Check for updates

RETRACTED: Genetic admixture history and forensic characteristics of Tibeto-Burman-speaking Qiang people explored *via* the newly developed Y-STR panel and genome-wide SNP data

Guanglin He[1,2,3,4]*†, Atif Adnan[2,4,5,6]*†,
Wedad Saeed Al-Qahtani[6], Fatmah Ahmed Safhi[7],
Hui-Yuan Yeh[3], Sibte Hadi[6], Chuan-Chao Wang[2,4]*,
Mengge Wang[8]*, Chao Liu[9]* and Jun Yao[5]*

[1]Institute of Rare Diseases, West China Hospital of Sichuan University, Sichuan University, Chengdu, China, [2]Department of Anthropology and Ethnology, Institute of Anthropology, School of Sociology and Anthropology, Xiamen University, Xiamen, China, [3]School of Humanities, Nanyang Technological University, Singapore, Singapore, [4]State Key Laboratory of Cellular Stress Biology, National Institute for Data Science in Health and Medicine, School of Life Sciences, Xiamen University, Xiamen, China, [5]Department of Forensic Genetics, School of Forensic Medicine, China Medical University, Shenyang, China, [6]Department of Forensic Sciences, College of Criminal Justice, Naif Arab University for Security Sciences, Riyadh, Saudi Arabia, [7]Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, [8]Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China, [9]Department of Forensic Genetics, School of Forensic Medicine, Southern Medical University, Guangzhou, China

Fine-scale patterns of population genetic structure and diversity of ethnolinguistically diverse populations are important for biogeographical ancestry inference, kinship testing, and development and validation of new kits focused on forensic personal identification. Analyses focused on forensic markers and genome-wide single nucleotide polymorphism (SNP) data can provide new insights into the origin, admixture processes, and forensic characteristics of targeted populations. Qiang people had a large sample size among Tibeto-Burmanspeaking populations, which widely resided in the middle latitude of the Tibetan Plateau. However, their genetic structure and forensic features have remained uncharacterized because of the paucity of comprehensive genetic analyses. Here, we first developed and validated the forensic performance of the AGCU-Y30 Y-short tandem repeats (STR) panel, which contains slowly and moderately mutating Y-STRs, and then we conducted comprehensive population genetic analyses based on Y-STRs and genome-wide SNPs to explore the admixture history of Qiang people and their neighbors. The validated results of this panel showed that the new Y-STR kit was sensitive and robust enough for forensic applications. Haplotype diversity (HD) ranging from 0.9932 to 0.9996 and allelic frequencies ranging from 0.001946 to 0.8326 in 514 Qiang people demonstrated that all included

markers were highly polymorphic in Tibeto-Burman people. Population genetic analyses based on Y-STRs [R_{ST}, F_{ST}, multidimensional scaling (MDS) analysis, neighboring-joining (NJ) tree, principal component analysis (PCA), and median-joining network (MJN)] revealed that the Qiang people harbored a paternally close relationship with lowland Tibetan-Yi corridor populations. Furthermore, we conducted a comprehensive population admixture analysis among modern and ancient Eurasian populations based on genome-wide shared SNPs. We found that the Qiang people were a genetically admixed population and showed closest relationship with Tibetan and Neolithic Yellow River farmers. Admixture modeling showed that Qiang people shared the primary ancestry related to Tibetan, supporting the hypothesis of common origin between Tibetan and Qiang people from North China.

Introduction

East Asia has more than eight language families or groups that include Tungusic, Mongolic, Turkic, Sino-Tibetan, Austronesian, Austroasiatic, Tai-Kadai, and Hmong-Mein, and hundreds of ethnic groups and harbored complex patterns of genetic diversity (GD). Recent comprehensive genetic analyses have shown that population genetic analyses based on geographically and ethnically different populations could provide new insights into the formation of modern populations, especially in some regions with rich ethnolinguistic diversity (Liu et al., 2020; Kutanan et al., 2021). Recent advances in population genomic studies have provided an updated landscape of the genetic basis of East Asians (Wang C. C. et al., 2021). The ChinaMap project based on whole-genome sequences of over 10,000 people showed a significant genetic differentiation between Han Chinese and other minorities and found seven substructured groups of Han Chinese (Li et al., 2021). Genome-wide single nucleotide polymorphism (SNP) data from northern, central, southern, and western China also illuminated the different ancestral source composition and admixture weight in the formation of geographically diverse modern Hans (The Genographic Consortium, Gan et al., 2008; He et al., 2021a,b,c; Liu et al., 2021a,b; Wang M. et al., 2021; Yao et al., 2021). Furthermore, recent ancient DNA studies provided new insights into the peopling of the Tibetan Plateau and modern Sino-Tibetan people (Jeong et al., 2016; He et al., 2021c; Liu et al., 2021b; Wang C. C. et al., 2021). Ancient genomes from Nepal showed their genetic connection to East Asians rather than South Asians; ancient northern East Asian genomes from Gansu, Henan, and Shaanxi provinces also showed a close genetic relationship between millet farmers and modern Tibetans (Wang C. C. et al., 2021). These genetic findings showed complex patterns of GD and the population structure of

East Asia. Similar patterns of the consistent association between African GD and four language families [Afro-Asiatic, Nilo-Saharan, Niger-Saharian (Niger-Congo), and Khoisan] were also identified in East Asia (Lachance et al., 2012). Altaic-speaking populations in North China showed a predominant ancestry related to ancient Mongolian ancestry and western Eurasian ancestry (He et al., 2019, 2022b). Other populations from Sino-Tibetan language families in central and southern East Asia harbored a more primary ancestry related to ancient Yellow River Basin and Yangtze River Basin ancestries (He et al., 2021a,b, 2022a). These population stratifications provided a new opportunity for the development and validation of new forensic kits focused on the fine-scale genetic background of East Asians with different genetic markers and higher resolution. This type of kit could provide a more comprehensive genetic landscape reconstruction of East Asians.

Genetic markers localized in non-recombining regions of the Y-chromosome have played an important role in forensic identification, kinship testing, and molecular anthropology. The initial peopling history of East Asians has been explored *via* Y-chromosome markers, such as the demic diffusion of Han Chinese rather than cultural diffusion (Wen et al., 2004a). Besides the function of population history reconstructions, Y-STRs are widely used for characterizing male contributions to male-female mixtures in forensic genetics, particularly in sexual assault cases or familial searches such as kinship caseworks involving male offspring, especially in deficient paternity cases where the putative father was unavailable (Roewer and Epplen, 1992; Roewer et al., 1992; Prinz et al., 1997; Calacal et al., 2005). Currently, many Y-STR commercial kits are available, such as Yfiler (Thermo Fisher Scientific, Foster City, CA, USA) (Mulero et al., 2006), Yfiler Plus (Thermo Fisher Scientific) (Gopinath et al., 2016), Goldeneye 20Y (Peoplespot R&D, Jiangsu, China), Goldeneye 26Y (Peoplespot

R&D, Jiangsu, China), DNATyper Y26 (Second Institute of the Ministry of Public Security) (Mo et al., 2019), PowerPlex Y23 (Promega, Germany) (Thompson et al., 2013), and AGCU Y24 (China–Germany United, China) (Meng et al., 2019). These kits usually contain various numbers of Y-STRs ranging from 17 to 27; among these, 3–7 rapidly mutating Y-STRs have been included (Ballantyne et al., 2010). Several studies have demonstrated that Y-STR haplotype diversity (HD) and male lineage differentiation can be improved by adding additional carefully chosen Y-STRs (Adnan et al., 2018). Rapidly mutating Y-STRs with mutation rates $>1 \times 10^{-2}$ have been used for differentiation of closely related male individuals and can help individualize them because of their high discrimination power (DP) (Ballantyne et al., 2012; Adnan et al., 2016; Rakha et al., 2018). However, when it comes to familial searching, these Y-STRs are not conclusive. Keep this in mind, the AGCU has introduced a new 6-dye system that can co-amplify 30 Y-STRs: DYS392, DYS389I, DYS447, DYS389II, DYS438, DYS527a/b, DYS522 labeled with FAM, DYS391, DYS456, DYS19, DYS388, DYS448, DYS385a/b labeled with HEX, DYS481, DYS437, DYS390, DYS460, DYS533, DYS458 labeled with TAMRA, DYS393 Y_GATA_H4, DYS439, DYS635, DYS444, DYS643 labeled with ROX while DYS549, DYS557, DYS520 are labeled with VIG dye (Supplementary Figure 1). The 30 Y-STRs have low mutation rates (1×10^{-3} to 1×10^{-5}), which make them good candidates for family lineage reconstruction and are helpful for construction of Y-STR DNA databases, forensic caseworks, and kinship cases because of having a reduced risk of homogeneity and heterogeneity. We tested the forensic performance of accuracy, stability, sensitivity, specificity, resistance to inhibition, and mixture samples in the Qiang population from Beichuan, Sichuan, China.

Sino-Tibetan-speaking populations were the largest population group in Asia and were separated into Tibeto-Burman speakers residing in the Tibetan Plateau and Sinitic-speaking people in the low altitude regions of East Asia. The Qiang people are an ancient group in mainland China and are associated with the spread of modern Sino-Tibetan language and cultures. The earliest documented knowledge for the Qiang people was obtained from the Oracle bone which is known as a "living fossil" in the history of the Chinese nation in almost 3,000 years ago. Many of these people in the past were designated as "Qiang." Later on, they were also influenced by Chinese culture and amalgamated with the Han population during the Ming and Qing dynasties. These people were again reclassified. Non-Han people who were living in the upper valley of Min river and Beichuan area are called Qiang (Wang, 2013). Qiang people speak their own language of Qiangic, a subfamily of Tibeto-Burman languages. They are one of 55 minority ethnic groups in China. All minorities are 8.49% of the total Chinese populations. Qiang people are 0.23% of minority populations with a population size of 310,000, and one-third

of their population is located in Beichuan Qiang Autonomous County (Guo et al., 2003). Beichuan Qiang Autonomous County is associated with Mianyang city in Sichuan province and is the only Qiang autonomous county in China.

A previous genetic analysis has mainly focused on the fine-scale population genetic structure of Han Chinese and Tibetan people using genome-wide SNP data; however, this important Qiang population in the Tibeto-Burman language group is uncharacterized until now (He et al., 2021c; Wang M. et al., 2022). There is limited knowledge of the GD and fine-scale admixture structure of Qiang people and their genetic origin, and the detailed evolutionary history reconstructed based on genome-wide SNP data also remains uncharacterized. Thus, here, we first validated a Y-STR profiling panel to obtain specific Y-STR data and conducted a genome-wide SNP study. This was aimed to determine the forensic efficiency and application of the newly developed AGCU Database Y30 kit, study the genetic relationships between Qiang people and their neighbors based on low-moderate mutated Y-STRs, and to determine the fine-scale genetic structure, admixture history, and development of modern Qiang people based on the genome-wide SNP data.

