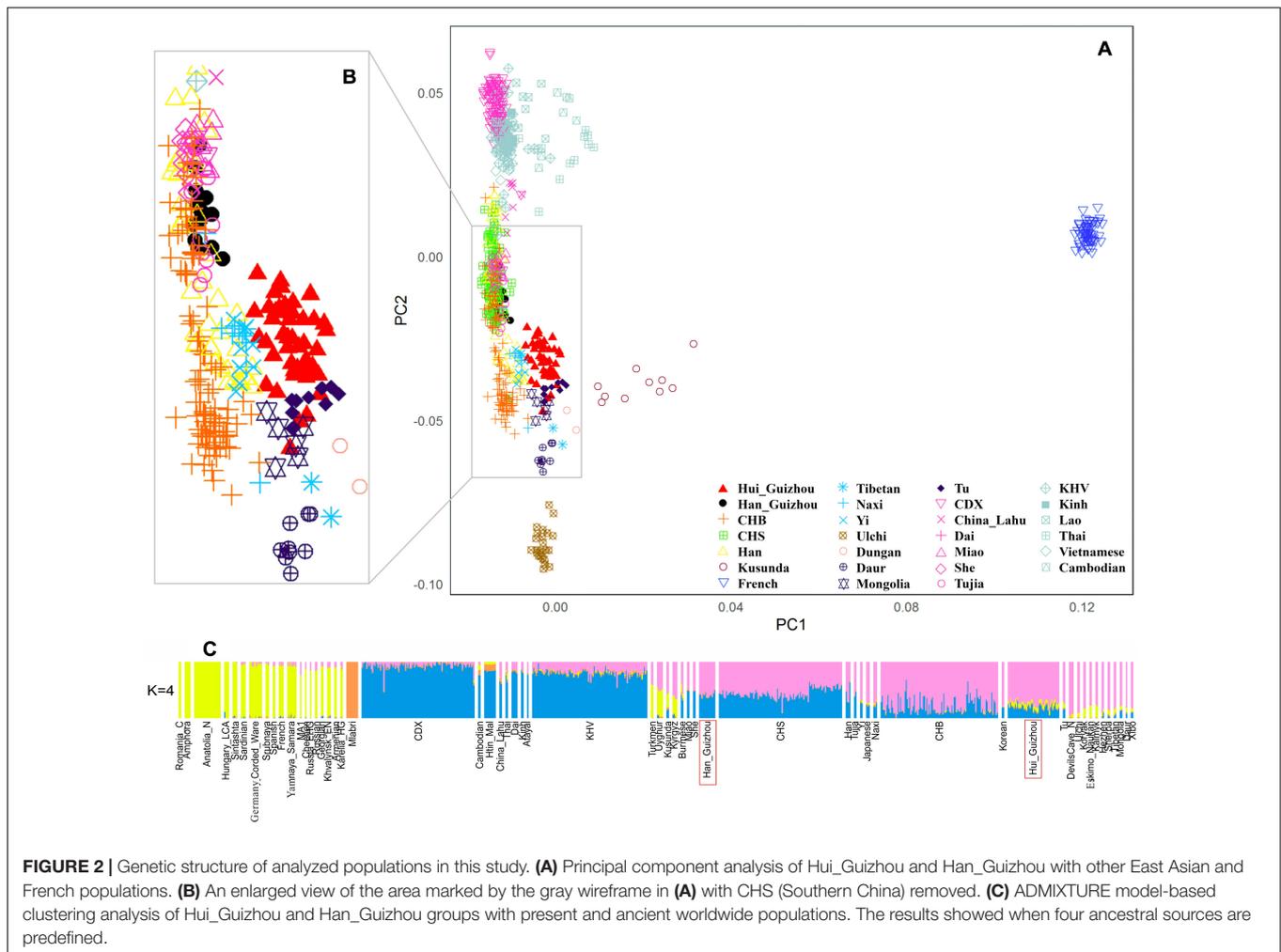
We first carried out a PCA to obtain a qualitative picture of how Guizhou Hui and Han individuals related to previously published East Asian populations from Human Origin Dataset (Patterson et al., 2006; Lazaridis et al., 2014). We used the names “Hui_Guizhou” and “Han_Guizhou” to refer to samples from Hui and Han ethnic groups in Guizhou province, respectively. The result was shown in **Figure 2A**. To facilitate observation, we enlarged the area of interest (gray wireframe in **Figure 2A**) on the left (**Figure 2B**) by removing CHS (Han Chinese in southern China) to avoid visual clutter. Based on the first and second principal components, Guizhou Hui and Han individuals

were divided into two distinct genetic clusters. In addition, we found that Hui_Guizhou people clustered closely with northern populations (such as Mongolic-speaking, Sino-Tibetan, Tu, and CHB (Han in Beijing, China) populations) rather than with southern populations. The genetic structure of Hui_Guizhou was not in accordance with their geographical distribution. Unlike Hui_Guizhou, Han_Guizhou overlapped with populations in southern China, such as CHS, She, Miao and Tujia.

