Journal of Zhejiang University-SCIENCE B 2026 27(4):402-415
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.B2400576

Genetic structure and admixture of the Yi and Qiang in southwestern China

Jun LIU[1,2*], Li JIANG[2*], Yongqiang KONG[3], Chunnian WANG[2,4], Shuo FENG[2], Xuanzhu CHEN[3], Deqin ZHANG[5], Chuanchao WANG[6,7,8✉], Caixia LI[2✉]

[1]*School of Forensic Medicine, Shanxi Medical University, Jinzhong 030600, China*
[2]*Key Laboratory of Forensic Genetics, Beijing Engineering Research Center of Crime Scene Evidence Examination, National Engineering Laboratory for Forensic Science, Institute of Forensic Science, Beijing 100038, China*
[3]*Jiangsu International Joint Research Center of Genomics, Jiangsu Key Laboratory of Phylogenomics and Comparative Genomics, School of Life Science, Jiangsu Normal University, Xuzhou 221116, China*
[4]*School of Computer Science, Shaanxi Normal University, Xi'an 710119, China*
[5]*Institute of Forensic Medicine, Guizhou Medical University, Guiyang 550004, China*
[6]*Department of Anthropology and Ethnology, Institute of Anthropology, Fujian Provincial Key Laboratory of Philosophy and Social Sciences in Bioanthropology, School of Sociology and Anthropology, Xiamen University, Xiamen 361005, China*
[7]*State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen 361005, China*
[8]*Ministry of Education Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai 200433, China*

Abstract: The Tibetan-Yi Corridor in southwestern China is well-known for the origins, migration, and evolution of Sino-Tibetan populations. Previous genetic studies have primarily focused on Han and Tibetan populations, thereby leaving the significant genetic diversity within the Tibeto-Burman groups under-researched. In this study, to explore the genetic structure and admixture history of Tibeto-Burman populations in southwestern China, we sequenced the human genomes of 100 individuals from the Qiang and Yi ethnic groups in Sichuan Province. These populations were found to have the closest genetic affinity with nearby Tibeto-Burman-speaking Tujia and Tibetan populations. The Qiang share more allele sites with northern Altaic-speaking populations, while the Yi have closer genetic relationships with southern Hmong-Mien populations. The dominant ancestry of the Yi and Qiang derived from Neolithic millet agriculturalists in the Yellow River Basin, with a smaller proportion from Neolithic coastal populations in southern China, supporting the hypothesis of a northern origin of Sino-Tibetan populations. The Yi have more southern genetic components than the Qiang, reflecting the differential genetic influences of southeastern coastal populations on these groups. In summary, this study elucidates the fine-scale genetic structure of Tibeto-Burman populations and their genetic relationships with other Chinese populations, laying the foundation for forensic genetic research in East Asian populations.

Key words: Tibeto-Burman; Genetic diversity; Genetic structure; Population genetic relationships

1 Introduction

Located in the eastern part of the Eurasian continent, Eastern Asia and its adjacent vast hinterland are bordered by the Ural Mountains in the west, the Qinghai-Xizang Plateau in the southwest, and the Bering Strait in the northeast, with a long history and complex development dynamics (Stoneking and Delfin, 2010). As an important hub for ancient human migrations, a site of numerous ancient human fossils, one of the centers of agricultural origins, and a gathering place for different linguistic groups, East Asia exemplifies the complex historical dynamics of population movements, biological evolution, and sociocultural exchanges spanning millennia (Bae et al., 2017; Wang CC et al., 2021). As such, this region is crucial for studying

---

✉ Caixia LI, licaixia@tsinghua.org.cn
Chuanchao WANG, chuanchaowang@fudan.edu.cn
[*] The two authors contributed equally to this work
Chuanchao WANG, https://orcid.org/0000-0003-4630-5232
Jun LIU, https://orcid.org/0009-0000-4891-2138

Received Nov. 18, 2024; Revision accepted Apr. 9, 2025;
Crosschecked Mar. 13, 2026; Published online Mar. 28, 2026

human origins (HOs) and ethnic evolution. China, the largest country in East Asia, hosts populations from seven language families: Sino-Tibetan, Tai-Kadai, Hmong-Mien, Austroasiatic, Altaic, Indo-European, and Austronesian, reflecting a complex genetic diversity (Cai, 2009). The Sino-Tibetan language family comprises the Chinese and Tibeto-Burman groups. The Han Chinese group is spread across various provinces, while the Tibeto-Burman groups mostly reside in the river valleys of eastern Xizang and western Sichuan in southwestern China, known as the Tibeto-Yi Corridor. This region has been a focal point for ethnological and linguistic studies and is key to understanding the origins, migration, and evolution of Sino-Tibetan populations.

Previous linguistic and archaeological evidence has shown that the origin and expansion of Sino-Tibetan populations are related to the Neolithic Yangshao and Majiayao cultures, as well as the rapid development of millet agriculture (Sagart et al., 2019; Zhang et al., 2019). Modern Tibeto-Burman and Han populations shared a common ancestry with ancient populations in the middle and upper reaches of the Yellow River Basin (Wang CC et al., 2021). The differentiation between the Han and Tibeto-Burman groups began around 6000 years ago during the Yangshao culture period (Victor, 1998). Many whole-genome studies in the Tibeto-Yi Corridor have focused on Tibetan and Han populations. It was revealed that populations in the Tibeto-Yi Corridor are a mix of Tibetans from the Qinghai-Xizang Plateau and East Asian populations from the surrounding low-altitude areas (Yao et al., 2017). From the paternal Y-chromosome perspective, populations carrying the haplogroup D-M174 from the Late Paleolithic and those carrying the haplogroup O2a2b1a1a-F5 from the middle and upper reaches of the Yellow River Basin underwent population expansion and migrated through the Tibeto-Yi Corridor to the Himalayas and Yunnan, mixing to form modern Tibeto-Burman groups (Wen et al., 2004; Shi et al., 2008; Yu and Li, 2021).

The Qiang people originated from primitive nomadic tribes in northwestern China, with a history dating back nearly 3000 years to the Shang Dynasty, as evidenced by oracle bone inscriptions. Historically, the Qiang primarily lived in the mountainous regions near the Min River. However, due to natural disasters and wars, they began to migrate and merged with the Han population during the Ming and Qing Dynasties, with some evolving into the present-day Qiang (Wang, 2013). The Qiang mainly reside in the Aba Tibetan and Qiang Autonomous Prefecture in Sichuan Province, making them the third-largest minority in Sichuan, with a population of approximately 300 000. They speak Qiangic, a language belonging to the Qiangic branch of the Tibeto-Burman family within the Sino-Tibetan language family. Previous genetic studies have shown that Qiang's maternal gene pool, based on mitochondrial DNA (mtDNA), has significant contributions from northern populations, with prevalent haplogroups A, D, G, and M8 (Torroni et al., 1994; Yao et al., 2002; Zhao et al., 2009; Qin et al., 2010; Qi et al., 2013; Wang et al., 2014). The Qiang have the highest Y-chromosome short tandem repeat (STR) diversity among the Sino-Tibetan groups in eastern Xizang, which may result from the Sino-Tibetan expansion (Kang et al., 2012). Additionally, archaeological findings indicate that the Yangshao culture, which originated around 7000 years ago, emerged in regions inhabited by the Qiang (Zhao et al., 2011).