Materials and methods

Sample preparation

Control DNA 9948, 007 (male DNA), and 9947A (female DNA) were used for sensitivity, precision evaluation, inhibitor study, and DNA mixture studies. Blood samples were collected from a total of 534 unrelated healthy individuals from the Qiang ethnic group (514 male and 20 female) from Beichuan Qiang Autonomous County, Mianyang city, Sichuan province, China (Figure 1A). All participants gave their informed consent either orally and with thumbprints (in case they could not write) or in writing after the study aims and procedures were carefully explained to them in their language. The study was approved by the ethical review board (dated March 20, 2019 with approval reference no. 2019-86-P) of the China Medical University and Xiamen University (XDYX2019009), China. Whole blood samples of pigs, cattle, sheep, chickens, ducks, fish, rats, and rabbits were purchased from Wuxi Zoological Garden (Wuxi, Jiangsu, China), which were collected to conduct the species specificity test. All the blood samples were stored at $-20°C$ before DNA extraction. DNA was isolated from blood using the ReliaPrep^{TM} Blood genomic DNA (gDNA) Miniprep System (Promega, Madison, WI, USA) according to the manufacturer's instructions. DNA quantitation for all the samples was conducted using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA), and the final concentration of DNA was diluted to -2 ng/µl. Moreover, 9948 was also used as positive/ control sample in all PCR batches.
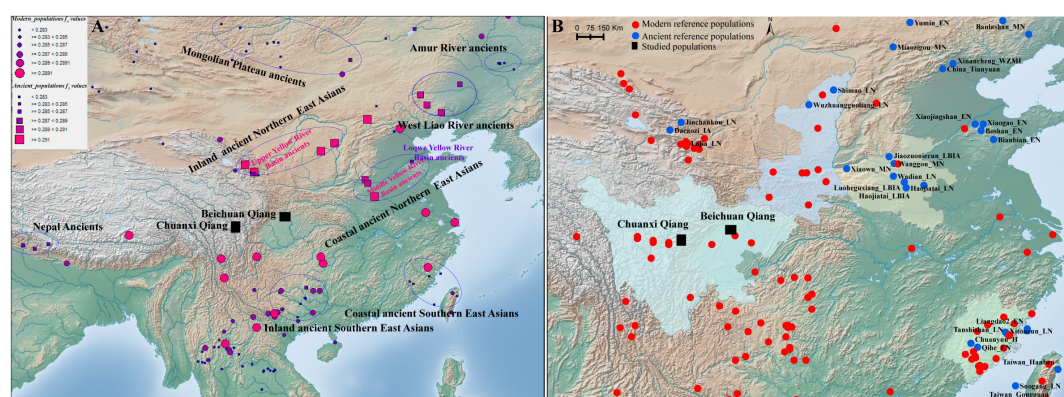
FIGURE 1

Geographical positions of our studied populations and reference populations. **(A)** Genetic affinity between Chuanxi Qiang and reference populations from the HO dataset. The color and size of the circle showed the genetic affinity between the Qiang people and the modern reference populations, and the color and size of the square denote the genetic affinity with ancient reference populations. **(B)** Geographical position of the studied populations and reference populations from the Illumina dataset.

Reference dataset

Y-STR data from different populations in the Y-STR-based analysis were collected from the YHRD database.[1] The genome-wide SNP data of Qiang people developed by us was used for this study and then merged with other modern and ancient genomes from China, Nepal, Southeast Asia, Japan, Mongolia, and Siberia (Yang et al., 2020; Mao et al., 2021; Wang T. et al., 2021) collected from the Allen Ancient DNA Resource (AADR, **Figure 1B**).[2]

Validation framework

PCR amplification and electrophoresis

A total of 514 unrelated individuals from the Qiang population were genotyped with the AGCU Database Y30 kit. PCR amplification was performed in a GeneAmp® PCR System 9700 thermal cycler (Thermo Fisher Scientific) on a gold-plated silver block. After optimization, the final reaction system was performed in a 25-μl volume containing 10 μl of master mix, 5 μl of primer set, 1 μl of 5 U/μl Taq DNA polymerase (AGCU ScienTech Incorporation, Wuxi, China), and 0.5–2 ng of gDNA under the following PCR conditions: initial denaturation of 95°C for 2 min, 30 cycles of denaturation, annealing, and extension step at 94°C for 30 s, 59°C for 60 s, and 72°C for 60 s, followed by a final extension step at 60°C for 30 min and final soak at 4°C. Amplification products were analyzed regarding AGCU Marker SIZ-500 internal size standard (AGCU ScienTech Incorporation, Wuxi, China) and

the AGCU Database Y30 kit Allelic Ladder using an ABI 3500 genetic analyzer (Thermo Fisher Scientific) with the POP-4 polymer (Thermo Fisher Scientific) according to the AGCU Database Y30 kit standard protocol. An injection time of 10 s, injection and run voltages of 3 KV, and run time 2,500 s were used for all electrophoresis runs on a 3,500 genetic analyzer. Gene-Mapper Software version 4.0 (Thermo Fisher Scientific) was used for genotype assignment.

Sensitivity test

Serial dilutions of control DNA 9948 were prepared using dH$_2$O (1, 0.5, 0.25, 0.125, 0.1, and 0.05 ng/μl). These were amplified in triplicate using a 25- to 10-μl reaction volume. Amplification processes were performed according to the manufacturer's instructions.

Species specificity test

gDNA 1–2.5 ng/μl of cat, sheep, chicken, duck, rabbit, mice, pig, and *Escherichia coli* was used in a 10-μl reaction volume for each animal and amplified according to the above-mentioned PCR conditions.

Consistency test

We selected 20 unrelated individuals from the Qiang population and genotyped them with the AGCU Database Y30 kit, and the same 20 unrelated individuals from the Qiang population were also genotyped with the Yfiler plus kit.

Stability test

To test the stability of the AGCU Database Y30 kit, the components of the kit were repeatedly frozen and thawed 5–10 times before amplification while the control group was kept at −20°C. Furthermore, common inhibitors that may be a part of the biological samples were mixed with control DNA 9948

for subsequent PCR experiments. The concentrations of used inhibitors were diluted as follows: heme (25, 50, 75, 100, and 125 µM), indigo (10, 12, 14, 16, and 18 µM), humic acid (40, 50, 60, 70, and 80 µg/µl), Ca^{2+} (0.5, 0.75, 1, 1.5, and 2 µM), hemoglobin (20, 30, 40, 50, and 60 µM), and EDTA (500, 750, 1,000, 1,500, and 2,000 µM) by Sigma-Aldrich (Saint Louis, MO, USA).

Gender-specific test

gDNA 1–2 ng/µl from 20 unrelated female individuals was amplified in a 10-µl reaction volume according to the PCR conditions of the AGCU Database Y30 kit.

Mixture sample analysis

Manually, male (9948)/male (007) mixture samples were prepared with concentration ratios of 1:1, 1:4, 1:9, 1:19, and 1:29, while male (9948)/female (9947A) samples were also prepared with the same concentration ratios and then amplified according to the PCR conditions of the AGCU Database Y30 kit.

PCR variation of primer concentration, annealing temperature, and cycle number testing

We used 10 and 25 µl reaction volumes with 1X and 2X primers to detect 1 ng of control DNA (9948) and repeated it thrice with annealing temperatures 56, 57, 58, 59, 60, 61, 62, and 63°C, and cycle numbers 29 and 30 were used.

Population and concordance studies based on non-high mutated STRs

Genotyping of Y-STR data and forensic statistical parameter estimation

Allelic and haplotype frequencies were computed with the direct counting method, and the GD of a single locus or HD of all linked loci was calculated using:

$$GD \; or \; HD = \frac{n}{n-1} \times \sum p_i^2$$

Where n is the male population size and p_i is the frequency of ith haplotype. Discrimination capacity (DC) was calculated as the ratio of unique haplotypes in the samples. Match probabilities (MPs) were calculated as $\sum p_i^2$, where p_i is the frequency of the i-th haplotype. We calculated both R_{ST} and F_{ST} values because in the generalized stepwise mutation model, R_{ST} offers relatively unbiased evaluations of migration rates and times of population divergence, while, on other hand, F_{ST} tends to show too much population similarity, predominantly when migration rates are low or divergence times are long. Genetic distances were evaluated between reference populations and the Qiang population on overlapping STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392,

DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and Y_GATA_H4) using the Arlequin Software v3.5 (Excoffier and Lischer, 2010). Reduced dimensionality spatial representation of the populations was performed based on R_{ST} values by multidimensional scaling (MDS) with IBM SPSS Statistics for Windows, version 23.0 (IBM Corp., Armonk, NY, USA). Heatmaps were generated using R_{ST} and F_{ST} values and the R program V3.4.1 platform (Plummer, 2003) with the help of a ggplot2 module.

Phylogenetic analysis

A neighbor-joining phylogenetic tree was constructed for the Qiang population and the reference populations based on a distance matrix of F_{ST} using the Mega7 software (Kumar et al., 2016). Y-DNA Haplogroup Predictor NEVGEN[3] was used to predict Y-SNP haplogroups from Y-STR haplotypes. Haplotypes with null alleles or duplicated variants in the Qiang population were excluded from the analysis.

The median-joining network

To define the genetic relationships among the Qiang population individuals, we used the stepwise mutation model and median joining-maximum parsimony algorithm using the program Network 5 as described in the Fluxus Engineering website[4] and weighting criteria for Y-STRs. Haplotypes with null alleles or duplicated variants in the Qiang population were excluded from the analysis.

Population admixture and evolutionary history reconstruction

Principal component analysis and model-based ADMIXTURE

Population admixture history reconstruction was conducted based on genetic variations of 600K SNPs included in the Human Origins array. The merged dataset of modern and ancient populations also included genetic variations of 600K SNPs. We conducted a principal component analysis (PCA) for 1,158 individuals from 105 modern and ancient populations using smartpca of EIGENSOFT v.6.1.4 (Patterson et al., 2006). Ancient reference East Asians were projected onto the basic background of modern East Asians with additional parameters (numoutlieriter: 0 and lsqproject: YES). We merged the genome-wide SNPs of 20 Qiang individuals with 1,986 reference individuals to conduct a model-based ADMIXTURE analysis. We used PLINK v.1.9 (Chang et al., 2015) to prune the raw dataset using the following parameters (–indep-pairwise 200 25 0.4) and ran ADMIXTURE (v.1.3.0) (Alexander et al., 2009) with

---

3   http://www.nevgen.org

4   http://www.fluxus-engineering.com

10-fold cross-validation and predefined ancestral sources. In-house script and PLINK v.1.9 (Chang et al., 2015) were used to calculate the pairwise F_{ST} genetic distances between the Qiang and other reference populations.

Allele sharing and admixture signatures in *f*-statistics

We estimated the shared genetic drift between the Qiang people and other modern and ancient populations using the *qp3Pop* in ADMIXTOOLS in the form of f_3 (Qiang, reference; Mbuti) and estimated the admixture evidence using *qp3Pop* in ADMIXTOOLS (Reich et al., 2009; Patterson et al., 2012) in the form of f_3 (source 1, source 2; Qiang). Additionally, we calculated f_4-values in the form of f_4 (reference 1, reference 2; Qiang, Mbuti) and f_4 (reference 1, Qiang; reference 2, Mbuti) to test the asymmetrical/symmetrical relationship between the Qiang people and the others. We used qpAdm to estimate the admixture proportion and ancestral sources of the Qiang people and the other neighbors. Both three-way and two-way admixture models were used here. We used nine populations (Mbuti, Ust_Ishim, Kostenki14, Papuan, Australian, Mixe, MA1, Jehai, and Tianyuan) as outlier groups. To further explore the phylogenetic framework of the Qiang people, we used the *qpGraph* program implemented in the ADMIXTOOLS 2 package (Patterson et al., 2012; Fu et al., 2015) to test all possible evolutionary models and choose the best models based on Z-scores and minimum likelihood.