The results of model-based ADMIXTURE clustering analysis (Alexander et al., 2009) were consistent with PCA showing that Hui_Guizhou population had a closer affinity on average to northern populations instead of southern groups (**Figure 2C** and **Supplementary Figure S2**). **Figure 2C** showed the ADMIXTURE results assuming $K = 4$ clusters (we selected this number because it was the most optimal K value using 10-fold cross-validation). The primary ancestry component assigned to the Hui_Guizhou population, shown in purple, also maximized in Neolithic DevilsCave hunter-gatherers (Siska et al., 2017; Sikora et al., 2019) and present-day Ulchi individuals in Russian Far East. ADMIXTURE also assigned to Hui_Guizhou samples an ancestry component of yellow color, which maximized West Eurasians but was absent in the southern Chinese populations. The ancestry composition of Hui_Guizhou shown in **Figure 2C** indicated that there were more northern related, followed by the southern related, and a few West Eurasian related ancestry components.

²<http://www.isogg.org/tree/>

³<http://www.phylotree.org/>



Han_Guizhou people displayed an analogous genetic profile with southern Chinese populations, like CHS, Miao, She and Tujia in terms of admixture proportions related to southern and northern components.

Population Continuity and Admixture in the Hui_Guizhou and Han_Guizhou

The outgroup f_3 -statistics (Raghavan et al., 2014) of the form f_3 (Mbuti; X, Y) were consistent with the patterns observed in the above PCA and ADMIXTURE analysis, suggesting that Hui_Guizhou shared more genetic drift with Tu and Tibeto-Burman speaking populations (**Supplementary Figure S1**). However, Han_Guizhou showed close genetic proximity with southern Chinese populations, especially Hmong-Mien speaking populations, Tujia and CHS. In addition, we used negative Z-scores of admixture f_3 -statistics of the form f_3 (Target; Source 1, Source 2) to explore possible sources for Hui_Guizhou and Han_Guizhou people in this study. The results in **Supplementary Table S2** showed that the top possible related sources of Hui_Guizhou consisted of Han related especially Han_Guizhou in this study, Hmong-Mien speaking populations, and ancient

and present-day West Eurasians. The top possible sources of the Han_Guizhou were Atayal, Dai and northern-related populations such as Tibetan, Ulchi and Hezhen groups. Inferred from the f_3 -statistics (**Supplementary Table S2**), although the genetic structure of Hui_Guizhou was closer to northern related populations, we still found the signal of possible genetic assimilation with indigenous southern related populations, like Han_Guizhou and Hmong-Mien speaking populations (**Supplementary Figure S3**).

To further explore the differentiation between the Hui_Guizhou and indigenous populations, we performed the f_4 statistics (Patterson et al., 2012) in the form of f_4 (Test, Mbuti; Hui_Guizhou, Han_Guizhou). **Supplementary Table S3** showed that present-day and ancient Europeans shared more alleles with Hui_Guizhou people than with Han_Guizhou, suggesting West Eurasian gene flow into Hui_Guizhou. We confirmed the results by replacing Han_Guizhou with other Han Chinese groups in datasets (**Supplementary Table S4**). In addition, we continued to calculate f_4 statistics in the form of f_4 (Test, Mbuti; Hui_Guizhou, She) and f_4 (Test, Mbuti; Hui_Guizhou, Miao), respectively (**Supplementary Tables S5, S6**). Similarly, East Asian populations shared more alleles with Miao and She groups

than with Hui_Guizhou, while Hui_Guizhou people tended to be closer to Western Eurasian-related populations.

The Ancestry Related Sources of Hui_Guizhou

We systematically explored diverse *qpAdm*-based admixture models for identifying plausible admixture sources for Hui_Guizhou people in this study. Firstly, we used Han_Guizhou and French as proxies for the East Asian and western related source populations in a two-way admixture. Hui_Guizhou individuals were estimated to have 6.2% French related and 93.8% Han_Guizhou related ancestry (**Figure 3A** and **Supplementary Table S7**). We then replaced Han_Guizhou with Han Chinese in published datasets to confirm the estimation (**Supplementary Table S8**). We observed a consistent result regardless of whether one of the proxies was Han Chinese or Han_Guizhou.

