



Fine-Scale Population Admixture Landscape of Tai–Kadai-Speaking Maonan in Southwest China Inferred From Genome-Wide SNP Data

Jing Chen^{1†}, Guanglin He^{2,3,4,5†}, Zheng Ren¹, Qiyang Wang¹, Yubo Liu¹, Hongling Zhang¹, Meiqing Yang¹, Han Zhang¹, Jingyan Ji¹, Jing Zhao^{2,3,4}, Jianxin Guo^{2,3,4}, Jinwen Chen^{2,3,4}, Kongyang Zhu^{2,3,4}, Xiaomin Yang^{2,3,4}, Rui Wang^{2,3,4}, Hao Ma^{2,3,4}, Le Tao^{2,3,4}, Yilan Liu^{2,3,4}, Qu Shen^{2,3,4}, Wenjiao Yang^{2,3,4}, Chuan-Chao Wang^{2,3,4*} and Jiang Huang^{1*}

OPEN ACCESS

Edited by:

Pavel Flegontov,
Harvard University, United States

Reviewed by:

Pittayawat Pittayaporn,
Chulalongkorn University, Thailand
Chao Ning,
Peking University, China

*Correspondence:

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

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 15 November 2021

Accepted: 27 January 2022

Published: 17 February 2022

Citation:

Chen J, He G, Ren Z, Wang Q, Liu Y,
Zhang H, Yang M, Zhang H, Ji J,
Zhao J, Guo J, Chen J, Zhu K, Yang X,
Wang R, Ma H, Tao L, Liu Y, Shen Q,
Yang W,
Wang C-C and Huang J (2022) Fine-
Scale Population Admixture
Landscape of Tai–Kadai-Speaking
Maonan in Southwest China Inferred
From Genome-Wide SNP Data.
Front. Genet. 13:815285.
doi: 10.3389/fgene.2022.815285

¹Department of Forensic Medicine, Guizhou Medical University, Guiyang, China, ²State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen, China, ³Department of Anthropology and Ethnology, School of Sociology and Anthropology, Institute of Anthropology, National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen, China, ⁴State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, China, ⁵Institute Of Rare Diseases, West China Hospital of Sichuan University, Chengdu, China

Guizhou Province harbors extensive ethnolinguistic and cultural diversity with Sino-Tibetan-, Hmong–Mien-, and Tai–Kadai-speaking populations. However, previous genetic analyses mainly focused on the genetic admixture history of the former two linguistic groups. The admixture history of Tai–Kadai-speaking populations in Guizhou needed to be characterized further. Thus, we genotyped genome-wide SNP data from 41 Tai–Kadai-speaking Maonan people and made a comprehensive population genetic analysis to explore their genetic origin and admixture history based on the pattern of the sharing alleles and haplotypes. We found a genetic affinity among geographically different Tai–Kadai-speaking populations, especially for Guizhou Maonan people and reference Maonan from Guangxi. Furthermore, formal tests based on the f_3/f_4 -statistics further identified an adjacent connection between Maonan and geographically adjacent Hmong–Mien and Sino-Tibetan people, which was consistent with their historically documented shared material culture (Zhang et al., *iScience*, 2020, 23, 101032). Fitted qpAdm-based two-way admixture models with ancestral sources from northern and southern East Asians demonstrated that Maonan people were an admixed population with primary ancestry related to Guangxi historical people and a minor proportion of ancestry from Northeast Asians, consistent with their linguistically supported southern China origin. Here, we presented the landscape of genetic structure and diversity of Maonan people and a simple demographic model for their evolutionary process. Further whole-genome-sequence-based projects can be presented with more detailed information about the population history and adaptive history of the Guizhou Maonan people.

Keywords: fine-scale genetic structure, Tai–Kadai-speaking Maonan, admixture history, ethnolinguistic diversity, east Asian