Ancestral source composition and admixture dates inferred from chromosome painting

We phased on modern genomes in the merged dataset using SHAPEIT v2 (Browning and Browning, 2011) with the following parameter settings (–burn 10 –prune 10 –main 30). The phased haplotypes and refined-ibd.17Jan20.102 (Browning and Browning, 2013) were used to calculate the pairwise IBD between the Qiang and modern reference populations. FineStructure v4.1.1 (Hellenthal et al., 2014), ChromoPainter v2, and ChromoCombine v2 (Hellenthal et al., 2014) were used to explore the fine-scale population structure and relationship of redefined genetic homogeneous populations. We finally used GLOBETROTTER to identify, date, and describe admixture events.

Results

Validation of the developmental 30-marker Y-STR amplification system

For sensitivity, we separately amplified control DNA 9948 and 007 with different concentrations (1, 0.5, 0.25, 0.125, 0.1, and 0.05 ng/μl). The results showed that even as low as 0.125 ng of DNA template yielded complete amplification of AGCU

Database Y30 kit loci (**Supplementary Figure 2**). When we reduced the DNA template concentration to 0.05 ng, 80% of the loci were not detected. So, this kit can effectively detect samples with a sensitivity of 0.125 ng; hence, the AGCU Database Y30 kit demonstrated a high level of sensitivity.

For accuracy and consistency, we used different thermocyclers (GeneAmp® PCR system 9700, Veriti™ Thermal Cycler, 2720 Thermal Cycler, and ProFlex PCR System; Applied Biosystems, Foster City, CA, USA) for the amplification of DNA templates from blood, saliva, filter paper, and FTA card. After amplification using the AGCU Database Y30 kit, amplicons were detected with ABI3130XL, ABI3500, and ABI3730XL (Applied Biosystems, Waltham, MA, USA). There was no significant difference in amplification efficiency, and the genotyping results are complete, clear, and stable. There were no obvious non-specific peaks seen with the AGCU Database Y30 kit. The genotyping results of the AGCU Database Y30 kit are almost the same with those of AmpFISTR® Yfiler plus® on overlapping STRs (**Supplementary Figure 3**), which indicates that the kit has good accuracy and consistency.

For identity, we extracted DNA from blood, saliva, and hair samples from the same source and amplified them using the AGCU Database Y30 kit. The results showed that the DNA typing results of the blood, saliva, and hair samples from the same source were all identical (**Supplementary Figure 4**), which means that the kit can precisely determine the genotype.

To check the stability of the AGCU Database Y30 kit, the components of the kit were repeatedly frozen and thawed 10 times, and PCR amplification was performed on the final status (10th). The capillary electrophoresis results did not show any significant difference in amplification efficiency between different carrier samples (FTA card and filter paper) and the control that shows us the stability of the kit.

Generally, biological samples obtained from the crime scenes have been encountered in the different natural environment and some mixed agents can cause inhibition in the process of polymerase chain reaction. Intermingling of such agents with biological samples can lead to failure of the amplification process, and we cannot get a full profile. The most common inhibitors that frequently cause problems are from biological materials carrying themselves like hematin from blood and humic acid from soil. Humic acid will bind to template DNA and hematin will alter Taq polymerase. To check the tolerance level of the AGCU Database Y30 kit, we added six different inhibitors as mentioned in the M and M section. The typing results showed that a complete profile was generated when we used 40 and 50 μg/μl of humic acid, 20 and 30 μM of hemoglobin, 500 and 750 μM of EDTA, 0.5 and 0.75 μM of Ca^{2+}, 25 and 50 μM of hematin, and 4 and 8 μM of indigo. Hence, the kit has shown a high tolerance level.

For specificity, we amplified 1–2.5 ng of genomic DNA from male human, cat, sheep, chicken, duck, rabbit, mice, pig, *E. coli*, and female human using the AGCU Database Y30 kit. We did

not observe any non-specific DNA amplification that showed high species specificity (**Supplementary Figure 5**). We also amplified 20 unrelated female individuals and did not observe any peak after capillary electrophoresis, which means that the multiplex assay was gender-specific.

Mixtures of DNA samples are frequent in forensic caseworks, mainly in sexual assault cases. We mixed control 9948 (male DNA) and control 9947A (female DNA) at various ratios of 1:1, 1:4, 1:9, 1:19, 1:29, which were amplified in triplicate. All the mixtures accurately detected the male fraction (**Supplementary Figure 6A**). For the male-to-male mixture samples, control DNA 9948 and control DNA 007 were used to prepare mixtures at ratios of 1:1, 1:4, 1:9, 1:19, and 1:29. These were amplified in triplicate. At the mixture ratio of DNA 007 is 1:19, the genotyping of each STR present in the AGCU Database Y30 kit was detected correctly (**Supplementary Figure 6B**).

PCR environment test results

Reaction volume and primer concentration

We used two different reaction volumes of 25 and 10 µl for amplification. The genotyping results showed that there were no allele dropout or non-specific peaks, and that there was not any significant difference in peak heights for both reaction volumes. When we used different primer concentrations (1X and 2X) for the amplification process, the genotyping results were consistent, and there was no allele loss or non-specific peaks (**Supplementary Figure 7**).

Annealing temperature and cycle numbers

Denaturation, annealing, and extension are three vital steps for a successful PCR. Annealing is the key step in PCR and can determine the fate of its success or failure. Every amplification kit has a specific PCR annealing temperature that can fluctuate from one thermocycler to another. Therefore, it is important to check the tolerance level of the amplification kit to annealing temperature variations. We used different annealing temperatures with variable cycle numbers and found that at 63°C annealing temperature with 29 amplification cycles, the genotyping results were unstable and that few allele peaks were lower than 100 (**Supplementary Figure 8**). At 59°C annealing temperature with 29 amplification cycles, we got a perfect electropherogram with good peak heights (**Supplementary Figure 9**).

Population studies and forensic parameters

Allelic frequencies and forensic parameters

We successfully genotyped 30 Y-STRs in 514 men in the Qiang population residing in the Sichuan province of China.

A total of 474 haplotypes were detected in the Qiang population. Haplotype data have already been made accessible *via* the Y-chromosome Haplotype Reference Database (YHRD) under accession number YA004683 (**Supplementary Table 1**). The gene or locus diversity (GD), allele frequencies, and number of observed alleles for 26 single-copy STR loci and 2 multi-copy STRs are summarized in **Supplementary Table 2**. The allele frequencies ranged from 0.0019 to 0.8326. Among the single-copy STRs, DYS481 showed the highest diversity of 0.8416 with thirteen different alleles, and DYS391 was the least diverse locus at 0.2856 with five different alleles. The multi-copy Y-STRs DYS385a/b and DYS527 showed more combinations of alleles when compared with the single-copy Y-STRs. Gene diversities were 0.9324 and 0.9284 with 57 and 42 different allele combinations for DYS385a/b and DYS527, respectively. To assess the slow to medium mutating nature of the AGCU Y30 database kit, we evaluated the haplotype resolution with seventeen different combinations (**Table 1**), including the "minimal haplotype 9 Y-STRs," "the extended 11 Y-STRs loci," "PowerPlex Y12 STRs loci," "Y-filer Plus 17 STRs," and so on, until 30 Y-STRs. A total of 474 haplotypes were observed in 30 Y-STRs with a HD of 0.9996 and a DC of 0.9222. Among the 474 haplotypes, 85.79% were unique. When we reduced the number of STRs to Yfiler 17 STRs, a total of 438 haplotypes were observed with an HD value 0.9991 and a DC value 0.8521. Among the 438 haplotypes, 76.26% were unique. The multi-copy Y-STRs (DYS385a/b and DYS527) displayed 211 haplotypes with a HD value of 0.9729 and a DC value 0.4105. Among the 211 haplotypes, 26.85% were unique. The kit showed stronger DP, which means it can serve both purposes: (i) individual search and (ii) familial search.

Phylogenetic analyses and population comparisons

To reinforce the knowledge of ethnohistorical records of the Qiang and other Chinese populations, we used two different methods (pairwise R_{ST} and F_{ST} genetic distances) based on their similarity with *a priori* expectations. F_{ST} is a standardized variance of haplotype frequency and assumes genetic drift as the agent that differentiates populations. R_{ST} is a standardized variance of haplotype size and takes into account both drift and mutation as causes of population differentiation, assuming a stepwise mutation model (Slatkin, 1995). The genetic distances (R_{ST} and F_{ST}) between the Qiang and other Chinese integrated populations are listed in **Supplementary Tables 3**, **4**. The Tibetan population from Amdo, China showed the closest genetic distance (0.108) with the Qiang population followed by the Manchu population (0.1205) from Jilin, China. On the other hand, the Yao population from Liannan, China showed the maximum genetic distance (0.4629) followed by the Korean population (0.3851) from Yanbian, China among the reference populations. According to the F_{ST} genetic distance, the Tujia population (0.0005) from Youyang,