Furthermore, we were particularly interested in East Asian related sources in Hui_Guizhou people. We were wondering if we could distinguish the Han_Guizhou related local southern component and Tibetan/Tungusic related northern ancestry. We proposed a more complex three-way model using Han_Guizhou, Tibetan, and French as three sources (**Supplementary Table S9** and **Figure 3B**). In the East Asian related sources, Han related ancestry had the largest proportion accounting for 69.6%, while Tibetan related ancestry also had a prominent proportion of 24%. We then replaced Tibetan with ancient DevilsCave (**Supplementary Table S10** and **Figure 3D**) and present-day Ulchi (**Supplementary Table S11** and **Figure 3E**) in the Russian Far East, as well as Daur (**Supplementary Table S12** and **Figure 3C**) in northern China. The admixture proportion of Neolithic DevilsCave related nomadic ancestry was estimated to be 17.4% in Hui_Guizhou. The Daur related ancestry was also estimated at a similar proportion of around 16.3% in Hui_Guizhou. However, we observed a difference in the ancestry proportion of present-day Ulchi related sources in Hui_Guizhou, which was estimated to be 7.4%. These results above indicated the ancestors of Hui people in Guizhou had a large amount of Tibetan/Tungusic related northern ancestry before they migrated to southwest China.

MtDNA and Y Chromosomal Haplogroup Analysis

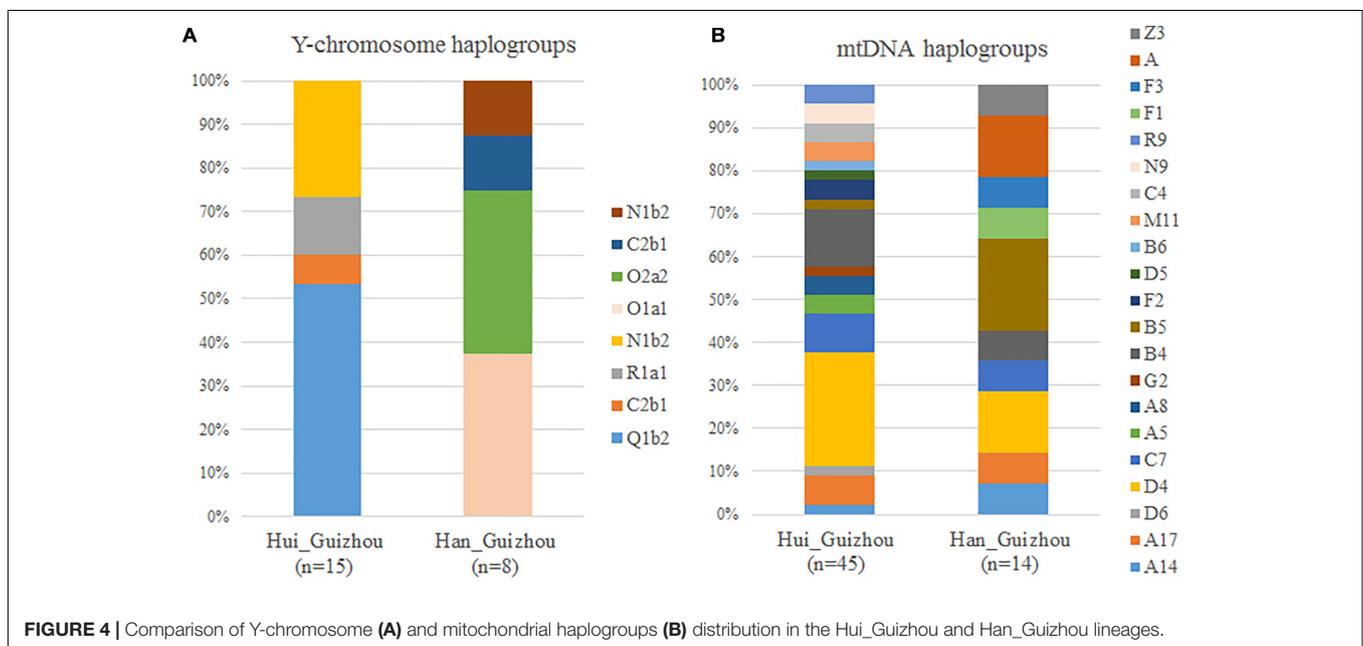
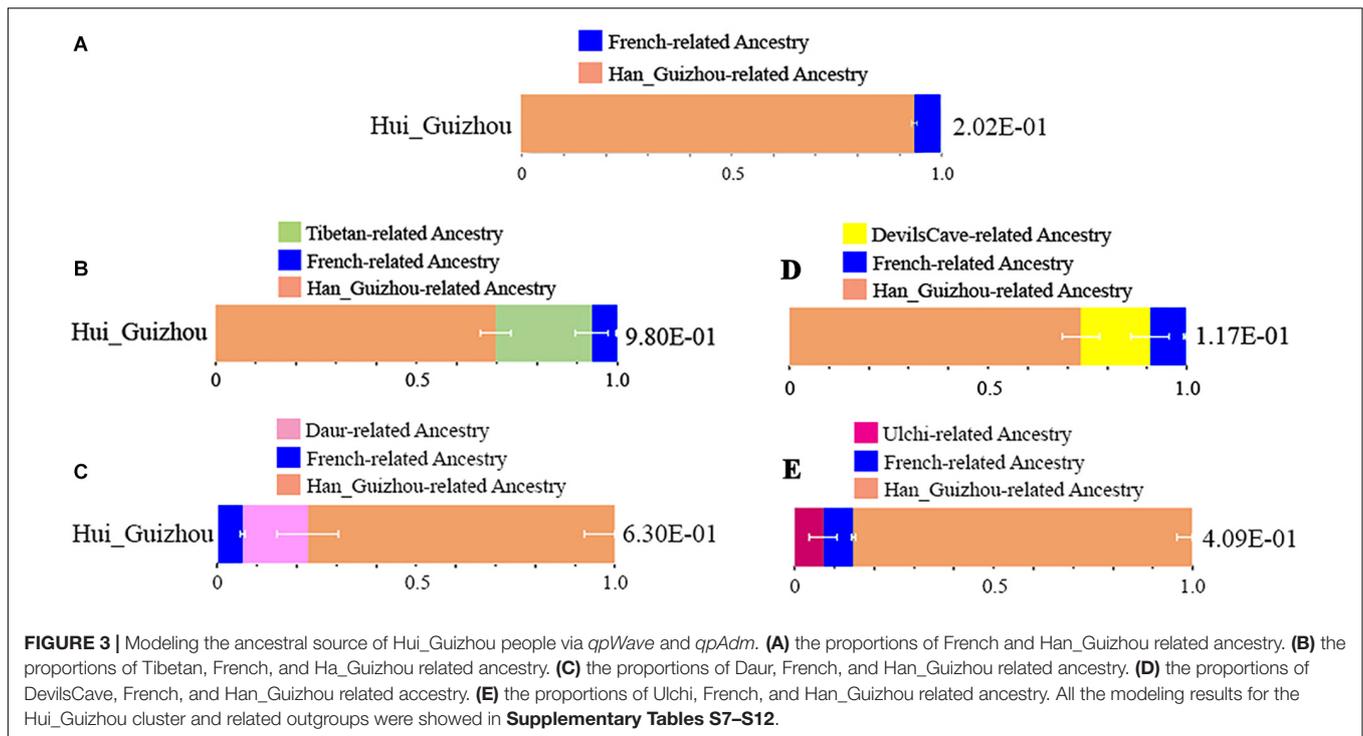