The Yi people, China's sixth-largest ethnic minority, primarily live between the highlands and coastal hills in the southwestern regions of Sichuan, Yunnan, and Guizhou. They speak their own language, Yi, which belongs to the Yi branch of the Tibeto-Burman family within the Sino-Tibetan language family. Historical records in both Chinese and Yi texts indicate that the Yi people had close ties with the ancient Qiang of western China, suggesting that the Yi mainly originated from the ancient Qiang. Studies based on X-chromosome STRs (X-STRs) have shown that the Yi have a close genetic distance to most Han populations; moreover, geographically proximate groups exhibit genetic homogeneity on the X-chromosome (Yao et al., 2017). However, population genetic analyses based on Y-STRs indicate diversity in Y-chromosome lineages among the Yi. Despite sharing similar ethnic cultures or geographical origins, the expected genetic homogeneity among the Yi cannot be fully observed (Fan et al., 2019). Genome-wide single-nucleotide polymorphism (SNP) data have revealed that the Yi share more alleles with ancient northern millet agricultural populations in China and have also been influenced by ancient southern populations (Sun et al., 2024). Existing archaeological data support the hypothesis that the Yi originated from the north, migrated along the Tibetan-Yi Corridor, and then mixed with southern residents (Jeong et al., 2014, 2017; Liu et al., 2022).

Previous genetic analyses have mainly focused on the Han and Tibetan populations, while research on the genetic diversity within the Tibeto-Burman groups, particularly the Qiang and Yi, and their internal genetic relationships remains insufficient. Therefore, in this study, we collected 100 blood samples from members of the Qiang and Yi groups from Sichuan Province, Southwest China, for DNA extraction followed by 5× depth whole-genome sequencing. To explore the genetic relationships between the Qiang and Yi within the Tibeto-Burman family and the genetic substructure within East Asia, we conducted principal component analysis (PCA), population ancestry component analysis, and *f*-statistics, *qpWave*, *qpAdm*, and TreeMix analyses. This study aims to provide in-depth information on admixture events among ethnic minorities in southwestern China.

2 Materials and methods

2.1 Sample collection and genotyping

We collected a total of 100 blood samples from the southwestern region of China, including 50 samples from the Qiang ethnic group in the northwestern Sichuan Province and 50 samples from the Yi ethnic group in the southwestern Sichuan Province. We performed whole-genome sequencing on the 100 samples at 5× depth using the Illumina Novaseq 6000, yielding 10 000 SNPs after genotype imputation (supplementary materials and methods). The variation data are deposited in the Genome Variation Map (GVM) (Li CP et al., 2021) in the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, and China National Center for Bioinformation (CNCB-NGDC Members and Partners, 2023) under accession number GVM000574.

2.2 Data processing and merging

We initially conducted quality control on the whole-genome sequencing data using Plink v1.9 (Chang et al., 2015), filtering SNPs with a detection rate below 5%, Hardy-Weinberg equilibrium (HWE) P value$<1\times10^{-6}$, and minor allele frequency (MAF)$<$ 0.02, yielding 6 812 907 SNPs. Kinship coefficients were calculated using KING v2.2.7 (Manichaikul et al., 2010). We removed three samples with a kinship within the third degree (Φ ranging from 0.0442 to 0.0884). Next, we merged the dataset from this study with public datasets using the mergeit tool of EIGENSOFT v6.1.4 (Price et al., 2006). The strandcheck parameter setting was Yes to remove the AT/CG SNPs, while all other parameters remained at their default settings. The public datasets primarily included the 1240K dataset and the HO dataset from the David Reich Lab (https://reich.hms.harvard.edu/datasets), along with other published data on modern and ancient East Asian populations (Patterson et al., 2012; Bergström et al., 2020; Ning et al., 2020; Yang et al., 2020; Wang CC et al., 2021). These sources cover populations from a partial region of mainland Southeast Asia (MSEA) (Kutanan et al., 2021), ancient individuals from the Qinghai-Xizang Plateau (Wang et al., 2023), and ancient populations from Sichuan and Yunnan Provinces (Tao et al., 2023). Specifically, the 1240K dataset includes both ancient and modern individuals, with data derived from either shotgun sequencing or in-solution target capture at 1 233 013 sites, as provided by the David Reich Laboratory. In contrast, the HO dataset is based on 597 569 SNPs genotyped using the Affymetrix HOs array. While the 1240K dataset offers higher SNP density, the HO dataset encompasses a wider range of present-day populations, rendering it particularly useful for studying modern genetic diversity. Thus, we generated two merged datasets for subsequent analyses, encompassing 140 313 SNPs from the merged HO dataset and 299 018 SNPs from the merged 1240K dataset. We used the former dataset in PCA, ADMIXTURE, F_{st}, and TreeMix analyses and the latter dataset in *f*-statistics, *qpWave*, and *qpAdm* analyses. Detailed information on the reference populations is listed in Table S1.

2.3 Principal component analysis

To explore the genetic structure of East Asian populations and investigate the clustering patterns between our study populations and reference populations, this study used PCA, which was conducted using the SmartPCA program in the EIGENSOFT software package (Price et al., 2006). On the basis of the HO dataset, we visualized clustering patterns across East Asia. The parameters were set as numoutlieriter=0 and lsqproject=YES. Modern populations were first plotted on a plane based on principal component 1 (PC1) and PC2 coordinates, followed by the projection of 96 ancient populations onto the same two-dimensional area.

2.4 ADMIXTURE

To account for the strong linkage disequilibrium of SNPs, we used Plink v1.9 (Chang et al., 2015) to prune them with the parameters "-indep-pairwise 200 25 0.4", resulting in the removal of highly linked SNPs and the reduction of the dataset to 28 481 SNPs. We conducted genetic ancestry and admixture analysis of East Asian populations using ADMIXTURE v1.3.0 (Alexander et al., 2009), setting the number of ancestral components (*K*) from 2 to 15 with random cycles. Ancestry Painter v5 (Feng et al., 2018) was utilized to visualize the results. We determined the optimal *K* value by analyzing the cross-validation error rate and examining the population and individual ancestry components for different *K* values.

2.5 Pairwise F_{st} and identical-by-descent

We measured genetic distances and relationships between the Qiang, Yi, and other modern East Asian populations using the SmartPCA program in EIGENSOFT (Price et al., 2006), setting the parameter fstonly as Yes. We calculated the lengths of pairwise shared identical-by-descent (IBD) segments using the Refined-IBD software (17Jan20.102.jar) (Browning and Browning, 2013), with the parameter set to length=0.1.

2.6 TreeMix

We inferred population separation and admixture patterns between the study populations and other East Asian reference populations using TreeMix v1.13 (Pickrell and Pritchard, 2012). To explore the genetic relationships between the Sino-Tibetan populations in Sichuan and 33 Chinese populations, we constructed likelihood phylogenetic trees with migration events ranging from 0 to 1.

2.7 *f*-statistics

To investigate genetic affinities and identify potential ancestral sources, we performed outgroup-f_3 and admixture-f_3 statistics using the *qp*3Pop program in the ADMIXTOOLS software package (Patterson et al., 2012; Kim et al., 2017). To explore substructure and gene flow directions within Sichuan, we merged the study population data with the 1240K database. We conducted four-population (f_4) statistics using the *qp*Dstat program in the ADMIXTOOLS software package. The underlying principles are detailed in the supplementary materials and methods.

2.8 *qpWave* and *qpAdm*

To assess the genetic homogeneity between the Qiang and Yi, we conducted pairwise *qpWave* tests using ADMIXTOOLS (Patterson et al., 2012). In addition, we used the *qpAdm* program (Patterson et al., 2012) to model the ancestry of the study populations and determine the precise admixture proportions between ancestral populations. Nine groups (Mbuti, Russia_Kostenki14, Malaysia_LN, Russia_DevilsCave_N.SG, Mixe, Russia_MA1_HG.SG, Russia_Ust_Ishim_HG, Papuan, and Australian) were selected to serve as outgroups.