INTRODUCTION

Guizhou Province is localized in the Yunnan–Guizhou Plateau and has a mountainous environment. This region has rich archaeologically attested cultures, historically documented ethnic groups, and languages. Populations belonging to different language families, including Tai–Kadai, Hmong–Mien, and Sino-Tibetan families, permanently resided here. Archaeological findings associated with the Chengtoushan, Daxi, and Shijiahe sites supported that southern China was the original birthplace of rice agriculture and the homelands of many languages. Further direct evidence from both archeology and language supported that the prosperity of rice farming led to the formation of the ancestral populations of present-day Tai–Kadai-, Hmong–Mien-, Austroasiatic-, and Austronesian-speaking populations and their used languages (Diamond and Bellwood, 2003; Zhao, 2011; Stevens and Fuller, 2017). In addition, linguistic evidence not only demonstrated that there was a common origin between Tai–Kadai and Austronesian language but also revealed that the Tai–Kadai language shared more language components (language borrowing) with surrounding Hmong–Mien and Sino-Tibetan families (Edmondson, 1988; Edmondson, 1997; Lu, 2008), which was also confirmed *via* the genome-wide SNP data (He et al., 2020). Importantly, a recent published genetic analysis based on the whole-genome SNP data from Southeast Asians also revealed the complex divergence processes, which showed that Austroasiatic people diverged from mainland Chinese populations approximately 15 thousand years ago (kya), Austronesian people diverged from mainland Sinitic-speaking Han and Tai–Kadai-speaking Dai around 10 kya, and Cordilleran people split from indigenous Taiwanese people at eight kya (Larena et al., 2021). The divergence time between the ancestors of Hmong-Mien and Tai-Kadai people keep unknown, but they experienced massive interaction with each other and with the southward Northeast Asians (Huang et al., 2020; Yang et al., 2020; Wang et al., 2021c; Wang C.C. et al., 2021). The initial landscape of the population history of southern China has been characterized *via* evidence from genetics and linguistics, whose detailed genetic history and admixture process with its neighbors need to be further explored based on the high-density genome-wide SNP data, especially for some geographically and ethnolinguistically diverse and specific indigenous populations.

The linguistic survey has found that the Tai–Kadai language was widely distributed in Southeast Asia, including Zhuang-Dai, Dong-Shui, Li, and Ge-Yang language sub-branches (Edmondson, 1988; Liang and Zhang, 1996; Edmondson, 1997; Kutanan et al., 2019; Liu et al., 2020; Kutanan et al., 2021). Linguistic findings showed that Maonan is a subgroup of the Dong-Shui language, mainly distributed in Guizhou, Guangdong, and Guangxi provinces (Lu, 2008). Historians hold the opinion that Tai–Kadai people in South China are one of the indigenous population with a long history (Wang, 2004; Huang, 2016; Zhang, 2016), which is also evidenced *via* the linguistic documents and ancient DNA evidence (Edmondson, 1988; Liang and Zhang, 1996; Lipson et al., 2018; McColl et al., 2018; Liu et al., 2020; Kutanan et al., 2021). Direct documents from historic materials showed that the Maonan ethnic group is related to the modern

Southern Chinese indigenous ethnic groups such as Bouyei, Mulam, and Gelao (Wang, 2004; Huang, 2016). However, the historical records for the origin of Maonan and the records of local Chronicles inscriptions and genealogies are unknown. From an archeological perspective (Ma et al., 2020), the presumed ancestral populations of modern Tai–Kadai and Hmong–Mien speakers are probably related to Daxi culture and Qujialing culture around the Yunnan–Guizhou Plateau. Therefore, the genetic investigation should be comprehensively carried out in areas with high ethnic and linguistic diversity to explore the genetic connection among modern ethnolinguistically different populations and the genetic interactions between Guizhou indigenous people with ancient northern and southern East Asians (Ning et al., 2020; Yang et al., 2020; Mao et al., 2021; Wang et al., 2021c; Wang C. C. et al., 2021).

Considering the importance of a comprehensive and deep genetic survey of South China, previous genetic studies based on forensic genetic markers have shed light on the basic genetic profile and demographic history among Tai–Kadai speakers from southern China in the past two decades. From the perspective of uniparentally inherited Y chromosome haplogroup and mitochondrial haplogroup, Chen et al. made a preliminary exploration focused on the forensic parameters and genetic structure of the Tai–Kadai-speaking populations in Guizhou and the population genetic relationship based on the short tandem repeat (STR) on the autosome and X/Y-chromosomes (Chen et al., 2018c). Further genetic studies had focused on the population admixture history and genetic diversity of Tai–Kadai-speaking Gelao and Bouyei by insertion/deletion polymorphisms (InDels) and ancestry-informative single-nucleotide polymorphism (AISNPs) (He et al., 2019a; He et al., 2019b; He et al., 2021b). The obtained research results showed a significant genetic interaction between Tai–Kadai- and Hmong–Mien-speaking populations (He et al., 2019a; He et al., 2019b; He et al., 2021b). However, these studies were conducted based on low-density genetic markers and were mainly focused on exploring forensic characteristics. The low resolution of low-density markers was restricted to provide a fine-scale population genetic structure that can show the detailed information for the population admixture, evolutionary, and adaptive history. Recent population genomic history characterization of ethnolinguistically diverse people in Guizhou, including Han, Chuanqing, Gejia, Dongjia, Xijia, Mongolian, Manchu, Bouyei, Sui, Tujia, Dong, and Gelao people, has provided new insights into their formation processes and complete landscape of genetic history (Lu et al., 2020; Chen et al., 2021; Bin et al., 2021; He et al., 2021a; Wang et al., 2021a; Wang et al., 2021b). However, the admixture history of another important Guizhou indigenous Maonan still remains unknown.