We showed the maternal mtDNA haplogroups of Hui_Guizhou and Han_Guizhou samples in **Supplementary Table S13** and **Figure 4B**. Although we found a high diversity of the lineages in the maternal gene pool of Hui_Guizhou and Han_Guizhou, these mtDNA haplogroups detected were all frequent in East Asian populations. We observed a genetic North-South admixture pattern in the haplogroup frequency profile of mtDNA. In detail, haplogroup D4, D5 and D6 accounted for 31.11% (14/45) in Hui_Guizhou and 14.29% (2/14) in Han_Guizhou individuals. Previous studies suggested that these haplogroups were predominant in populations from northern China (Tanaka et al., 2004; Wen et al., 2004; Kong et al., 2011; Li et al., 2019). Haplogroup G2 and Z found in Hui_Guizhou and

Han_Guizhou were also relatively abundant in northern China (Tanaka et al., 2004; Li et al., 2007). In addition, haplogroup B, F and R were mainly found in southwest Chinese and Southeast Asian populations (Tanaka et al., 2004; Li et al., 2007), contributed to 26.67% (12/45) of Hui_Guizhou and 42.86% (6/14) of Han_Guizhou people. The other lineages, like haplogroup A and N9, contributed to 17.78% (8/45) in Hui_Guizhou and 28.57% (4/12) in Han_Guizhou populations, which had a prevailing northern Asia dispersal (Tanaka et al., 2004; Li et al., 2019). Haplogroup C and its sublineages were considered to have expanded in Northeast Asia (Tanaka et al., 2004), accounting for 13.33% (6/45) of Hui_Guizhou and 7.14% (1/14) of Han_Guizhou people.