3 Results

3.1 Population genetic structure

PCA revealed significant genetic diversity within East Asian populations, with genetic substructure corresponding to geographic and linguistic divisions (Fig. 1). We identified three primary clusters: (1) the Altaic and Sino-Tibetan cluster, (2) the southern Tai-Kadai and Austroasiatic cluster, and (3) the southern Hmong-Mien and Austronesian cluster. Our studied populations could be positioned between the high-altitude Tibeto-Burman and low-altitude southern populations, indicating a close genetic relationship between modern Tibeto-Burman groups and ancient Yellow River Basin farmers.

In the PCA plot shown in Fig. 1a, many individuals cluster closely together, resulting in a blurred view. To address this issue, we excluded ancient populations and conducted a clearer PCA analysis (Fig. 1b). The study populations appear to be between the Tibeto-Burman and southern groups, including Austroasiatic, Austronesian, Tai-Kadai, and Hmong-Mien-related populations. Some samples overlap with previously published Tibeto-Burman populations, forming an elongated genetic cluster. This suggests a genetic substructure within Sichuan, with the studied populations likely resulting from an admixture between the Tibeto-Burman-related and southern groups.

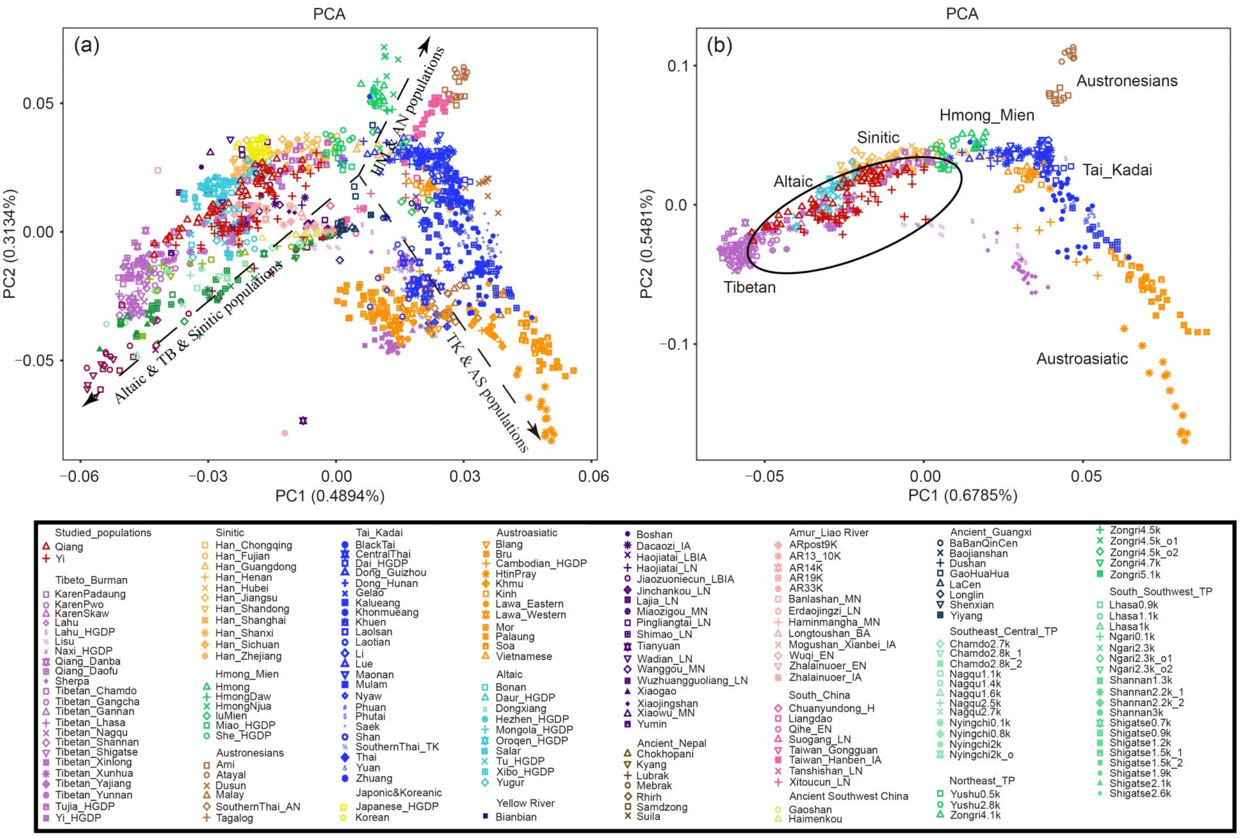

Fig. 1 Principal component analysis (PCA) of the study populations and other ancient and modern East Asian populations. (a) Results of PCA for 1386 East Asian individuals from 192 ancient and modern populations. (b) Results of PCA for 588 East Asian individuals from 53 modern populations. The study populations are encircled. TB: Tibeto-Burman; HM: Hmong-Mien; AN: Austronesian; TK: Tai-Kadai; AS: Austroasiatic.

We further conducted an unsupervised ADMIXTURE clustering analysis and found the lowest cross-validation error at *K*=6. The genetic components of the Qiang and Yi populations were similar, showing a high proportion of components from the Qinghai-Xizang Plateau (blue) and ancient northern Yellow River Basin farmers (tan). Modern Qiang and Yi populations appeared to have close genetic relationships with both Qinghai-Xizang Plateau populations and ancient Yellow River Basin farmers (Fig. 2). Additionally, both populations have been influenced by southern groups such as Hmong-Mien, Austronesian, and Tai-Kadai speakers. When comparing the genetic components of the Qiang and Yi from the Sichuan region, we found that the Qiang have more ancestral components from Yellow River Basin farmers, while the Yi have more ancestral components from Qinghai-Xizang Plateau populations. Furthermore, the Yi have a higher proportion of southern population components, showing closer genetic ties to groups like the Lahu.

3.2 Population relationships with modern East Asians

Using the Mbuti population as an outgroup, we performed outgroup-f_3 (Qiang/Yi, *X*; Mbuti) statistical analysis for the studied Yi and Qiang populations, revealing that the Qiang share more alleles and have closer genetic affinities with Chinese-speaking Han and other Sino-Tibetan populations (Fig. 3a). Meanwhile, the Yi share more alleles and have closer genetic affinities with the Han, Yi_human genome diversity project (HGDP), Tujia, and She populations (Fig. 3b). With *X* representing Tibeto-Burman populations, the Yi samples share more alleles and exhibit closer genetic affinity with Yi_HGDP, followed by the Tujia and Qiang. Conversely, the Qiang share more alleles and exhibit closer genetic affinity with the Tujia, followed by the Qiang_Danba and Tibetan (Fig. S1).