TABLE 1   Forensic statistical parameters of different marker composition in different Y-STR sets.
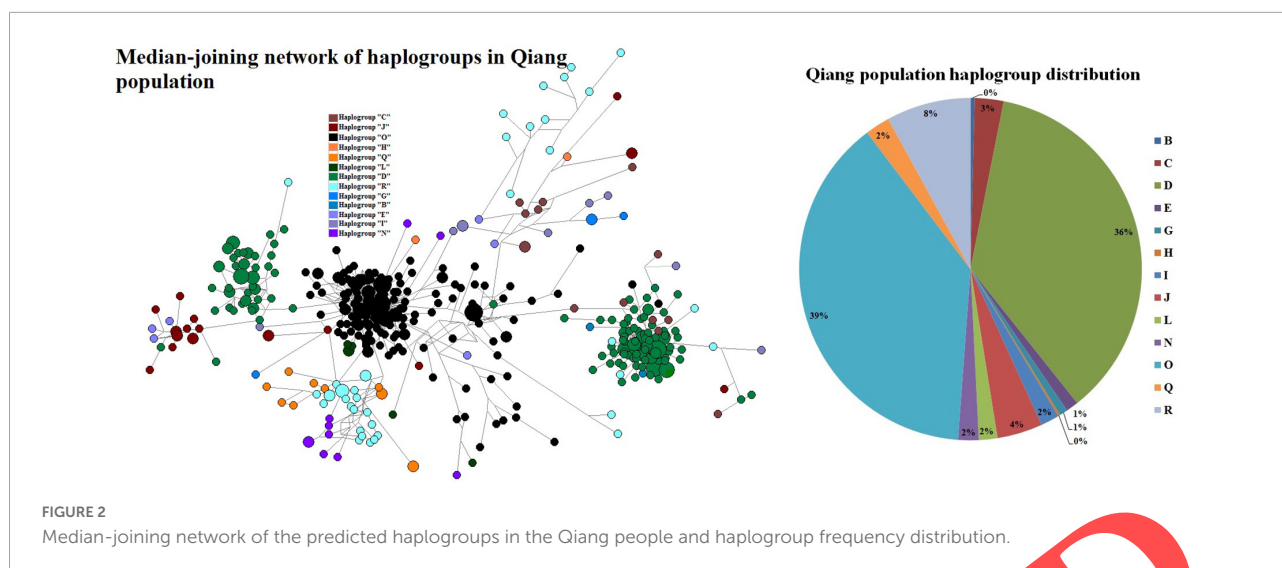
Marker panels	SZ	RMP	GD	HD	UHT	% UHT	DC
MH 9 Loci	514	0.0088	0.9932	315	243	47.2763	0.6128
SWGDAM 11 STRs	514	0.0059	0.996	357	287	0.5584	0.6946
PowerPlex Y12 loci	514	0.0057	0.9962	365	300	0.5837	0.7101
Y-filer 17 loci	514	0.0028	0.9991	438	385	0.749	0.8521
Y-filer +DYS388	514	0.0028	0.9991	438	385	0.749	0.8521
Y-filer +DYS388 & DYS444	514	0.0028	0.9992	443	392	0.7626	0.8619
Y-filer +DYS388 + DYS444 & DYS447	514	0.0025	0.9994	455	392	0.7626	0.8852
Y-filer +DYS388 + DYS444 + DYS447 & DYS460	514	0.0025	0.9995	459	414	0.8054	0.893
Y-filer +DYS388 + DYS444 + DYS447 + DYS460 & DYS481	514	0.0024	0.9995	463	421	0.8191	0.9008
Y-filer +DYS388 + DYS444 + DYS447 + DYS460 + DYS481& DYS520	514	0.0024	0.9995	464	422	0.821	0.9027
Y-filer +DYS388 + DYS444 + DYS447 + DYS460 + DYS481 + DYS520 & DYS522	514	0.0024	0.9995	466	426	0.8288	0.9066
Y-filer +DYS388 + DYS444 + DYS447 + DYS460 + DYS481 + DYS520 + DYS522 & DYS527	514	0.0024	0.9996	469	432	0.8405	0.9125
Y-filer +DYS388 + DYS444 + DYS447 + DYS460 + DYS481 + DYS520 + DYS522 + DYS527 & DYS533	514	0.0024	0.9996	471	436	0.8482	0.9163
Y-filer +DYS388 + DYS444 + DYS447 + DYS460 + DYS481 + DYS520 + DYS522 + DYS527 + DYS533 & DYS549	514	0.0023	0.9996	474	441	0.858	0.9222
Y-filer +DYS388 + DYS444 + DYS447 + DYS460 + DYS481 + DYS520 + DYS522 + DYS527 + DYS533 + DYS549 & DYS557	514	0.0023	0.9996	474	441	0.858	0.9222

Chongqing, China showed the closest genetic distance followed by the Han population (0.0006) from Anshan, China, while the Tuva population (0.0257) from Kanas, China showed the greatest genetic distance followed by the Manchu population (0.0207) from Jilin, China. Phylogenetic relationships between the Qiang population and the reference populations were assessed by MDS analysis based on R_{ST} distances derived from the Y-STR data (Supplementary Figure 10). We observed three loosely bound clusters, and the Qiang population was placed with the Tibetan population on the right side of the plot. The Uighur, Mosuo, Salar, Yi, and Kazakh populations clustered in the middle, while the Han, Tujia, Maio, Gin, and other populations clustered in the left side of the plot. In the neighboring-joining (NJ) tree, the Qiang population was also placed along with the Tibetan, Tuva, Manchu, and Daur populations (Supplementary Figure 11). Genetic distances, MDS plot, and the phylogenetic tree showed that the genetic affinity among the studied populations was consistent with linguistic, ethnic, and geographical classifications.

Ancestry information of Qiang ethnic groups using Y-STRs

Ancestry informative markers (AIMs) play an important role in genealogy. So, we used the NEVGEN software to predict

haplogroups from STRs. We used the predicted results from the software mainly considering the high confidence with our knowledge from previous Y-chromosome studies. There were four haplogroups (D, J, O, and R) that accounted for 87%, while just two haplogroups (D and O) accounted for 75% of the Qiang population (Figure 2). Haplogroup J accounts for 4% of the Qiang population from China. Haplogroup J is predominantly found in the Arabian Peninsula. The origin of this haplogroup is from the Middle Eastern area known as the Fertile Crescent, comprising Palestine, Jordan, Syria, Lebanon, and Iraq around 42.9KYA (Semino et al., 2004). Merchants from the Arabian Peninsula brought this genetic marker along with Silk Road to East Asia (Mahal and Matsoukas, 2018). Haplogroup J was only observed in several groups from the extreme northwest of China, which supports the Silk Road spread phenomenon. Haplogroup R originated from the north of Asia about 27 KYA. It is the most frequent haplogroup in Europe and Russia and, in some parts, it is at a frequency of 80%. Some researchers suggest that one of its branches originated in the Kurgan culture (Smolenyak and Turner, 2004). In the Qiang population, the frequency of haplogroup R was 8%. Previous studies (Su et al., 2000b; Underhill et al., 2001; Tajima et al., 2002; Hammer et al., 2006) have indicated that paternal haplogroup D originated in Central Asia. According to Hammer et al. (2006), haplogroup D

FIGURE 2
Median-joining network of the predicted haplogroups in the Qiang people and haplogroup frequency distribution.

originated between Tibet and the Altai mountains. They also speculated that multiple waves of human migrations swept into Eastern Eurasia. Haplogroup D, which is widely distributed across East Asia including most Tibeto-Burman-, Tai-Kadai-, and Hmong-Mien-speaking populations. This haplogroup was predicted to be 36% of the Qiang population. According to Zhong et al. (2011), haplogroup D is present in 2.49% of Han Chinese men. The frequencies of this haplogroup seem to be higher than average in the northern and western regions of China (8.9% in Shaanxi Han, 5.9% in Gansu Han, 4.4% in Yunnan Han, 3.7% in Guangxi Han, 3.3% in Hunan Han, and 3.2% in Sichuan Han). According to The Genographic Consortium et al. (The Genographic Consortium, Yan et al., 2011), the frequency of haplogroup D in the Han population from Eastern China (Jiangsu, Zhejiang, Shanghai, and Anhui) is 1.94%. The Tibetan population is thought to contain an admixture of two ancient populations represented by two major East Asian Y-chromosome lineages, the O and D haplogroups (Shi et al., 2008). Haplogroup O is the most common haplogroup in China and is prevalent throughout East and Southeast Asia. The frequency of haplogroup O in the Chinese Han populations ranges from 29.7 (Han from Guangxi) (The Genographic Consortium, Gan et al., 2008) to 74.3% (Han from Changting, Fujian) (Wen, 2004). Haplogroup O is the most commonly found haplogroup in most East Asian populations, such as approximately 40% of Manchu, Koreans, and Vietnamese (Hammer et al., 2006), 34% in Filipino men (Jin et al., 2009), 55% in Malaysian men (Su et al., 2000a), 44% in Tibetan men (Wen et al., 2004b), 25% in Indonesian men (Li et al., 2008), and 20% in Japanese men (Hammer et al., 2006). In Qiang's population, its frequency was predicted to be approximately 39%. Haplogroup C has a very wide distribution and might represent one of the earliest settlements in East Asia. Haplogroup C was at a low frequency in the Qiang population.

Fine-scale admixture history and evolutionary frameworks of Qiang people

Overview of the genetic structure of Qiang people inferred from genome-wide SNP data

To further formally identify, describe, and date admixture events and characterize the detailed genetic landscape of Qiang people, we collected population data of 600,000 SNPs in 20 Qiang individuals from the Human Origins array and merged it with 1,609 modern genomes and 377 ancient genomes from 59 eastern Eurasian populations. We first conducted a PCA analysis focused on the Chinese modern and ancient populations, in which ancient populations were projected onto the modern genetic landscape. We replicated five genetic lines with geographically restricted language families (Figure 3). Ancient southern East Asians from Guangxi and Fujian clustered closely with Austronesian, Tai-Kadai, and Hmong-Mien people related to modern northern East Asians. Ancient northern East Asians from the Yellow River Basin grouped closely to northern Hans and their neighbors, and ancient northeastern Asians showed a close relationship with modern Tungusic people. The Qiang samples were localized closely with modern Tibetans but some of them with Han Chinese, which showed the existence of admixture events supporting the theory that Tibetans and Hans were the ancestral populations of Qiang people. The Qiang people possessing a close genetic affinity with Hans were further confirmed in the following formal tests, which can provide a better landscape of population admixture and evolutionary history of the indigenous Qiang people.

The ancestry composition and corresponding admixture proportion of the Qiang people were analyzed using the model-based ADMIXTURE software (Figure 4). We found that 7.5% ancestries were related to Hmong-Mien-speaking
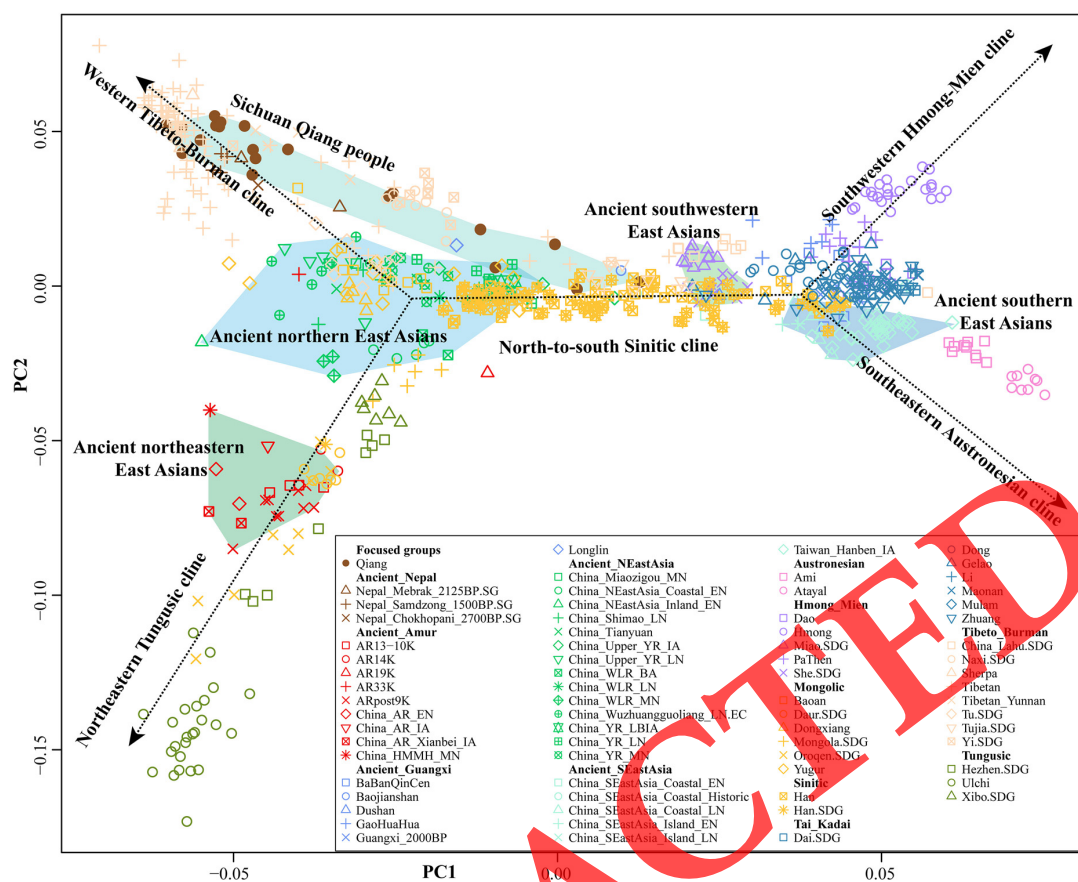
FIGURE 3

Principal component analysis results of 802 East Asian individuals from 76 modern and ancient populations. Each modern population was color-coded by language family, and ancient populations were color-coded by shape.