Maonan people used the Maonan language belonging to the Dongshui branch in the Tai–Kadai language family (Edmondson, 1988, 1997; Lu, 2008), which was widely distributed in Guizhou and Guangxi provinces. Historians supported that the Maonan people were one of the major descendants of ancient indigenous tribes in coastal southern China, and that they were especially associated with the hanging coffin burial custom (Zhang et al.,

2020). Based on the ancient DNA from the mitochondrial chromosome from historic hanging coffin people in Fujian, Guangxi, ancient people related to Tai–Kadai populations migrated westward from Fujian Wuyi Mount in the historical times and then across Southwest China to Southeast Asia (Zhang et al., 2020). Historical materials based on the hanging coffin customs also showed similar patterns of population migrations. Recent population genetic studies have included the Maonan people in Guangxi as the reference populations in the ancient DNA study. However, their fine-scale population history and their genetic relationship with the surrounding Hmong–Mien and Sino-Tibetan-speaking populations have not been fully characterized, which are especially focused on the sharing of genome-wide haplotype data. Genetic studies of populations from the Yunnan–Guizhou Plateau regions have found enriched genetic diversity and complex mixed population genetic history (He et al., 2019a; He et al., 2019b; Zhang et al., 2019; Liu et al., 2021a; He et al., 2021b; Liu et al., 2021b; Chen et al., 2021). The detailed relationship between Maonan and modern and ancient neighboring populations needs to be characterized further. Thus, we conducted a genetic analysis based on the array-based genotyping of approximately 700 K SNPs in Tai–Kadai-speaking Maonan people in Guizhou to reconstruct its genetic diversity and evolutionary relationship with surrounding populations. Then, we merged our data with modern and ancient available East Asian data to explore their fine-scale population genetic structure and evolutionary history.

MATERIAL AND METHODS

Sample Collection and DNA Preparation

We collected saliva samples from 41 unrelated Maonan individuals in Pingtang County in Guizhou Province, Southwest China (Supplementary Figure S1). Participants whose parents and grandparents are indigenous people and reside in the sampling palaces at least three generations should have non-consanguineous marriage at the same ethnical group. The study was approved by the Medical Ethics Committee of Guizhou Medical University, and the recommendations provided by the revised Helsinki Declaration of 2000 were followed. All the participants signed written informed consent prior to participating in the study. We genotyped the genome-wide SNP data using the Infinium Global Screening Array, which included approximately 700 K SNPs and covered SNPs from the autosome, Y-chromosome, and mitochondrial DNA. We used a similar in-house commonly used standard to conduct quality control filter procedures (Wang et al., 2021b). Then, we merged the genome-wide data of 41 Guizhou Maonan individuals with previously published present-day and ancient East Asian and Southeast Asian populations from Human Origins (HO) and 1240 k datasets included in the Allen Ancient DNA Resource (AADR) and our recently published genome-wide SNP data based on the Illumina platform (Lu et al., 2020; Wang et al., 2020; Wang et al., 2021a; Liu et al., 2021b; Wang et al., 2021b; He et al., 2021c; Chen et al., 2021; Yao et al., 2021).

Principal Component Analysis

We carried out principal component analysis (PCA) via the *smartpca* program of the EIGENSOFT v.6.1.4 package (Patterson et al., 2006) based on the merged Human Origin dataset. All default parameters were used with the additional parameter of `lsqproject: YES`, in which ancient DNA was projected based on the genetic landscape of the modern East Asians from Hmong–Mien, Tai–Kadai, Austronesian, Austroasiatic, and Sino-Tibetan speakers.