To determine the genetic affinities and population differentiation, we analyzed the genetic differentiation

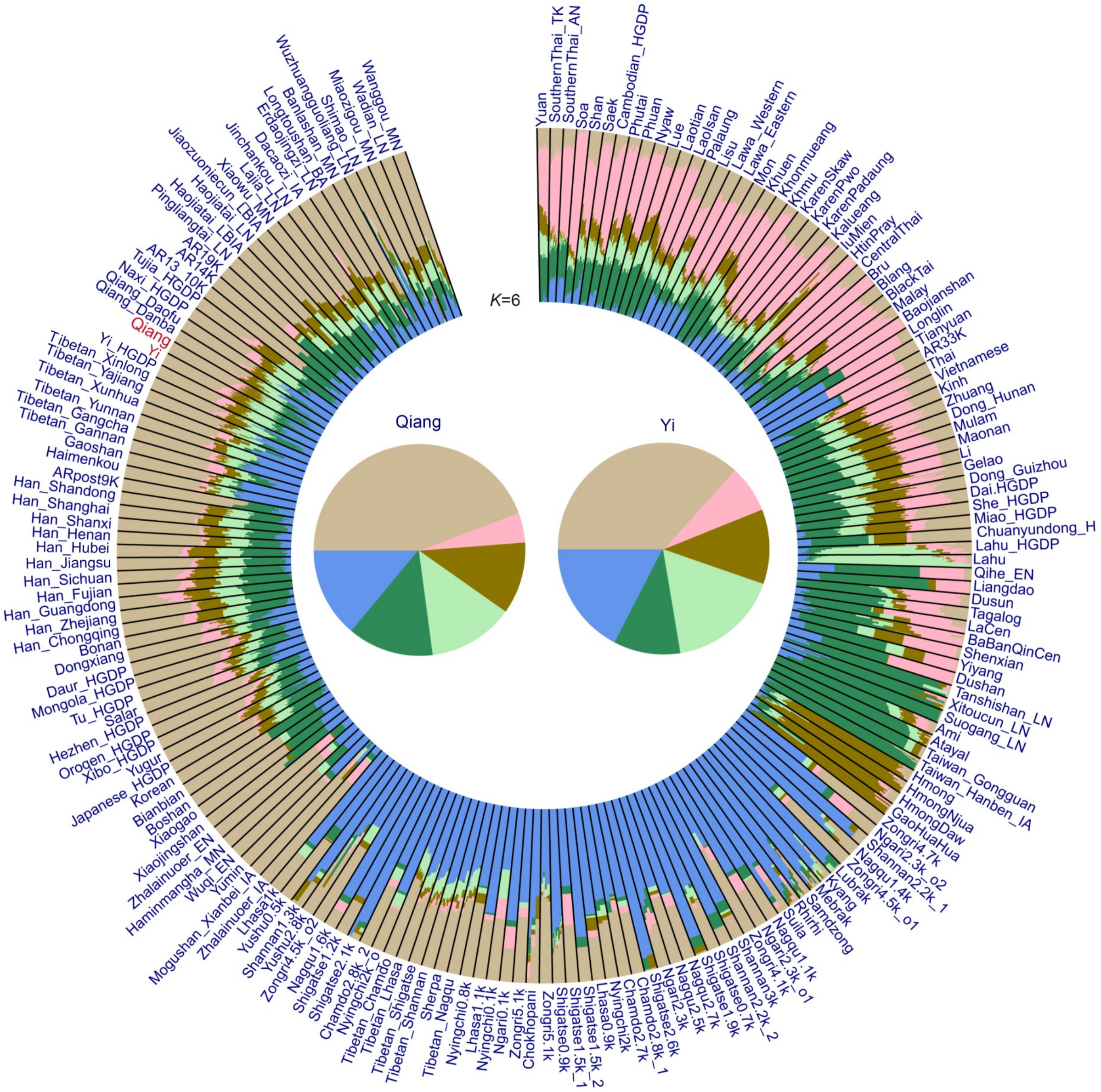

Fig. 2 ADMIXTURE results of the study populations and reference populations. The outer circle shows the ADMIXTURE results chosen based on the minimum cross-validation error of six ancestral populations (*K*=6).

index of the studied populations and surrounding groups, calculating their pairwise F_{st} distances. We found that the Qiang and Yi populations are more closely related to the Han and Tibetan populations (Fig. S2a). Population comparisons based on F_{st} genetic differences showed that the Qiang have closer genetic ties to the nearby Yugur and Mongolian populations, while the Yi are more closely related to the Xinlong Tibetans and Yi_HGDP. Pairwise IBD shared segment lengths further revealed close genetic relationships between the Qiang and geographically proximate Tibetan- and Mongolic-speaking Tu populations (Fig. S2b). Similarly, the Yi exhibit close genetic relationships with the Tibetans, Naxi, and Yi_HGDP.

Both the outgroup-f_3 statistics and F_{st} values suggested that the Qiang have closer genetic relationships with geographically proximate northern Tibeto-Burman populations, while the Yi exhibit closer genetic affinity with geographically proximate southern Tibeto-Burman populations (Figs. S3 and S4). In the maximum likelihood tree inferred by TreeMix, the Yi, Qiang, Naxi, and Tibetan cluster together. Moreover, we detected potential gene flow from the Lahu to the Yi (Fig. S5).

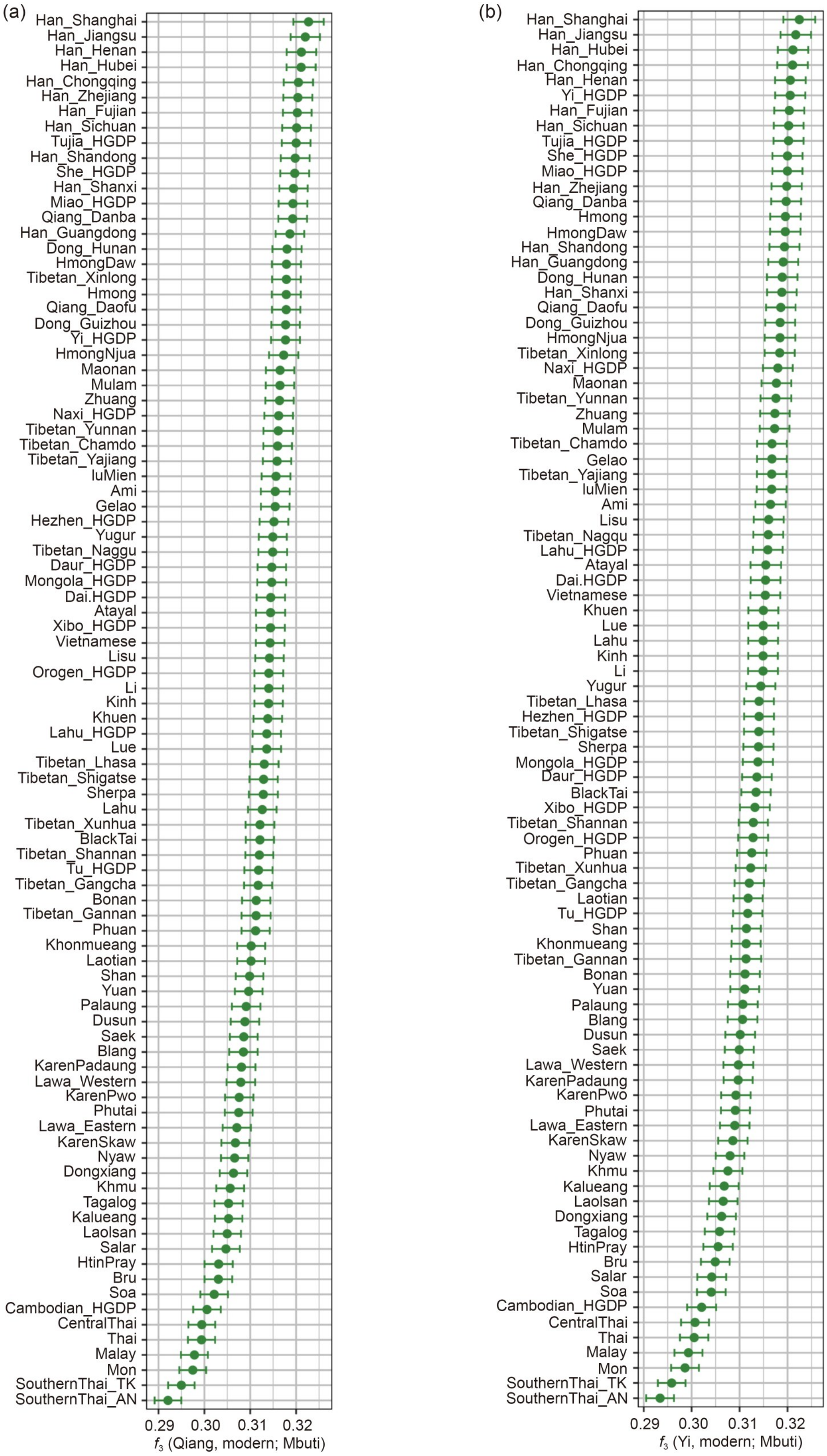

Fig. 3 Shared genetic drift based on outgroup-f_3 statistics analysis. (a, b) Outgroup-f_3 statistics (Qiang/Yi, modern East Asians; Mbuti) for the studied Qiang (a) and Yi (b) populations. A larger f_3 value suggests more gene flow, indicating closer genetic affinity.