Hmong, 8.3% were related to Austroasiatic-speaking Htin, 69.6% were related to middle Neolithic Miaozigou people, and 0.146% ancestries maximized in Austronesian-speaking Atayal. Tibetans and ancient Nepalese people shared most ancestry composition of Qiang's gene pool, suggesting their close genetic relationship with Yellow River Basin farmers. These patterns of genetic affinity were further supported by the pairwise qpWave analysis results (**Supplementary Figure 12**). Additional genetic input from southern China related to ancient Guangxi and Fujian people also played an important role in the formation of the Sichuan Qiang people. To further dissect the genetic difference between geographically diverse Qiang people, we merged our data with 1,875 individuals from 231 Eurasian populations and conducted a model-based ADMIXTURE analysis based on the best fitted model with seven ancestral sources (**Figure 5**). We found that Danba Qiang and Daofu Qiang people shared a similar pattern of admixture history and that they derived most of their ancestry from the ancient Nepalese population (83.4∼83%). Similar to the first fitted model, Atayal-related southern East Asians also contributed to the Qiang people's gene pool (15.8∼16.2%).

Pairwise F_{ST} genetic distances can provide direct evidence for genetic differentiation through inter-populations and intra-populations heterozygosities. We first identified the closest genetic relationship between Danba Qiang and Daofu Qiang (F_{ST}: 0.0043, **Supplementary Table 5**). We also identified a close genetic relationship between Chuanxi Qiang and ancient Mongolian and Xiongnu. The estimated F_{ST} values between the Qiang people and the modern East Asian reference populations showed that Qiang had the least genetic distances with Naxi and Yi, followed by Hans. Compared to the ancient populations, we found that the Qiang people had a closer genetic relationship with middle and upper Yellow River Basin people. We also used "outgroup-f_3-statistics" to measure the genetic affinity of geographically different populations and found that two geographically different populations (Daofu and Danba) shared a close genetic relationship, followed by Tibetans, ancient Nepalese, and Neolithic Lajia people (**Figure 6** and **Supplementary Table 6**). Generally, the patterns between Danba Qiang and their reference populations are consistent with the pattern observed and focused on Daofu Qiang. Finally, we calculated the "pairwise outgroup-f_3" value between the

FIGURE 4

Model-based ADMIXTURE results of 1,788 East Asian individuals from 133 modern and ancient populations. Each population was visualized with equal width for better presentation for the population label, which is not associated with population size. K equal to six has the least cross-validation error.

meta-Qiang and their reference populations (Figure 1). We identified consistent patterns of genetic affinity between the Qiang and reference modern and ancient East Asians in the two indexes. The modern Qiang people showed a close genetic relationship with Tibetans and ancient Yellow River Basin farmers who are geographically close to each other.

Formal testing for the admixture process of Qiang people

We calculated "admixture-f_3-statistics" in the form of f_3 (source1, source2; Qiang) to formally explore the formation process of Qiang populations, in which the statistically negative values (Z-scores less than -3) indicated that the targeted populations were a mixed product of the ancestral surrogate related to sources 1 and 2. Most Tibeto-Burman-speaking populations, such as highland Tibetan and Yunnan Lahu, showed no statistically significant signals in the admixture f_3-statistics. Here, we found many source pairs with admixed evidence for Qiang people with one source from northern East Asians (Tibetan and Yellow River Basin farmers) and the other one related to the southern East Asians (Tai-Kadai and Austronesian speakers, Supplementary Table 7). Additionally, we conducted the f_4-statistics in the form of f_4 (reference

1, reference 2; Qiang, Mbuti) to test the asymmetrical or symmetrical sharing ancestry signals between the Qiang people and modern and ancient East Asians. Among the modern reference populations, we found that the Qiang people share most alleles with Yi, Naxi, Tibetans, and Hans compared with the other reference populations. When compared with ancient East Asians from southern East Asia or Siberia, Qiang shares more alleles with middle and late Neolithic people from the Yellow River Basin (Figure 7). Finally, we performed f_4-statistics in the form of f_4 (reference 1, Qiang; reference 2, Mbuti) and found many signals generating negative values with reference 2 for modern Sino-Tibetan people or ancient Yellow River farmers, suggesting that additional gene flow from these populations contributed to modern Qiang people (Supplementary Figure 13). Generally, our f_3/f_4 analyses showed that the Qiang people were a mixed population with ancestral sources derived from their northern and southern neighbors.

We formally estimated the ancestry source composition and admixture coefficient of Sichuan Qiang people and their geographically close populations. Considering the identified two-way admixture signatures from northern and southern China, we first used the two-way admixture qpAdm models to
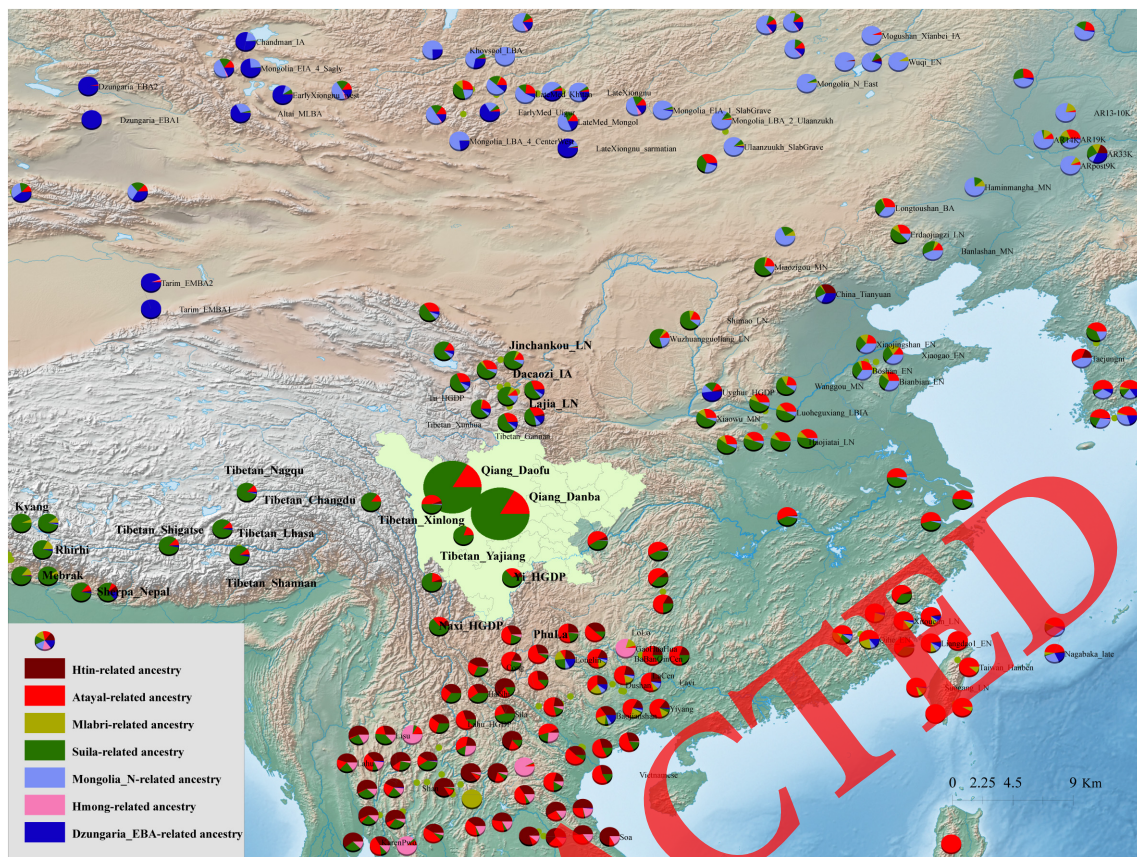
FIGURE 5
Ancestry composition of 233 modern and ancient Eurasian populations based on the model-based ADMIXTURE model. We used the merged HO and Illumina datasets here. K equal to six has the least cross-validation error.
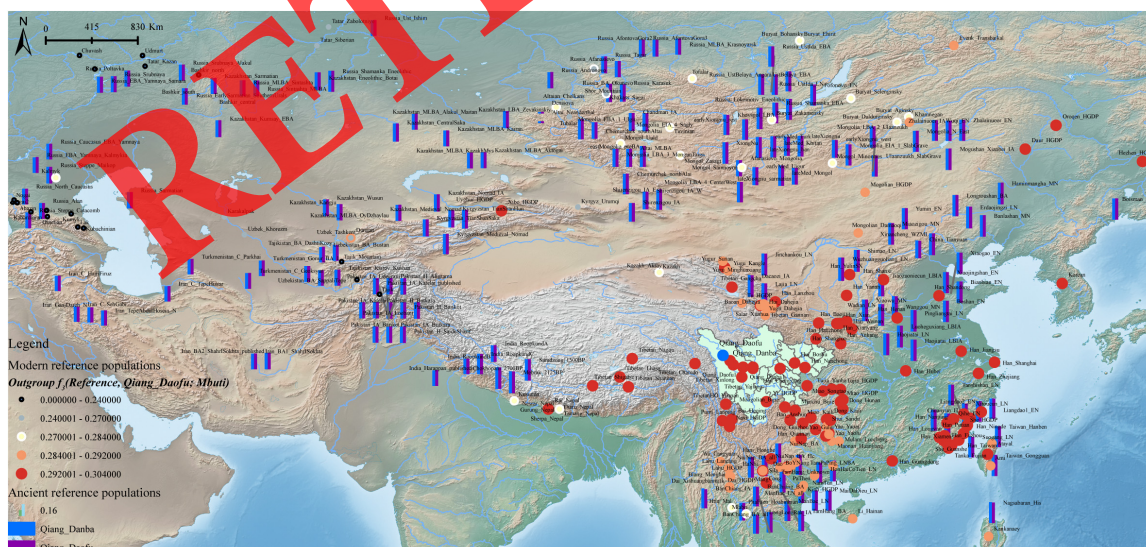

FIGURE 6
Shared genetic drift between Danba and Daofu Qiang people and Eurasian reference populations. Bar showed the genetic affinity between the Qiang people and the ancient Eurasian populations. Circle size and color showed the genetic affinity between Daofu Qiang and modern reference populations.