To trace the paternal history and genealogical ancestors of Hui and Han Chinese in Guizhou, we performed Y-chromosome haplogroups analysis. On the paternal side, a striking feature of the Y chromosomal profile of Hui_Guizhou was the high frequency of haplogroup Q1b2b1b2a-L330-F1893 (**Supplementary Table S13** and **Figure 4A**), accounting for 53.33% (8/15) of the total Hui samples. Haplogroup Q1b2b1b2a-L330-F1893 was a subclade of Q-L330. Q-L330 was the major sub-lineage of the Q-M242 samples in populations of southern Siberia and the adjacent region of the Mongolia Plateau, like the Altaians, Tuvans, and Kets (Dulik et al., 2012; Karmin et al., 2015). Previous studies had revealed the haplogroup Q-M242 might have originated in southern Siberia and diffused gradually to other parts of Eurasia since the Paleolithic Age (Dulik et al., 2012; Balanovsky et al., 2017; Huang et al., 2018). The second frequent haplogroup was N1b2a2-M1811 accounting for 26.67% of Hui_Guizhou people. N1b2a2-M1811 was the sub-clades of N1b-F2930, which was prevalent in the present-day Sino-Tibetan populations (Ilumäe et al., 2016). R1a1a1b2-F992 and R1a1a1b1a2b3-FGC4499/Y2192 were subclades of haplogroup R1a1a-M17, which likely migrated from the Central Asia steppe (Sharma et al., 2009; Underhill et al., 2010). Unlike the profile of Y-chromosome haplogroups in Hui_Guizhou people, Han_Guizhou samples had high-frequency subclades of haplogroup O1a and O2a, which are the major local paternal lineages in East and Southeast Asia (van Oven et al., 2012; Yan et al., 2014). In particular, the haplogroups O1a1a2a1-CTS701, O1a1a1b1-Z23406, and O1a1a1a1a1a1a-A12439 in Han_Guizhou people were sublineages of O1a-M119, which was prevalent along the southeast coast of China (Wang and Li, 2013). The subhaplogroup of O2a2b1a1a-F8, F42 were suggested to be one of the three super-grandfathers for present-day Chinese that experienced star-like expansions in the Neolithic Era about 5.4 thousand years ago (Yan et al., 2014). Compared with Han Chinese in Guizhou, Hui_Guizhou people on the paternal perspective had more lineages that were frequent in North Asia and Central Asia.

DISCUSSION

The Hui people are a relatively small population (0.5% of the population of the province) but have a long history living in Guizhou province in southwest China. According to historical



records, the Hui people came to Weining County in Guizhou province during the period of the early formation of Hui population in China (Bao, 1982; Yu, 2015). However, the origin and migration history of Hui people has always been obscure, due to the complicated ethnic origins and the lack of genome-wide data. Therefore, research on the Hui people in Guizhou is helpful for us to understand the early migration pattern and historical practice of the Hui people in China. In this study, we generated genome-wide SNP data from Hui and local Han

Chinese people in Guizhou province. We merged our data with previously published datasets involving ancient and present-day populations and comprehensively integrated the genetic evidence of maternal, paternal and autosomal results to infer the origin and admixture history of the Hui people in Guizhou.

Male-Driven Migration of Guizhou Hui

From the paternal perspective, we found the Y-chromosome lineages that are frequent in North Asia and Central Asia

(Yan et al., 2014; Ilumäe et al., 2016) reached a higher frequency in Guizhou Hui people than in Guizhou Han Chinese. However, the paternal profile of Guizhou Han was quite different from Guizhou Hui by mainly having lineages that are predominant in East and Southeast Asia. However, on the maternal side, we found Guizhou Hui and Han shared almost all the mtDNA haplogroups. The inconsistency of paternal and maternal genetic profile indicated the migration practice of Guizhou Hui people was probably a sex-biased male-driven process. We noted the small number of male individuals in Guizhou Hui samples was a limitation of the study. However, the observed sex-biased pattern in Hui people were consistent with previous studies (Wang et al., 2019; Xie et al., 2019), as well as historical records (Bao, 1982; Mu, 1985; Xie and Shan, 2002) that the ancestors of Hui people migrated to China were mainly men. Their marriage was carried out in a relatively closed system of endogamy and the intermarriage usually involved indigenous Han women converting to Islam Hui when married to the Hui males (Gladney, 1998).