To examine the genetic relationships between the Qiang, Yi, and modern East Asians, we further conducted f_4 statistics (Qiang/Yi, modern Asians; Mbuti). The results indicated that the Yi, compared to the Qiang, share more alleles with Tibeto-Burman populations and have received additional gene flow from Austroasiatic populations. In contrast, the Qiang share more alleles and exhibit closer genetic affinity with Altaic populations compared to the Yi (Table S2).

3.3 Population relationships with ancient East Asians

The outgroup-f_3 statistics revealed that the Yi and Qiang populations share more alleles and exhibit closer genetic affinity with Late Neolithic to Bronze Age agricultural populations from southwestern China and Neolithic to Iron Age agricultural populations from the Yellow River Basin (Fig. S6).

We used f_4 statistics (Qiang/Yi, ancient Asians; Mbuti) to examine the asymmetric shared ancestry signals between the studied populations and ancient East Asians. The results showed that the Yi share more alleles with ancient populations from Guangxi and the Qinghai-Xizang Plateau compared to the Qiang. Conversely, the Qiang share more alleles with ancient populations from the Amur River Basin compared to the Yi (Table S3).

The symmetric f_4 statistics (ref1, ref2, Qiang/Yi; Mbuti) indicated that both the Qiang and Yi share more alleles with Yellow River Basin farmers and ancient populations from Gaoshan City and Haimenkou (Fig. S7). The asymmetric f_4 statistics (ref1, Qiang/Yi, ref2; Mbuti) obtained statistically significant negative Z values, suggesting closer genetic links between the Qiang and Yi and Yellow River Basin farmers compared to other ancient populations. In addition, there is evidence of gene flow from southern ancient populations in Taiwan relative to their potential northern ancestral sources, indicating that the Qiang and Yi were also influenced by southern ancient groups (Fig. 4).

To explore the genetic homogeneity, we also conducted pairwise *qpWave* analysis between the studied

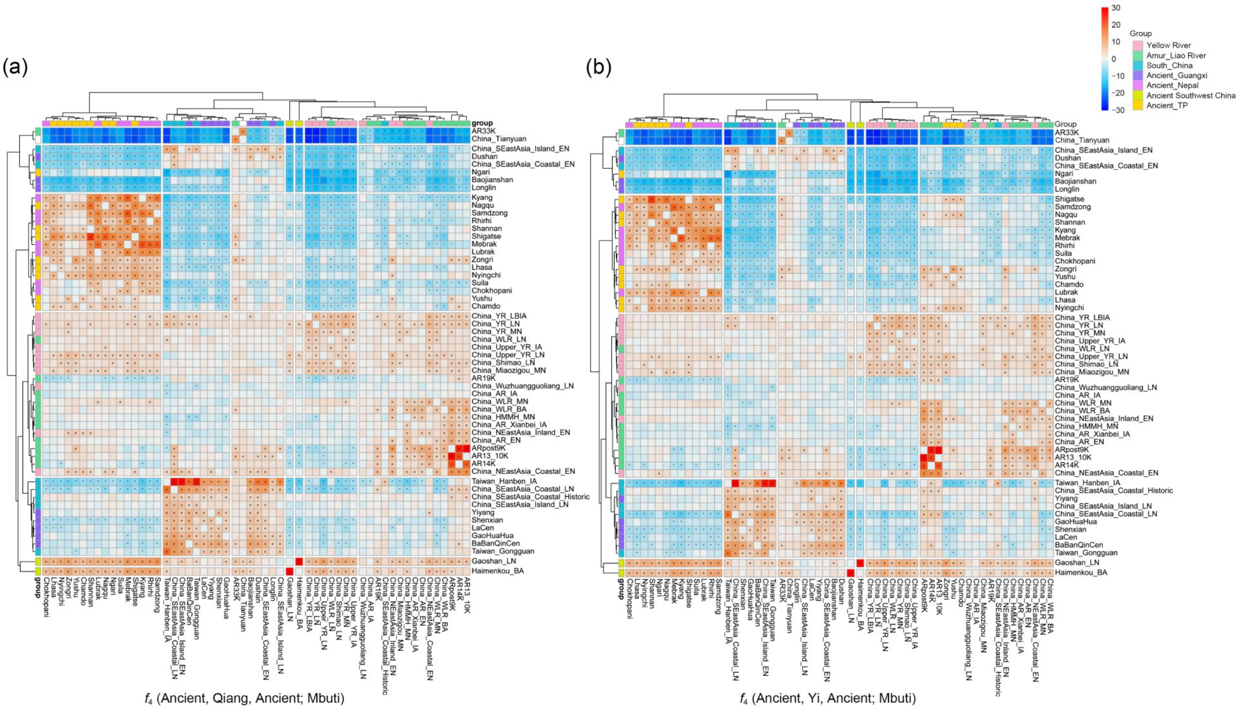

Fig. 4 f_4 statistics results. (a, b) f_4 statistics (ref1, Qiang/Yi, ref2; Mbuti) estimates genetic affinities between the studied Qiang (a) and Yi (b) populations and ancient populations. Statistically significant population pairs are marked with an asterisk (*). Red indicates a closer relationship between the study population and the right-side group compared to the bottom group. Negative values in blue indicate a closer relationship between the study population and the bottom group compared to the right-side group. TP: Qinghai-Xizang Plateau; YR: Yellow River; EN: Early Neolithic; MN: Middle Neolithic; LN: Late Neolithic; BA: Bronze Age; IA: Iron Age; LBIA: Late Bronze and Iron Ages; WLR: West Liao River; AR: Amur River; HMMH: Haminmangha.

and ancient East Asian populations. The results indicated a signal of homogeneity between the Qiang and ancient northern populations. Additionally, genetic heterogeneity was observed between the Yi, Qiang and geographically proximate Late Neolithic to Bronze Age agricultural populations from the southwestern region, suggesting different population admixture histories (Fig. S8).

3.4 Admixture model of the studied Qiang and Yi populations

We calculated the admixture-f_3 statistics focusing on Yi and Qiang populations, considering all modern and ancient populations from East Asia and Southeast Asia as potential sources (Tables S4 and S5). When the Sichuan Yi were set as the target population, the Z-value results indicated that the Sichuan Yi likely formed through admixture between the Tibetan groups and southern Tai-Kadai-, Austronesian-, and Austroasiatic-speaking groups. Moreover, evidence of admixture in the Qiang was found, with one source being Tibetan groups and the other from southern East Asians, including Tai-Kadai-, Austronesian-, and Austroasiatic-speaking groups.

Next, we used *qpAdm* to model the genetic formation of the Yi and Qiang populations. For the two-way admixture model, we selected ancient populations from the Neolithic Yellow River Basin (China_YR_MN/China_Upper_YR_LN) as northern sources and Iron Age Hanben populations from Taiwan, China (Taiwan_Hanben_IA), along with historical populations (BaBanQinCen) as potential southern sources. As shown in (Fig. 5 and Table S6), the Yi have a northern ancestry proportion ranging from 59.6% to 77.1%, with the highest southern ancestry proportion at 40.4% when using Taiwan_Hanben_IA as the source. Meanwhile, the Qiang have a northern ancestry proportion ranging from 75.3% to 88.7%, with the highest southern ancestry proportion at 24.7% when using Taiwan_Hanben_IA as the southern source. Overall, both studied populations exhibit a higher northern ancestry proportion than southern ancestry. Using consistent northern and southern sources, the Yi show a higher southern ancestry proportion, while the Qiang show a higher northern ancestry proportion. These results suggest that the admixture events occurred at different historical periods, aligning with the historical records of southward migration.