FIGURE 7

Results from f_4-statistics showed the genetic affinity between Qiang and northern East Asians. Statistically significant population pairs were marked with stars, and the red color showed that Qiang had a close relationship with the right populations compared to the bottom populations. Blue color with negative values means a close genetic relationship with bottom populations compared with the right populations.

portray their ancestry admixture processes using Early Neolithic people from Amur River Basin (China_AR_EN), Neolithic people from the West Liao River Basin (China_WLR_LN), and Bronze/Iron Age populations from the Yellow River Basin (China_YR_LBIA) as the northern sources and using Neolithic people from Fujian and Guangxi and historic people from Guangxi as the potential southern sources. Our spatiotemporal analysis focused on the admixture process. We found that the Qiang and other northern Tibeto-Burman-speaking populations (Tibetan and Sherpa) derived their primary ancestry from the ancient northern East Asians in the predefined two-way admixture models. Southern Tibeto-Burman people harbored more ancestry related to the southern East Asians, such as Sila. It was modeled as an admixture result of 0.297 ancestries related to Yellow River farmers and the remaining ancestry appeared to be

the BaBanQinCen, which is consistent with the admixture patterns observed in Cong people (**Figure 8A**). To further validate whether both inland and coastal southern East Asians contributed to the formation of the Sichuan Qiang people, we used a three-way admixture model with China_AR_EN as the northern source, Baojianshan as the inland southern source, and Hanben_IA as the coastal source. We found that the Qiang people could be modeled as an admixture of the 0.0564 ± 0.095 ancestry from China_AR_EN, the 0.102 ± 0.08 ancestry from Baojianshan, and the 0.334 ± 0.165 ancestry from Hanben (**Figure 8B**). The admixture models could fit the formation of the Naxi people and Yunnan Tibetans well. Finally, we also tested other models with different north-south source pairs for the Qiang people and confirmed that the major ancestry of the Qiang people descended from ancient northern East Asians (**Figure 8C**).
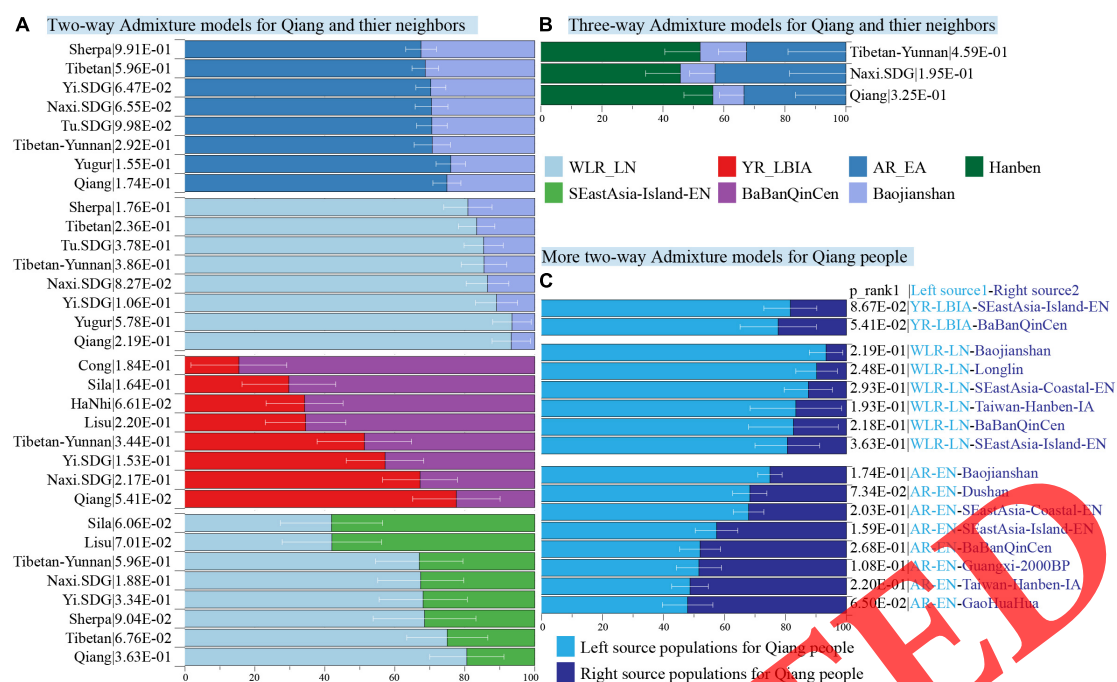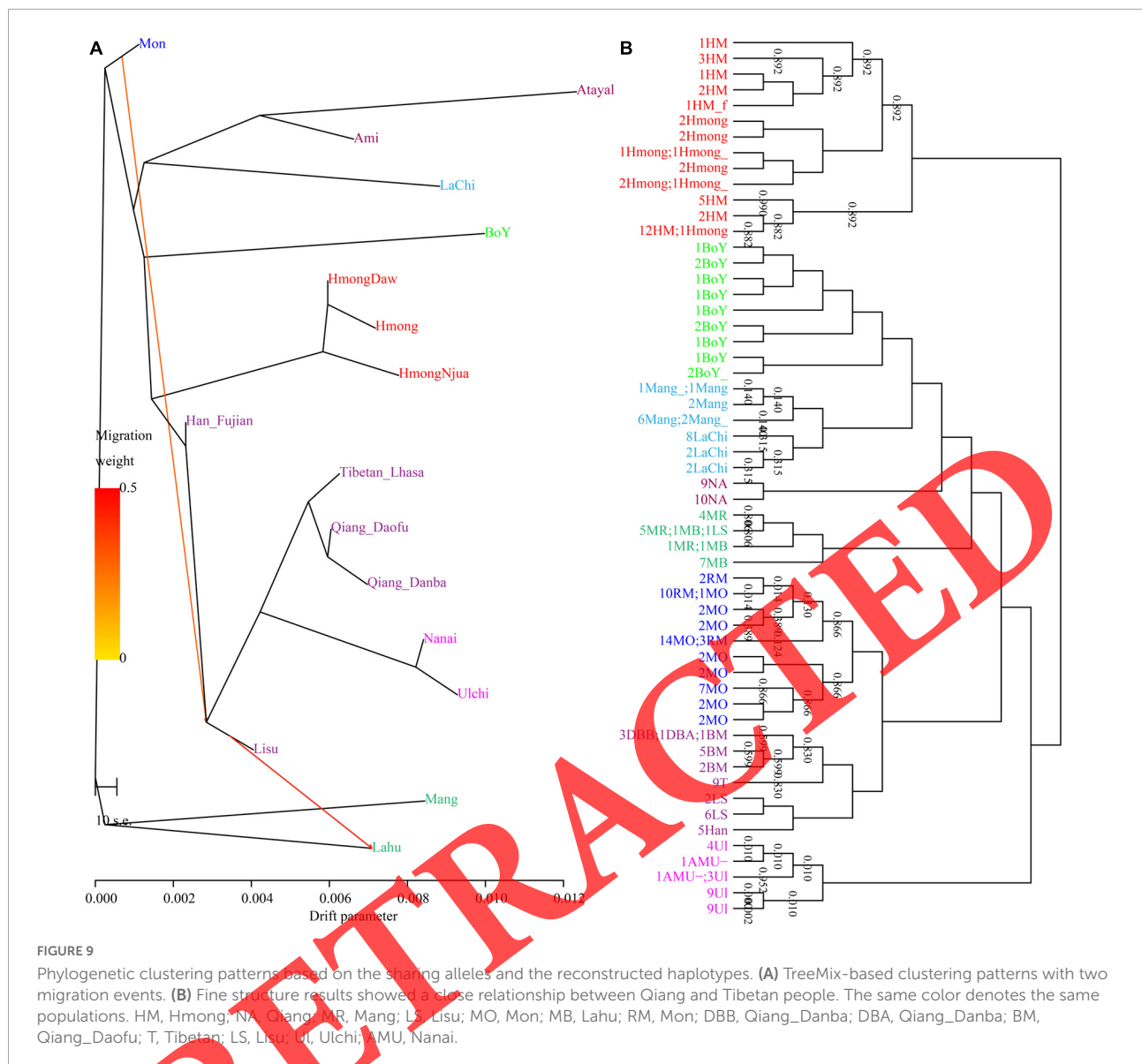
FIGURE 8

Results of the qpAdm-based models show the formation of modern Tibeto-Burman-speaking populations with different admixture models. **(A)** QpAdm models of Tibeto-Burman people with two source surrogates from northern and southern China. **(B)** Three-way qpAdm admixture models of Tibeto-Burman people with sources from inland and coastal southern East Asians and one northern East Asian source. **(C)** Two-way admixture models with different sources fitted the gene pool of Qiang people. Error bar denoted the standard error, and different colors showed different ancestral sources. WLR, West Liao River; LN, Late Neolithic; YR, Yellow River; LBIA, Late Bronze Age and Iron Age; AR, Amur River; EA, Early Neolithic Age; EN, Early Age; IA, Iron Age.