Genetic Assimilation With Indigenous Han People

We found genetic evidence in Hui to support their intense admixture with indigenous Han people. The maternal lineages had shown Guizhou Han shared the majority of mtDNA haplogroups with Guizhou Hui people, such as B5, D4, C7, B4, A17, A14, and F. In addition, the *qpAdm*-based admixture models also demonstrated that Guizhou Hui shared more alleles with Guizhou Han than with other East Asian populations. The Guizhou Han related ancestry ranged from 70 to 85% in Guizhou Hui samples. These results indicate there have been massive assimilations of indigenous Han populations in the formation of Guizhou Hui people. The genetic affinity between Guizhou Hui and Guizhou Han is consistent with the previous evidence that the genetic profile of Muslim Hui people was closest to the indigenous Han Chinese populations (Xie and Shan, 2002; Zhang et al., 2010; Lan et al., 2018; Xie et al., 2019). Although the culture of Hui people has distinct differences with Han Chinese due to their practice of Islam, the majority of Hui people speak Han Chinese languages nowadays showing the close affinity with Han Chinese.

Limited West Eurasian Related Ancestry in Guizhou Hui People

On the genome-wide side, in addition to the predominant East Asian related ancestry, we also detected there was Western Eurasian related ancestry in Guizhou Hui people. The results of *qpADM*-based admixture models indicated that Guizhou Hui individuals were estimated to have about 6% French related West Eurasian ancestry. We here used French as a proxy to estimate the admixture proportion, but we note the West Eurasian ancestry might be diverse and probably from different regions of West Eurasia. The origin and diversification of Hui groups in China via demic diffusion involving the mass movement of people or simple cultural diffusion is a long-going debate. Here we show the proportion of West Eurasian related

ancestry in quite low in present-day Hui people, supporting the formation of Hui involving massive assimilation of indigenous East Asians.

The North to South Migration of Guizhou Hui People

According to the historical records, since Yuan, Ming and Qing dynasties, Muslim Hui people had spread from Central Asia, Arabia, and Persia to China along the Silk Road Economic Belt (Mu, 1985; Xie and Shan, 2002; Yao et al., 2004). However, it was obscure from where and how the ancestors of Guizhou Hui came to southwest China. Our genetic evidence indicated the Guizhou Hui people showed an affinity with northern populations, such as Tibetan and Tungusic/Mongolic-speaking populations. We found the Guizhou Hui samples could be well modeled as a three-way admixture between Guizhou Han (69.6%), Tibetan (24%), and French (6.4%), or as a mixture of Guizhou Han (77.1%), Daur (16.3%), and French (6.6%). This was not caused by the recent expansion and gene flow of Han Chinese since we can get a similar admixture proportion when using the Neolithic DevilsCave samples as a source in place of present-day Tibetan/Daur. We have also found evidence to support this northern ancestry from paternal Y-chromosome of Guizhou Hui people. We proposed that the Muslim Hui people probably migrated into China via the northern route, then migrated southward into Guizhou province with the massive admixture with the indigenous ethnic groups, especially Han Chinese people, to form the present-day Guizhou Hui group.

In the future work, a larger number of samples from diverse Hui people living in different regions of China were considered to comprehensively reveal the genetic diversity and population history of the Hui ethnic groups.

CONCLUSION

The Guizhou Hui group has a long history and unique cultural features. However, the origin and migration history of Hui people have always been obscure due to the lack of genome-wide data. Research on the Guizhou Hui people will help to provide clues to the long debate on the origin and diversification of the Muslim Hui people in China. In this study, from a genome-wide perspective, we synthetically integrated the genetic results of maternal mtDNA, paternal Y-chromosome and autosomal DNA. We proposed that the present-day Guizhou Hui people were formed by male-dominated migration of West Eurasian related people to East Asia with massive assimilation of indigenous East Asians, especially Han Chinese.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the zenodo repository, accession number (doi: 10.5281/zenodo.4275668).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Guizhou Medical University and Xiamen University (Approval Number: XDYX2019009) and were in accordance with the recommendations provided by the revised Helsinki Declaration of 2000. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JH and C-CW designed this study. JZ and C-CW wrote the manuscript. QW, ZR, HZ, JJ, YLiu, MY, and JH collected the samples. QW, ZR, HZ, JJ, YLiu, MY, YLi, GC, and JH conducted the experiment. JZ, JS, GH, JG, XY, JC, KZ, RW, and C-CW analyzed the data. All authors reviewed the manuscript.