4 Discussion

East Asia, home to a quarter of the global population, exhibits high genetic diversity. While research on the genetic structure and population admixture history of East Asians has significantly intensified over the past 20 years (Stoneking and Delfin, 2010; Chiang et al., 2018), representative populations remain less studied compared to European populations. Understanding ethnic and linguistic genetic diversity can advance molecular anthropology, medicine, and population evolution studies (Liu et al., 2020; Kutanan et al., 2021). Herein, we sequenced whole genomes from 100 individuals of the Qiang and Yi groups from Sichuan, China. After excluding three samples due to kinship relationships, analysis was performed on 97 individuals. Combined with previously published data (Gan et al., 2008; He et al., 2021; Li L et al., 2021; Liu et al., 2021; Wang CC et al., 2021; Wang MG et al., 2021; Yao et al., 2021), we explored the genetic structure

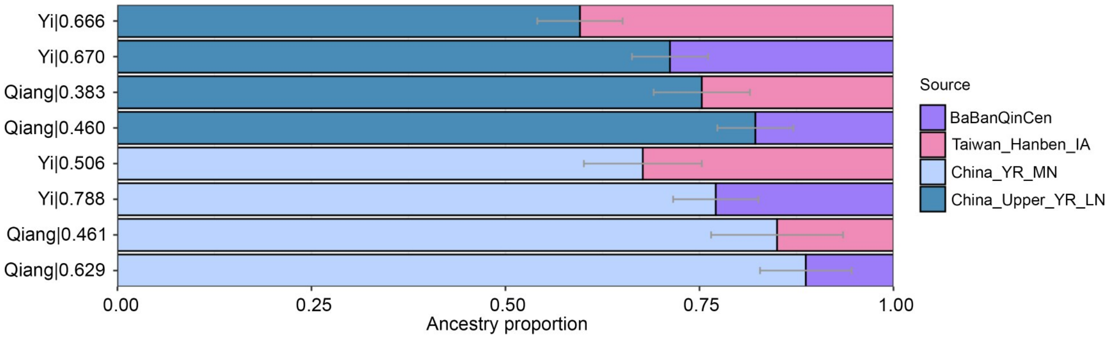

Fig. 5 Two-way admixture model of ancestral proportions in the study populations based on *qpAdm*. The proportions of different ancestries in the Qiang and Yi are estimated by *qpAdm*. IA: Iron Age; YR: Yellow River; MN: Middle Neolithic; LN: Late Neolithic.

and admixture history of these populations in southwestern China.

We found that genetic variation largely follows a geographical pattern, with the Qiang clustering closely with nearby Yi, Tibetan, and Han populations. Previous analyses have shown that modern Tibetans' primary ancestors were millet farmers from the middle and upper Yellow River Basin in northern China (Lu et al., 2016; Wang CC et al., 2021). Our results support the theory that the dominant ancestry of the Qiang people derived from northern East Asia, consistent with the shared origin of the Qiang, Tibetan, and Han populations from the Yellow River Basin. Previous mtDNA studies of Qiang populations have further confirmed this observation, demonstrating substantial contributions from northern haplogroups, including A, C, D, and G, to their maternal gene pool (Wang et al., 2014). Moreover, *qpWave* analyses revealed genetic homogeneity between the Qiang population and ancient northern groups, a finding that aligns with paleogenomic studies of the Qinghai-Xizang Plateau populations. As such, the Qiang can be fully modeled as descendant populations of Iron Age agriculturalists from the upper Yellow River region (Wang et al., 2023). We also found close genetic relationships between the Qiang and historical Xiongnu and Mongolian populations, reflecting ancient contributions from populations in China's northwestern regions (Zhao et al., 2011; Li et al., 2017). The two-way admixture model based on *qpAdm* confirmed the north-south admixture in the Qiang's formation, with the primary ancestral component from Neolithic millet agriculturalists in the Yellow River Basin and additional influences from historical populations in Guangxi and related southern Tai-Kadai and Austronesian groups.

The PCA and admixture results yielded strong connections between the Yi and ancient populations from the Yellow River Basin. The f_4 statistics revealed that, compared to the Qiang, the Yi have been more influenced by the southern groups such as Austroasiatic, Tai-Kadai, and coastal southern Chinese populations; therefore, the Yi's ancestral components originate from northern ancient populations represented by the Yellow River Basin and southern ancient populations represented by Guangxi. Furthermore, mtDNA studies of the Yi population revealed that their maternal gene pool is composed of both northern-prevalent haplogroups (A and D) and southern-prevalent haplogroups (B, F, and M7b), reflecting a dual ancestry from northern and southern East Asia (Li et al., 2007). This supports previous archaeological data and suggests that the Yi's ancestors migrated southward through the Tibetan-Yi Corridor during historical periods due to famine and war, mixing with southern indigenous residents (Jeong et al., 2014, 2017; Wang et al., 2014; Liu et al., 2022).

Previous research has shown strong genetic connections between the Qiang, Yi, and Han populations, indicating their significant admixture. However, our study revealed genetic differences among Tibeto-Burman groups, highlighting a substantial genetic substructure and heterogeneity between the Qiang and Yi populations in Sichuan. The *qpAdm* admixture model indicated that the Sichuan Yi have a higher proportion of southern genetic components compared to the Qiang, suggesting a longer history of admixture with southern groups. Similarly, a study of multiple ethnic groups in the Tibeto-Yi Corridor revealed that three populations inhabiting the southernmost region of the Tibeto-Yi Corridor, the Lahu, Jino, and Hani, harbor over 50% of Dai-related ancestral components, highlighting their closer genetic affinity to southern East Asian populations (Zhang et al., 2022). Conversely, the Qiang have a higher proportion of northern genetic components, suggesting a more conservative lifestyle in recent years. The genetic differences between the Yi and Qiang populations of Sichuan are mainly due to varying proportions of southern genetic influences, pointing to a genetic gradient of southeast coastal population expansion into western China. The results overall indicate that Sichuan is a complex region influenced by different populations from various areas, each contributing different proportions of genetic components.

The origin of Sino-Tibetan populations remains intensely debated, with three prevalent hypotheses that are partially supported by linguistic and historical findings: the northern India origin hypothesis (Blench, 2014), the southwestern China/Tibeto-Yi Corridor hypothesis (Sagart et al., 2019), and the northern China origin hypothesis (Aikhenvald et al., 2001; Zhang et al., 2019). Our findings support the northern China origin hypothesis, which is consistent with previous genome-wide analyses of the Qiang and Yi. Previous studies on Y-chromosome haplogroups D-M174 and O-M117 have also suggested that Tibeto-Burman speakers are a mix of Neolithic migrants from the northern Yellow River Basin and southern indigenous residents (Wen et al., 2004; Shi et al., 2008; Yu and Li, 2021).

Moreover, modern southwestern Tibeto-Burman populations (Qiang and Yi) retain significant genetic components related to ancient populations from Gaoshan City and Haimenkou, indicating a stable genetic structure in the region for at least three to four thousand years (Tao et al., 2023). Genetic analysis uncovers that the Qiang and Yi are not genetically homogeneous with geographically proximate Late Neolithic to Bronze Age agricultural populations in southwestern China. Genetic, historical, and linguistic evidence collectively suggest that, despite a common genetic origin, Tibeto-Burman populations exhibited different genetic admixture patterns after separating from the Han (Wang CC et al., 2021).