The length and number of ancestral chromosome fragments document more traces of the human population in the process of evolutionary history under different driving forces, including natural selection, mutation, drift, and recombination. We subsequently conducted a population genetic analysis based on phased haplotypes (**Figure 9**). The shared IBD showed that the Qiang people had the closest genetic relationship with modern Naxi, Yi, and Tibetans (**Figure 9A**). The PCA based on the co-ancestry matrix showed a similar cluster pattern, which is also consistent with the identified genetic affinity by allele sharing in the *f*-statistics (**Figure 9B**). The fine-scale population substructure based on the genetic similarity of re-classified homogeneous populations also confirmed the genetic affinity between Qiang and other lowland Tibeto-Burman speakers. Finally, we painted the targeted chromosomes of the Qiang people using both northern and southern East Asians as potential ancestral sources (surrogate populations) and the ChromoPainter-based copy vector. We identified the admixture model of one-date-multiway as the best model guesses for the Sichuan Qiang people. GLOBETROTTER-based admixture dates showed that the Qiang people underwent multiple admixture events in historic times with the surrounding populations, a northern source related to Tibetans and southern sources related to southwestern indigenes. Lastly, we reconstructed the phylogenetic relationship among Chinese populations (**Figures 10A,B**). Consistent with the population clustering patterns in the PCA results (**Figure 10B**), the reconstructed phylogenetic relationship showed that the Qiang people clustered closely with northern East Asians related to Tibetan, Sherpa, and Tu people (**Figure 10A**), suggesting their common population history.
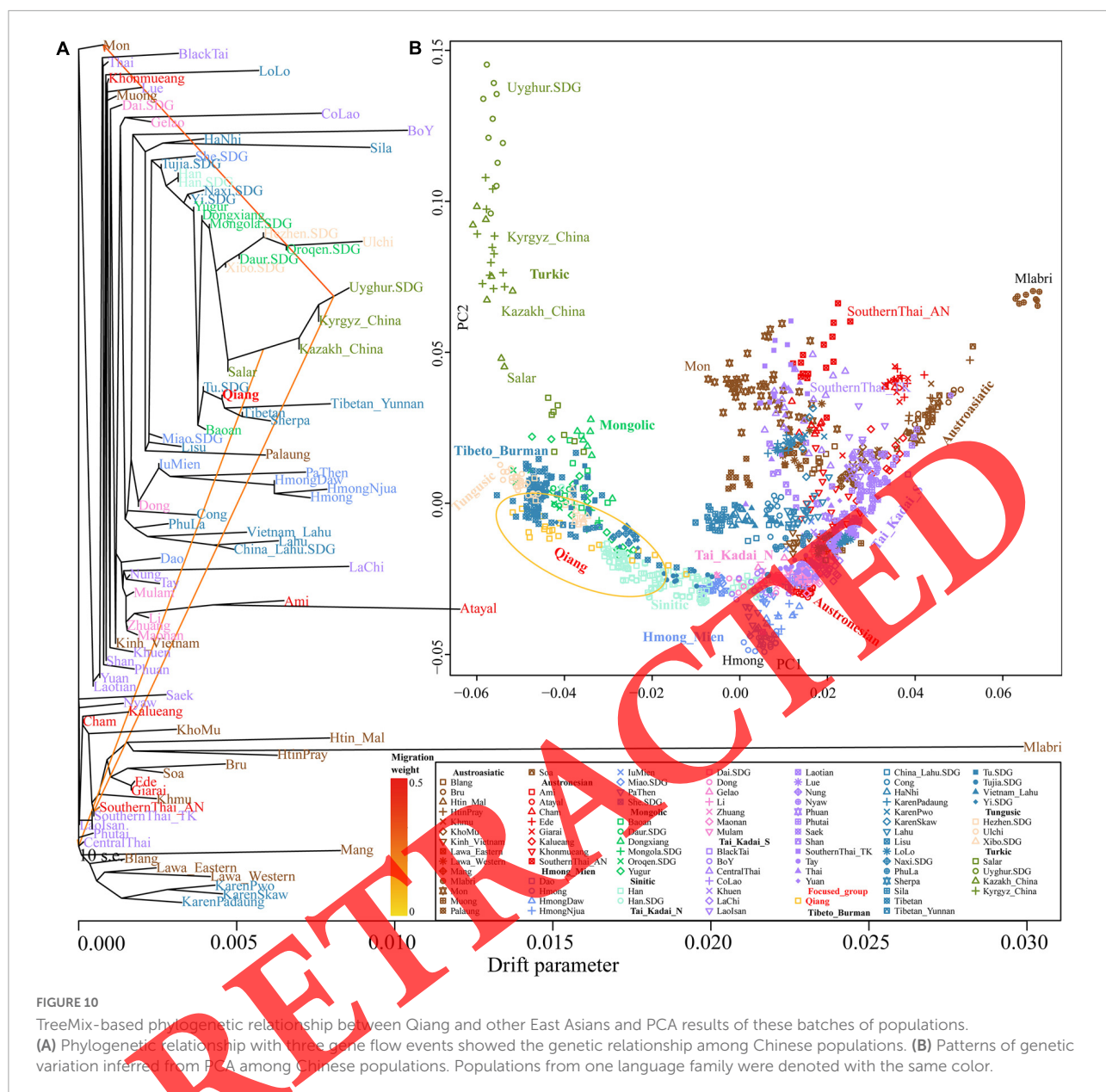
Discussion

The origin and admixture history of the Qiang people has been a mystery in China. Culturally documented evidence supported the common origins of Tibetan and Qiang people; some also suggested a close relationship between the Qiang and Han people. However, no population genetic study focused on the fine-scale population history of Qiang populations has been reported. In this study, first, we developed and comprehensively validated a 30-plexY-STR multiplex assay and found that it was robust enough for forensic genetics applications. Second, we genotyped the Qiang population for use as forensic genetic reference data. We identified high diversity and high efficacy for forensic application. The population genetic analysis based on the Y-STR data showed a close relationship between the Qiang and Tibetan people. Third, we performed a comprehensive

FIGURE 9
Phylogenetic clustering patterns based on the sharing alleles and the reconstructed haplotypes. **(A)** TreeMix-based clustering patterns with two migration events. **(B)** Fine structure results showed a close relationship between Qiang and Tibetan people. The same color denotes the same populations. HM, Hmong; NA, Qiang; MR, Mang; LS, Lisu; MO, Mon; MB, Lahu; RM, Mon; DBB, Qiang_Danba; DBA, Qiang_Danba; BM, Qiang_Daofu; T, Tibetan; LS, Lisu; Ul, Ulchi; AMU, Nanai.

population genomic analysis focused on the genetic structure and admixture history of modern Qiang people.

The findings from PCA and ADMIXTURE analyses showed that the Qiang people clustered closely to the geographically close Yi, Tu, and Tibetans. A previous genetic analysis has shown that Tibetan populations harbored the pre-Neolithic ancestry related to the early Asian and additional Neolithic ancestries from northern Chinese millet farmers (Lu et al., 2016). Geographically diverse Tibetans also showed a differentiated population structure, such as Tibetans in the Tibetan-Yi corridor showing more ancestry related to southern East Asians (He et al., 2021c; Liu et al., 2021b). We also identified different genetic structures between the Qiang and highland Tibetans. The Qiang people in Sichuan province showed a mixed landscape with major ancestry from Yellow River farmers and minor ancestry

related to southern East Asians (Dai and Atayal), suggesting that Qiang possessed more genetic influence from southern Chinese populations. Our ADMIXTURE and qpAdm results consistently supported that most of the ancestry of the Qiang people was derived from northern East Asians, which was consistent with the common origin of Qiang, Tibetan, and Han Chinese populations from the Yellow River Basin in North China. The direct evidence for the close relationship between Qiang and Neolithic millet farmers was obtained from the identified closest relationship with Qiang and Late Neolithic Lajia and Jinchankou people from the Ganqing region by shared genetic drift testing. We also identified a close genetic relationship between Qiang and historic Xiongnu and Mongolians, which was consistent with the ancient populations residing in wide regions of northwest China who contributed genetic materials

FIGURE 10
TreeMix-based phylogenetic relationship between Qiang and other East Asians and PCA results of these batches of populations.
(A) Phylogenetic relationship with three gene flow events showed the genetic relationship among Chinese populations. **(B)** Patterns of genetic variation inferred from PCA among Chinese populations. Populations from one language family were denoted with the same color.

to modern Qiang people. We also noted that the genetic history of only two geographically different Qiang populations was explored, and we suggested that more genome-wide SNP data from geographically different regions can provide deeper and more comprehensive insights into the complex landscape of the Qiang people.

Conclusion

This study first performed a developmental validation study focused on the AGCU Database Y30 kit and population genetic analyses based on Y-STR diversity and genome-wide SNP variation. The study showed that the Y30 kit was a robust assay for forensic caseworks and can be used as a powerful tool for forensic identification and population genetics research. The population genetic analysis based on the Y-STR and genome-wide SNP data showed that the Qiang people had the closest genetic affinity with the geographically close Tibeto-Burman-speaking Naxi, Yi, and Tibetan populations. Further evidence from shared ancestry in the f-statistics and qpAdm-based admixture models showed that the Qiang people derived most of their ancestry from Neolithic Yellow River millet farmers, and that the remaining ancestry was derived from the Neolithic-to-historic Guangxi people. This is consistent with the reconstructed population evolution of modern Tibetan people based on modern and ancient people. Furthermore, we also identified obvious recent admixture

models in the one-date-multiway model based on phasing modern chromosome painting. This study supported the view that Qiang people originated from millet farmers in the Yellow River Basin and underwent continuous gene flow events from southern East Asians from ancient to the historic time period.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

This study was approved by the Ethical Review Board (Dated 20th March 2019 with approval reference no. 2019-86-P) of the China Medical University, and Xiamen University (XDYX2019009). Written informed consent to participate in this study was provided by the participants or their legal guardian/next of kin.

Author contributions

GH, AA, C-CW, MW, CL, and JY designed the study. GH, AA, C-CW, and MW wrote the manuscript. GH, AA, WA-Q, FS, H-YY, and SH collected the samples, conducted the experiment, and analyzed the data. All authors reviewed the manuscript.

Funding

Acknowledgments

Conflict of interest

The 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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fevo.2022.939659/full#supplementary-material

References

Adnan, A., Rakha, A., Kasim, K., Noor, A., Nazir, S., Hadi, S., et al. (2018). Genetic characterization of Y-chromosomal STRs in Hazara ethnic group of Pakistan and confirmation of DYS448 null allele. *Int. J. Legal Med.* 133, 789–793. doi: 10.1007/s00414-018-1962-x

Adnan, A., Ralf, A., Rakha, A., Kousouri, N., and Kayser, M. (2016). Improving empirical evidence on differentiating closely related men with RM Y-STRs: A comprehensive pedigree study from Pakistan. *Forensic Sci. Int. Genet.* 25, 45–51. doi: 10.1016/j.fsigen.2016.07.005

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664.

Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., et al. (2010). Mutability of Y-chromosomal microsatellites: Rates, characteristics, molecular bases, and forensic implications. *Am. J. Hum. Genet.* 87, 341–353. doi: 10.1016/j.ajhg.2010.08.006

Ballantyne, K. N., Keerl, V., Wollstein, A., Choi, Y., Zuniga, S. B., Ralf, A., et al. (2012). A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci. Int. Genet.* 6, 208–218. doi: 10.1016/j.fsigen.2011.04.017

Browning, B. L., and Browning, S. R. (2011). A fast, powerful method for detecting identity by descent. *Am. J. Hum. Genet.* 88, 173–182.

Browning, B. L., and Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* 194, 459–471. doi: 10.1534/genetics.113.150029

Calacal, G. C., Delfin, F. C., Tan, M. M. M., Roewer, L., Magtanong, D. L., Lara, M. C., et al. (2005). Identification of exhumed remains of fire tragedy victims using conventional methods and autosomal/Y-chromosomal short tandem repeat DNA profiling. *Am. J. Forensic Med. Pathol.* 26, 285–291. doi: 10.1097/01.paf.0000177338.21951.82

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8

Excoffier, L., and Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

Fu, Q., Hajdinjak, M., Moldovan, O. T., Constantin, S., Mallick, S., Skoglund, P., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524, 216–219. doi: 10.1038/nature14558

Gopinath, S., Zhong, C., Nguyen, V., Ge, J., Lagacé, R. E., Short, M. L., et al. (2016). Developmental validation of the Yfiler§ Plus PCR amplification kit: An enhanced Y-STR multiplex for casework and database applications. *Forensic Sci. Int.* 24, 164–175. doi: 10.1016/j.fsigen.2016.07.006

Guo, J. T. J. J., Ji si, R. k. h. s. h. k. j. t., Wei, G. j. m., and Zhansi, J. j. f. (2003). *2000 nian ren koupu cha Zhong guo min zu ren kouziliao = Tabulation on nationalities of 2000 population census of China.* Beijing: Min zu chu ban she.

Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M., et al. (2006). Dual origins of the Japanese: Common ground for hunter-gatherer and farmer Y chromosomes. *J. Hum. Genet.* 51, 47–58. doi: 10.1007/s10038-005-0322-0

He, G., Adnan, A., Rakha, A., Yeh, H.-Y., Wang, M., Zou, X., et al. (2019). A comprehensive exploration of the genetic legacy and forensic features of Afghanistan and Pakistan Mongolian-descent Hazara. *Forensic Sci. Int.* 42:e1–e12. doi: 10.1016/j.fsigen.2019.06.018

He, G., Li, Y., Wang, M., Zou, X., Yeh, H., Yang, X., et al. (2021a). Fine-scale genetic structure of Tujia and central Han Chinese revealing massive genetic admixture under language borrowing. *J. Syst. Evol* 59, 1–20. doi: 10.1111/jse.12670

He, G., Wang, M., Li, Y., Zou, X., Yeh, H., Tang, R., et al. (2021b). Fine-scale north-to-south genetic admixture profile in Shaanxi Han Chinese revealed by genome-wide demographic history reconstruction. *J. Syst. Evol* 4, 955–972. doi: 10.1111/jse.12715

He, G., Wang, M., Zou, X., Chen, P., Wang, Z., Liu, Y., et al. (2021c). 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. doi: 10.3389/fgene.2021.725243

He, G., Wang, M., Zou, X., Yeh, H.-Y., Liu, C., Liu, C., et al. (2022b). Extensive ethnolinguistic diversity at the crossroads of North China and South Siberia reflects multiple sources of genetic diversity. *J. Syst. Evol.* doi: 10.1111/jse.12827

He, G., Li, Y., Zou, X., Yeh, H., Tang, R., Wang, P., et al. (2022a). The northern gene flow into southeastern East Asians inferred from genome-wide array genotyping. *J. Syst. Evol.* jse.12826. doi: 10.1111/jse.12826

Hellenthal, G., Busby, G. B. J., Band, G., Wilson, J. F., Capelli, C., Falush, D., et al. (2014). A genetic atlas of human admixture history. *Science* 343, 747–751. doi: 10.1126/science.1243518

Jeong, C., Ozga, A. T., Witonsky, D. B., Malmström, H., Edlund, H., Hofman, C. A., et al. (2016). Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7485–7490. doi: 10.1073/pnas.1520844113

Jin, H.-J., Tyler-Smith, C., and Kim, W. (2009). The peopling of Korea revealed by analyses of mitochondrial DNA and Y-chromosomal markers. *PLoS One* 4:e4210. doi: 10.1371/journal.pone.0004210

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Kutanan, W., Liu, D., Kampuansai, J., Srikummool, M., Srithawong, S., Shoocongdej, R., 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, 3459–3477. doi: 10.1093/molbev/msab124

Lachance, J., Vernot, B., Elbers, C. C., Ferwerda, B., Froment, A., Bodo, J.-M., et al. (2012). Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* 150, 457–469. doi: 10.1016/j.cell.2012.07.009

Li, H., Wen, B., Chen, S.-J., Su, B., Pramoonjago, P., Liu, Y., et al. (2008). Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evol. Biol.* 8:146. doi: 10.1186/1471-2148-8-146

Li, L., Huang, P., Sun, X., Wang, S., Xu, M., Liu, S., et al. (2021). The ChinaMAP reference panel for the accurate genotype imputation in Chinese populations. *Cell Res.* 31, 1308–1310. doi: 10.1038/s41422-021-00564-z

Liu, D., Duong, N. T., Ton, N. D., Van Phong, N., Pakendorf, B., Van Hai, N., et al. (2020). Extensive ethnolinguistic diversity in vietnam reflects multiple

sources of genetic diversity. *Mol. Biol. Evol.* 37, 2503–2519. doi: 10.1093/molbev/msaa099

Liu, Y., Wang, M., Chen, P., Wang, Z., Liu, J., Yao, L., et al. (2021a). Combined low-/high-density modern and ancient genome-wide data document genomic admixture history of high-altitude east Asians. *Front. Genet.* 12:582357. doi: 10.3389/fgene.2021.582357

Liu, Y., Yang, J., Li, Y., Tang, R., Yuan, D., Wang, Y., et al. (2021b). Significant east asian affinity of the Sichuan Hui genomic structure suggests the predominance of the cultural diffusion model in the genetic formation process. *Front. Genet.* 12:626710. doi: 10.3389/fgene.2021.626710

Lu, D., Lou, H., Yuan, K., Wang, X., Wang, Y., Zhang, C., et al. (2016). Ancestral origins and genetic history of Tibetan highlanders. *Am. J. Hum. Genet.* 99, 580–594. doi: 10.1016/j.ajhg.2016.07.002

Mahal, D. G., and Matsoukas, I. G. (2018). The geographic origins of ethnic groups in the Indian subcontinent: Exploring ancient footprints with Y-DNA haplogroups. *Front. Genet.* 9:4. doi: 10.3389/fgene.2018.00004

Mao, X., Zhang, H., Qiao, S., Liu, Y., Chang, F., Xie, P., et al. (2021). The deep population history of northern East Asia from the late pleistocene to the holocene. *Cell* 184, 3256–3266.e13. doi: 10.1016/j.cell.2021.04.040

Meng, H., Guo, Y., Jin, X., Chen, C., Cui, W., Shi, J., et al. (2019). Internal validation study of a newly developed 24-plex Y-STRs genotyping system for forensic application. *Int. J. Legal Med.* 133, 733–743. doi: 10.1007/s00414-019-02028-x

Mo, X.-T., Zhang, J., Ma, W.-H., Bai, X., Li, W.-S., Zhao, X. C., et al. (2019). Developmental validation of the DNATyper™ Y26 PCR amplification kit: An enhanced Y-STR multiplex for familial searching. *Forensic Sci. Int* 38, 113–120. doi: 10.1016/j.fsigen.2018.10.008

Mulero, J. J., Chang, C. W., Calandro, L. M., Green, R. L., Li, Y., Johnson, C. L., et al. (2006). Development and validation of the AmpFlSTRRYfilertm PCR Amplification Kit: A male specific, single amplification 17 Y-STR multiplex System. *J. Forensic Sci.* 51, 64–75. doi: 10.1111/j.1556-4029.2005.00016.x

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

Plummer, M. (2003). *JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling.* Available online at: https://cran.r-project.org/web/packages/rjags/index.html (accessed April 19, 2022).

Prinz, M., Boll, K., Baum, H., and Shaler, B. (1997). Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Sci. Int.* 85, 209–218. doi: 10.1016/S0379-0738(96)02096-8

Rakha, A., Oh, Y. N., Lee, H. Y., Hussain, S., Waryah, A. M., Adnan, A., et al. (2018). Discriminating power of rapidly mutating Y-STRs in deep rooted endogamous pedigrees from Sindhi population of Pakistan. *Legal Med.* 34, 17–20. doi: 10.1016/j.legalmed.2018.08.001

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., and Singh, L. (2009). Reconstructing Indian population history. *Nature* 461, 489–494. doi: 10.1038/nature08365

Roewer, L., Amemann, J., Spurr, N. K., Grzeschik, K.-H., and Epplen, J. T. (1992). Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum. Genet.* 89, 389–394. doi: 10.1007/BF00194309

Roewer, L., and Epplen, J. T. (1992). Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Sci. Int.* 53, 163–171. doi: 10.1016/0379-0738(92)90193-Z

Semino, O., Magri, C., Benuzzi, G., Lin, A. A., Al-Zahery, N., Battaglia, V., et al. (2004). Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: Inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am. J. Hum. Genet.* 74, 1023–1034. doi: 10.1086/386295

Shi, H., Zhong, H., Peng, Y., Dong, Y., Qi, X., Zhang, F., 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. doi: 10.1186/1741-7007-6-45

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139, 457–462.

Smolenyak, M., and Turner, A. (2004). *Trace your roots with DNA: Using genetic tests to explore your family tree.* Emmaus, Pa: Rodale.

Su, B., Xiao, C., Deka, R., Seielstad, M. T., Kangwanpong, D., Xiao, J., et al. (2000b). Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* 107, 582–590. doi: 10.1007/s004390000406

Su, B., Jin, L., Underhill, P., Martinson, J., Saha, N., McGarvey, S. T., et al. (2000a). Polynesian origins: Insights from the Y chromosome. *Proc. Natl. Acad. Sci. U.S.A* 97, 8225–8228. doi: 10.1073/pnas.97.15.8225

Tajima, A., Pan, I.-H., Fucharoen, G., Fucharoen, S., Matsuo, M., Tokunaga, K., et al. (2002). Three major lineages of Asian Y chromosomes: Implications for the peopling of east and southeast Asia. *Hum. Genet.* 110, 80–88. doi: 10.1007/s00439-001-0651-9

The Genographic Consortium, Gan, R.-J., Pan, S.-L., Mustavich, L. F., Qin, Z.-D., Cai, X.-Y., et al. (2008). Pinghua population as an exception of Han Chinese's coherent genetic structure. *J. Hum. Genet.* 53, 303–313. doi: 10.1007/s10038-008-0250-x

The Genographic Consortium, Yan, S., Wang, C.-C., Li, H., Li, S.-L., and Jin, L. (2011). An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *Eur. J. Hum. Genet.* 19, 1013–1015. doi: 10.1038/ejhg.2011.64

Thompson, J. M., Ewing, M. M., Frank, W. E., Pogemiller, J. J., Nolde, C. A., Koehler, D. J., et al. (2013). Developmental validation of the PowerPlex§ Y23 system: A single multiplex Y-STR analysis system for casework and database samples. *Forensic Sci. Int.* 7, 240–250. doi: 10.1016/j.fsigen.2012.10.013

Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon Lahr, M., Foley, R. A., et al. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65, 43–62. doi: 10.1046/j.1469-1809.2001.6510043.x

Wang, C. C., Yeh, H.-Y., Popov, A. N., Zhang, H.-Q., Matsumura, H., Sirak, K., et al. (2021). insights into the formation of human populations in East Asia. *Nature* 591, 413–419. doi: 10.1038/s41586-021-03336-2

Wang, F. (2013). On the oracle bone inscriptions "Qiang". *J. Qilu Normal Univ.* 28, 68–71.

Wang, M., Du, W., Tang, R., Liu, Y., Zou, X., Yuan, D., et al. (2022). Genomic history and forensic characteristics of Sherpa highlanders on the Tibetan Plateau inferred from high-resolution InDel panel and genome-wide SNPs. *Forensic Sci. Int.* 56, 102633. doi: 10.1016/j.fsigen.2021.102633

Wang, M., Yuan, D., Zou, X., Wang, Z., Yeh, H.-Y., Liu, J., 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. doi: 10.3389/fgene.2021.727821

Wang, T., Wang, W., Xie, G., Li, Z., Fan, X., Yang, Q., et al. (2021). Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. *Cell* 184, 3829–3841.e21. doi: 10.1016/j.cell.2021.05.018

Wen, B. (2004). The origin of Mosuo people as revealed by mtDNA and Y chromosome variation. *Sci. China Ser. C* 47:1. doi: 10.1360/02yc0207

Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y., et al. (2004a). Genetic evidence supports demic diffusion of Han culture. *Nature* 431, 302–305. doi: 10.1038/nature02878

Wen, B., Xie, X., Gao, S., Li, H., Shi, H., Song, X., et al. (2004b). Analyses of genetic structure of Tibeto-burman populations reveals sex-biased admixture in Southern Tibeto-Burmans. *Am. J. Hum. Genet.* 74, 856–865. doi: 10.1086/386292

Yang, M. A., Fan, X., Sun, B., Chen, C., Lang, J., Ko, Y.-C., et al. (2020). Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* 369, 282–288. doi: 10.1126/science.aba0909

Yao, H., Wang, M., Zou, X., Li, Y., Yang, X., Li, A., 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, 631–651. doi: 10.1007/s00438-021-01767-0

Zhong, H., Shi, H., Qi, X.-B., Duan, Z.-Y., Tan, P.-P., Jin, L., et al. (2011). Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol. Biol. Evol.* 28, 717–727. doi: 10.1093/molbev/msq247