FUNDING

This work was funded by the Guizhou Province Education Department, Characteristic Region Project, Qian Education KY No. (2021)065, the National Natural Science Foundation of China (NSFC 31801040), the Nanqiang Outstanding Young Talents Program of Xiamen University (X2123302), the Fundamental Research Funds for the Central Universities (ZK1144), the Guizhou Province Engineering Technology Research Center Project [Qian High-Tech of Development and

REFERENCES

- 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393
- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Balanovsky, O., Gurianov, V., Zaporozhchenko, V., Balaganskaya, O., Urasin, V., Zhabagin, M., et al. (2017). Phylogeography of human Y-chromosome haplogroup Q3-L275 from an academic/citizen science collaboration. *BMC Evol. Biol.* 17(Suppl. 1):18. doi: 10.1186/s12862-016-0870-2
- Bao, J. X. (1982). When the Hui people moved to Guizhou. *Guizhou Historical Stud.* 1, 61–63.
- Du, R., and Yip, V. F. (1993). *Ethnic Groups in China*. New York, NY: Science Press.
- Dulik, M. C., Zhadanov, S. I., Osipova, L. P., Askapuli, A., Gau, L., Gokcumen, O., et al. (2012). Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *Am. J. Hum. Genet.* 90, 229–246. doi: 10.1016/j.ajhg.2011.12.014
- Gladney, D. C. (1998). *Ethnic Identity in China: The Making of a Muslim Minority Nationality*. San Diego, CA: Harcourt Brace College.
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211. doi: 10.1038/nature14317
- Huang, Y. Z., Pamjav, H., Flegontov, P., Stenzl, V., Wen, S. Q., Tong, X. Z., et al. (2018). Dispersals of the Siberian Y-chromosome haplogroup Q in Eurasia. *Mol. Genet. Genom.* 293, 107–117. doi: 10.1007/s00438-017-1363-8
- Illumäe, A. M., Reidla, M., Chukhryaeva, M., Järve, M., Post, H., Karmin, M., et al. (2016). Human Y Chromosome Haplogroup N: a non-trivial time-resolved

Reform Commission No. (2016)1345], the Guizhou Scientific Support Project [Qian Science Support (2019) 2825], the Guizhou “Hundred” innovative talents project [Qian Science Talent Platform (2020) 6012], the Guizhou Scientific Support Project [Qian Science Support (2020) 4Y057], the Guizhou Science Project [Qian Science Foundation (2020) 1Y353], the Guizhou Medical University Academic Sprout Cultivation Project [Qian Science Platform Talent (2018)5779-X], and the Opening Foundation of Shanghai Key Lab of Forensic Medicine, Key Lab of Forensic Science, Ministry of Justice, China (Academy of Forensic Science) (KF202009).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.618614/full#supplementary-material>

Supplementary Figure 1 | Heatmap showed the shared genetic drift of Hui_Guizhou and Han_Guizhou with other present and ancient populations, measured by Outgroup f_3 statistics (Source1, Source2; Mbuti). The Hui_Guizhou and Han_Guizhou groups marked by the red and blue wireframe, respectively. Lighter colors indicate more shared drift.

Supplementary Figure 2 | Results of admixture f_3 -statistics of the form $f_3(X, Target; Mbuti)$. We only plotted the f_3 -values of top 50 populations related to Hui_Guizhou people.

Supplementary Figure 3 | Admixture analysis with 56 East Asian and Western related populations assuming 2–10 ancestral populations. Our sample groups were marked by red wireframe.

- phylogeography that cuts across language families. *Am. J. Hum. Genet.* 99, 163–173. doi: 10.1016/j.ajhg.2016.05.025
- Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M. A., Järve, M., Talas, U. G., et al. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25, 459–466. doi: 10.1101/gr.186684.114
- Kong, Q. P., Sun, C., Wang, H. W., Zhao, M., Wang, W. Z., Zhong, L., et al. (2011). Large-scale mtDNA screening reveals a surprising matrilineal complexity in east Asia and its implications to the peopling of the region. *Mol. Biol. Evol.* 28, 513–522. doi: 10.1093/molbev/msq219
- Lan, Q., Chen, J., Guo, Y., Xie, T., Fang, Y., Jin, X., et al. (2018). Genetic structure and polymorphism analysis of Xinjiang Hui ethnic minority based on 21 STRs. *Mol. Biol. Rep.* 45, 99–108. doi: 10.1007/s11033-018-4143-6
- Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409–413. doi: 10.1038/nature13673
- Li, H., Cai, X., Winograd-Cort, E. R., Wen, B., Cheng, X., Qin, Z., et al. (2007). Mitochondrial DNA diversity and population differentiation in southern East Asia. *Am. J. Phys. Anthropol.* 134, 481–488. doi: 10.1002/ajpa.20690
- Li, Y. C., Ye, W. J., Jiang, C. G., Zeng, Z., Tian, J. Y., Yang, L. Q., et al. (2019). River valleys shaped the maternal genetic landscape of han chinese. *Mol. Biol. Evol.* 36, 1643–1652. doi: 10.1093/molbev/msz072
- Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., et al. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538, 201–206. doi: 10.1038/nature18964
- Mu, D. Q. (1985). Research on the origin of Hui nationality. *J. Henan Univ.* 4, 105–110.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., et al. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093. doi: 10.1534/genetics.112.145037
- Patterson, N., Price, A. L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2:e190. doi: 10.1371/journal.pgen.0020190

- Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., et al. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505, 87–91. doi: 10.1038/nature12736
- Sharma, S., Rai, E., Sharma, P., Jena, M., Singh, S., Darvishi, K., et al. (2009). The Indian origin of paternal haplogroup R1a1* substantiates the autochthonous origin of Brahmins and the caste system. *J. Hum. Genet.* 54, 47–55. doi: 10.1038/jhg.2008.2
- Sikora, M., Pitulko, V. V., Sousa, V. C., Allentoft, M. E., Vinner, L., Rasmussen, S., et al. (2019). The population history of northeastern Siberia since the Pleistocene. *Nature* 570, 182–188. doi: 10.1038/s41586-019-1279-z
- Siska, V., Jones, E. R., Jeon, S., Bhak, Y., et al. (2017). Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci. Adv.* 3:e1601877. doi: 10.1126/sciadv.1601877
- Tanaka, M., Cabrera, V. M., González, A. M., Larruga, J. M., Takeyasu, T., Fuku, N., et al. (2004). Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res.* 14, 1832–1850. doi: 10.1101/gr.2286304
- Underhill, P. A., Myres, N. M., Rootsi, S., Metspalu, M., Zhivotovskiy, L. A., King, R. J., et al. (2010). Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a. *Eur. J. Hum. Genet.* 18, 479–484. doi: 10.1038/ejhg.2009.194
- van Oven, M., and Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mut.* 30, E386–E394. doi: 10.1002/humu.20921
- van Oven, M., van den Tempel, N., and Kayser, M. (2012). A multiplex SNP assay for the dissection of human Y-chromosome haplogroup O representing the major paternal lineage in East and Southeast Asia. *J. Hum. Genet.* 57, 65–69. doi: 10.1038/jhg.2011.120
- Wang, C. C., and Li, H. (2013). Inferring human history in East Asia from Y chromosomes. *Invest. Genet.* 4:11. doi: 10.1186/2041-2223-4-11
- Wang, C. C., Lu, Y., Kang, L., Ding, H., Yan, S., Guo, J., et al. (2019). The massive assimilation of indigenous East Asian populations in the origin of Muslim Hui people inferred from paternal Y chromosome. *Am. J. Phys. Anthropol.* 169, 341–347. doi: 10.1002/ajpa.23823
- Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y., et al. (2004). Genetic evidence supports demic diffusion of Han culture. *Nature* 431, 302–305. doi: 10.1038/nature02878
- Xie, M., Song, F., Li, J., Lang, M., Luo, H., Wang, Z., et al. (2019). Genetic substructure and forensic characteristics of Chinese Hui populations using 157 Y-SNPs and 27 Y-STRs. *For. Sci. Int. Genet.* 41, 11–18. doi: 10.1016/j.fsigen.2019.03.022
- Xie, X. D., and Shan, X. M. (2002). DNA evidence of the origin of the Hui people. *Hui Res.* 47, 75–78.
- Yan, S., Wang, C. C., Zheng, H. X., Wang, W., Qin, Z. D., Wei, L. H., et al. (2014). Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLoS One* 9:e105691. doi: 10.1371/journal.pone.0105691
- Yao, H. B., Wang, C. C., Tao, X., Shang, L., Wen, S. Q., Zhu, B., et al. (2016). Genetic evidence for an East Asian origin of Chinese Muslim populations Dongxiang and Hui. *Sci. Rep.* 6:38656.
- Yao, Y. G., Kong, Q. P., Wang, C. Y., Zhu, C. L., and Zhang, Y. P. (2004). Different matrilineal contributions to genetic structure of ethnic groups in the silk road region in china. *Mol. Biol. Evol.* 21, 2265–2280. doi: 10.1093/molbev/msh238
- Yu, Y. X. (2015). Review and Reflection on the Research of Hui in Guizhou. *Soc. Sci. Ningxia* 6, 121–124.
- Zhang, F., Li, H., Huang, L., Lu, Y., and Hu, S. (2010). Similarities in patrilineal genetics between the han chinese of central china and chaoshanese in southern China. *Commun. Contemp Anthropol.* 4, e2/5-13. doi: 10.4236/coca.2010.41002

Conflict of Interest: GC was employed by company WeGene.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with one of the authors C-CW.

Copyright © 2021 Wang, Zhao, Ren, Sun, He, Guo, Zhang, Ji, Liu, Yang, Yang, Chen, Zhu, Wang, Li, Chen, Huang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.