5 Conclusions

To elucidate the genetic relationships between Tibeto-Burman and other Chinese populations, this study systematically examines the genetic structure of the Qiang and Yi, laying a foundation for forensic genetic studies of East Asian populations. We analyzed high-density SNP data from 50 Qiang and 50 Yi individuals in southwestern China, revealing that the Qiang and Yi have the closest genetic relationships with nearby Tibeto-Burman Tujia and Tibetan populations. Further evidence from *f*-statistics and *qpAdm* admixture models indicated that most ancestors of the Qiang and Yi were Neolithic millet agriculturalists from the Yellow River Basin, with the remaining ancestry from ancient coastal southern populations. The northern genetic components of these populations link to the Neolithic culture of millet agriculturalists in the Yellow River Basin, supporting the hypothesis that Sino-Tibetan populations originated from northern China. Our findings also highlight varying proportions of genetic influence from southeastern coastal populations on different groups. The 5× sequencing data may introduce potential biases, although we have made efforts to minimize this issue through imputation and stringent quality control measures. Future research should collect high-coverage sequencing data from underrepresented populations to fully confirm our findings and gain new insights into the formation of Tibeto-Burman populations, further exploring fine-scale population genetic structures and enhancing the application value of whole-genome sequencing in forensic science.

Data availability statement

The genome-wide variation data were sourced from the public dataset of Allen Ancient DNA Resource (AADR) (Mallick et al., 2024). The newly generated variation data reported in this paper have been submitted to the National Genomics Data Center (NGDC; https://ngdc.cncb.ac.cn) under accession number GVM000574.

Acknowledgments

This study was supported by the National Key R&D Program of China (No. 2022YFC3341004), the National Natural Science Foundation of China (Nos. 82171870, T2425014, and 32270667), the Natural Science Foundation of Fujian Province of China (No. 2023J06013), and the Major Project of the National Social Science Foundation of China (No. 21&ZD285).

We thank Prof. Wibhu KUTANAN (Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand), Prof. Mark STONEKING (Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany), and Dr. Dang LIU (Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany) for sharing genome-wide SNP data from Vietnam, Thailand, and Laos. We also thank all volunteers who participated in this project.

Author contributions

Jun LIU, Li JIANG, Chuanchao WANG, and Caixia LI wrote the manuscript, and conceived and designed the study. Jun LIU and Li JIANG performed most of the experiments with help from Xuanzhu CHEN and Deqin ZHANG. Jun LIU, Yongqiang KONG, Chunnian WANG, and Shuo FENG contributed to data analyses. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Jun LIU, Li JIANG, Yongqiang KONG, Chunnian WANG, Shuo FENG, Xuanzhu CHEN, Deqin ZHANG, Chuanchao WANG, and Caixia LI declare that they have no conflicts of interest.

The Ethics Committee of the Ministry of Public Security Identification Center reviewed and approved this study (Approval No. 2021-006). All participants provided written informed consent to participate.

Declaration on the use of generative AI tools

During the preparation of this work, the authors used ChatGPT in order to improve language and readability and to check for grammatical errors. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Aikhenvald AY, Dixon RMW, 2001. Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics. Oxford University Press, Oxford, UK.
https://doi.org/10.1093/oso/9780198299813.001.0001

Alexander DH, Novembre J, Lange K, 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, 19(9):1655-1664.
https://doi.org/10.1101/gr.094052.109

Bae CJ, Douka K, Petraglia MD, 2017. On the origin of modern humans: Asian perspectives. *Science*, 358(6368):eaai9067.
https://doi.org/10.1126/science.aai9067

Bergström A, McCarthy SA, Hui RY, et al., 2020. Insights into human genetic variation and population history from 929 diverse genomes. *Science*, 367(6484):eaay5012.
https://doi.org/10.1126/science.aay5012

Blench R, 2014. Rethinking Sino-Tibetan phylogeny from the perspective of North East Indian languages. Thomas OS, Nathan WH (Eds.), Trans-Himalayan Linguistics. Walter de Gruyter GmbH, Berlin, p.71-104.
https://doi.org/10.1515/9783110310832

Browning BL, Browning SR, 2013. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*, 194(2):459-471.
https://doi.org/10.1534/genetics.113.150029

Cai XY, 2009. Early Human Entry Into East Asia and the Process of Characteristic Formation in East Asian Populations as Revealed by the Y Chromosome. PhD Dissemination, Fudan University, Shanghai, China (in Chinese).
https://doi.org/10.1038/s41422-020-0322-9

Chang CC, Chow CC, Tellier LC, et al., 2015. Second-generation plink: rising to the challenge of larger and richer datasets. *GigaScience*, 4:7.
https://doi.org/10.1186/s13742-015-0047-8

Chiang CWK, Mangul S, Robles C, et al., 2018. A comprehensive map of genetic variation in the world's largest ethnic group—Han Chinese. *Mol Biol Evol*, 35(11):2736-2750.
https://doi.org/10.1093/molbev/msy170

Fan GY, An YR, Peng CX, et al., 2019. Forensic and phylogenetic analyses among three Yi populations in Southwest China with 27 Y chromosomal STR loci. *Int J Legal Med*, 133(3):795-797.
https://doi.org/10.1007/s00414-018-1984-4

Feng QD, Lu DS, Xu SH, 2018. *AncestryPainter*: a graphic program for displaying ancestry composition of populations and individuals. *Genomics Proteomics Bioinf*, 16(5):382-385.
https://doi.org/10.1016/j.gpb.2018.05.002

Gan RJ, Pan SL, Mustavich LF, et al., 2008. Pinghua population as an exception of Han Chinese's coherent genetic structure. *J Hum Genet*, 53(4):303-313.
https://doi.org/10.1007/s10038-008-0250-x

He GL, Wang MG, Zou X, et al., 2021. Peopling history of the Tibetan Plateau and multiple waves of admixture of Tibetans inferred from both ancient and modern genome-wide data. *Front Genet*, 12:725243.
https://doi.org/10.3389/fgene.2021.725243

Jeong C, Alkorta-Aranburu G, Basnyat B, et al., 2014. Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat Commun*, 5:3281.
https://doi.org/10.1038/ncomms4281

Jeong C, Peter BM, Basnyat B, et al., 2017. A longitudinal cline characterizes the genetic structure of human populations in the Tibetan Plateau. *PLOS One*, 12(4):e0175885.
https://doi.org/10.1371/journal.pone.0175885

Kang LL, Lu Y, Wang CC, et al., 2012. Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the eastern Himalayas. *Ann Hum Genet*, 76(1):92-99.
https://doi.org/10.1111/j.1469-1809.2011.00690.x

Kim K, Omori R, Ito K, 2017. Inferring epidemiological dynamics of infectious diseases using Tajima's *D* statistic on nucleotide sequences of pathogens. *Epidemics*, 21:21-29.
https://doi.org/10.1016/j.epidem.2017.04.004

Kutanan W, Liu D, Kampuansai J, et al., 2021. Reconstructing the human genetic history of mainland southeast Asia: insights from genome-wide data from Thailand and Laos. *Mol Biol Evol*, 38(8):3459-3477.
https://doi.org/10.1093/molbev/msab124

Li BB, Zhong FG, Yi HS, et al., 2007. Genetic polymorphism of mitochondrial DNA in Dong, Gelao, Tujia, and Yi ethnic populations from Guizhou, China. *J Genet Genomics*, 34(9):800-811.
https://doi.org/10.1016/s1673-8527(07)60091-5

Li CP, Tian DM, Tang BX, et al., 2021. Genome variation map: a worldwide collection of genome variations across multiple species. *Nucleic Acids Res*, 49(D1):D1186-D1191.
https://doi.org/10.1093/nar/gkaa1005

Li JW, Zeng W, Zhang Y, et al., 2017. Ancient DNA reveals genetic connections between early Di-Qiang and Han Chinese. *BMC Evol Biol*, 17:239.
https://doi.org/10.1186/s12862-017-1082-0

Li L, Huang PD, Sun XH, et al., 2021. The ChinaMap reference panel for the accurate genotype imputation in Chinese populations. *Cell Res*, 31:1308-1310.
https://doi.org/10.1038/s41422-021-00564-z
https://doi.org/10.1093/molbev/msz072

Liu CC, Witonsky D, Gosling A, et al., 2022. Ancient genomes from the Himalayas illuminate the genetic history of Tibetans and their Tibeto-Burman speaking neighbors. *Nat Commun*, 13:1203.
https://doi.org/10.1038/s41467-022-28827-2

Liu D, Duong NT, Ton ND, et al., 2020. Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. *Mol Biol Evol*, 37(9):2503-2519.
https://doi.org/10.1093/molbev/msaa099

Liu Y, Wang MG, Chen PY, et al., 2021. Combined low-/high-density modern and ancient genome-wide data document genomic admixture history of high-altitude East Asians. *Front Genet*, 12:582357.
https://doi.org/10.3389/fgene.2021.582357

Lu DS, Lou HY, Yuan K, et al., 2016. Ancestral origins and

genetic history of Tibetan highlanders. *Am J Hum Genet*, 99(3):580-594.
https://doi.org/10.1016/j.ajhg.2016.07.002

Mallick S, Micco A, Mah M, et al., 2024. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. *Sci Data*, 11:182.
https://doi.org/10.1038/s41597-024-03031-7

Manichaikul A, Mychaleckyj JC, Rich SS, et al., 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22):2867-2873.
https://doi.org/10.1093/bioinformatics/btq559

CNCB-NGDC Members and Partners, 2023. Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2024. *Nucleic Acids Res*, 52(D1):D18-D32.
https://doi.org/10.1093/nar/gkad1078

Ning C, Li TJ, Wang K, et al., 2020. Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat Commun*, 11:2700.
https://doi.org/10.1038/s41467-020-16557-2

Patterson N, Moorjani P, Luo Y, et al., 2012. Ancient admixture in human history. *Genetics*, 192(3):1065-1093.
https://doi.org/10.1534/genetics.112.145037

Pickrell JK, Pritchard JK, 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet*, 8(11):e1002967.
https://doi.org/10.1371/journal.pgen.1002967

Price AL, Patterson NJ, Plenge RM, et al., 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*, 38:904-909.
https://doi.org/10.1038/ng1847

Qi XB, Cui CY, Peng Y, et al., 2013. Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the Tibetan Plateau. *Mol Biol Evol*, 30(8): 1761-1778.
https://doi.org/10.1093/molbev/mst093

Qin ZD, Yang YJ, Kang LL, et al., 2010. A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *Am J Phys Anthropol*, 143(4):555-569.
https://doi.org/10.1002/ajpa.21350

Sagart L, Jacques G, Lai YF, et al., 2019. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proc Natl Acad Sci USA*, 116(21):10317-10322.
https://doi.org/10.1073/pnas.1817972116

Shi H, Zhong H, Peng Y, et al., 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol*, 6:45.
https://doi.org/10.1186/1741-7007-6-45

Stoneking M, Delfin F, 2010. The human genetic history of East Asia: weaving a complex tapestry. *Curr Biol*, 20(4): R188-R193.
https://doi.org/10.1016/j.cub.2009.11.052

Sun YT, Wang MG, Sun QX, et al., 2024. Distinguished biological adaptation architecture aggravated population differentiation of Tibeto-Burman-speaking people. *J Genet Genomics*, 51(5):517-530.
https://doi.org/10.1016/j.jgg.2023.10.002

Tao L, Yuan HB, Zhu KY, et al., 2023. Ancient genomes reveal millet farming-related demic diffusion from the Yellow River into southwest China. *Curr Biol*, 33(22):4995-5002.e7.
https://doi.org/10.1016/j.cub.2023.09.055

Torroni A, Miller JA, Moore LG, et al., 1994. Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol*, 93(2):189-199.
https://doi.org/10.1002/ajpa.1330930204

Victor HM, 1998. The Bronze Age and Early Iron Age peoples of eastern central Asia. *Early China*, 23:339-352.
https://doi.org/10.1017/S0362502800001000

Wang CC, Wang LX, Shrestha R, et al., 2014. Genetic structure of Qiangic populations residing in the western Sichuan corridor. *PLOS One*, 9(8):e103772.
https://doi.org/10.1371/journal.pone.0103772

Wang CC, Yeh HY, Popov AN, et al., 2021. Genomic insights into the formation of human populations in East Asia. *Nature*, 591:413-419.
https://doi.org/10.1038/s41586-021-03336-2

Wang FJ, 2013. On the oracle bone inscriptions “Qiang”. *J Qilu Normal Univ*, 28(3):68-71 (in Chinese).

Wang HR, Yang MA, Wangdue S, et al., 2023. Human genetic history on the Tibetan Plateau in the past 5100 years. *Sci Adv*, 9(11):eadd5582.
https://doi.org/10.1126/sciadv.add5582

Wang MG, Yuan DD, Zou X, et al., 2021. Fine-scale genetic structure and natural selection signatures of southwestern Hans inferred from patterns of genome-wide allele, haplotype, and haplogroup lineages. *Front Genet*, 12:727821.
https://doi.org/10.3389/fgene.2021.727821

Wen B, Xie XH, Gao S, et al., 2004. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet*, 74(5):856-865.
https://doi.org/10.1086/386292

Yang MA, Fan XC, Sun B, et al., 2020. Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science*, 369(6501):282-288.
https://doi.org/10.1126/science.aba0909

Yao HB, Tang SW, Yao XT, et al., 2017. The genetic admixture in Tibetan-Yi Corridor. *Am J Phys Anthropol*, 164(3): 522-532.
https://doi.org/10.1002/ajpa.23291

Yao HB, Wang MG, Zou X, et al., 2021. New insights into the fine-scale history of western-eastern admixture of the northwestern Chinese population in the Hexi Corridor via genome-wide genetic legacy. *Mol Genet Genomics*, 296(3):631-651.
https://doi.org/10.1007/s00438-021-01767-0

Yao YG, Kong QP, Bandelt HJ, et al., 2002. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet*, 70(3):635-651.
https://doi.org/10.1086/338999

Yu XE, Li H, 2021. Origin of ethnic groups, linguistic families, and civilizations in China viewed from the Y chromosome. *Mol Genet Genomics*, 296(4):783-797. https://doi.org/10.1007/s00438-021-01794-x

Zhang MH, Yan S, Pan WY, et al., 2019. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature*, 569:112-115. https://doi.org/10.1038/s41586-019-1153-z

Zhang Z, Zhang YL, Wang YN, et al., 2022. The Tibetan-Yi region is both a corridor and a barrier for human gene flow. *Cell Rep*, 39(4):110720. https://doi.org/10.1016/j.celrep.2022.110720

Zhao M, Kong QP, Wang HW, et al., 2009. Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci USA*, 106(50):21230-21235. https://doi.org/10.1073/pnas.0907844106

Zhao YB, Li HJ, Li SN, et al., 2011. Ancient DNA evidence supports the contribution of Di-Qiang people to the Han Chinese gene pool. *Am J Phys Anthropol*, 144(2):258-268. https://doi.org/10.1002/ajpa.21399

Supplementary information

Materials and methods; Tables S1–S6; Figs. S1–S8