ADMIXTURE Analysis

To prune the strong linkage disequilibrium, we first used PLINK tools (Chang et al., 2015) with the additional parameters (`--indep-pairwise 200 25 0.4`) to obtain the unlinked SNP data among Eurasian modern and ancient populations. Model-based clustering analysis was performed via ADMIXTURE (Alexander et al., 2009), and we ran ADMIXTURE with default parameters with the predefined ancestry sources or clusters ranging from $K = 2$ to 20. We assessed an optimal K value based on the lowest cross-validation error values using 10-fold cross-validation with different random seeds.

Admixture- f_3 -Statistics and Outgroup- f_3 -Statistics

We used ADMIXTOOLS (Patterson et al., 2012) to compute f_3 -statistic values and estimate standard errors by a block jackknife and default parameters. We used the *qp3Pop* program of EIGENSOFT to calculate the outgroup- f_3 -statistics in the form $f_3(\text{Population 1, Population 2; Mbuti})$ using the default parameters, and this index was used for evaluating the shared genetic drift between Population 1 and Population 2 since their separation from the outgroup population of Mbuti. Then we also used the *qp3pop* to perform the admixture- f_3 -statistics in the form $f_3(\text{Source 1, Source 2; Targeted population})$ to explore the admixture signals in Maonan samples with different East Asian and Southeast Asian ancestral source candidates. The value with $|Z\text{-score}| > 3$ denoted that Source 1 and Source 2 could generate the potential admixture signal for the target population.

f_4 Statistics

We used the *qpDstat* program in ADMIXTOOLS (Patterson et al., 2012) with default parameters to assess whether W or X harbored more ancestry related to population Y in the $f_4(W, X; Y, \text{Outgroup})$, which can be used to determine the signals and directions of admixture, and the primary source of gene flow to Guizhou Maonan and other modern and ancient reference East Asians.

Pairwise *qpWave* and *qpAdm* Estimation

We used *qpWave/qpAdm* as implemented in the ADMIXTOOLS (Patterson et al., 2012) package with default parameters and estimated standard errors to detect the minimum number of ancestral populations, and quantitatively estimate corresponding admixture proportions. We used ancient Northeast Asian-related ancestry as the northern sources and Guangxi- or Taiwan-related ancestry as the southern sources to perform the two population

qpAdm model. 1500-year-old BaBanQinCen people are the major ancestral southern sources in our admixture models as it was reported as the direct ancestral sources of modern Tai–Kadai people (Wang et al., 2021c). BaBanQinCen was a meta-population, which comprised two individuals from the Balong site (BalongKD10 and BalongKD07), two individuals from the Banda site (BandaKD15 and BandaKD11), one individual from Qinchang (QinchangKD13 and QinchangKD14), and one individual from Cenxun (CenxunKP05). We used the Mbuti, Ust_Ishim, Kostenki14, Papuan, Australian, Mixe, MA1, Jehai, and Tianyuan as outgroups. We also conducted pairwise qpWave analysis among Tai–Kadai, Hmong–Mien, Sinitic, and ancient Guangxi people to explore their genetic homogeneity. Admixture times were estimated using ALDER with the sources from northern and southern East Asia (Loh et al., 2013).

TreeMix

We ran TreeMix version 1.13 (Pickrell and Pritchard, 2012) to infer the patterns of population splits and admixtures between our target populations and multiple ancestral populations. First, we explored the genetic relationship between Maonan and 15 Chinese populations based on the Illumina array, which was also used in the following haplotype-based analysis. Second, we constructed the TreeMix-based phylogenetic tree among 39 populations to explore the genetic relationship with more reference populations.

Haplotype-Based Fine-Scale Population Structure

We used SHAPEIT v2 (Browning and Browning, 2011) to phase the genome-wide data of Maonan and other Chinese populations in Guizhou and the neighboring regions. Then we conducted the ChromoPainter and FineSTRUCTURE analysis (Hellenthal et al., 2014) to explore the coancestry matrix. We also used R packages implemented in the FineSTRUCTURE to perform the PCA analysis and explore the phylogenetic relationship of studied individuals and populations.

Uniparental Haplogroups

Based on this Illumina array, we genotyped the lineage-informative SNPs (LISNPs) in mitochondrial DNA and Y-chromosome. The haplogroup assignment was used as the in-house manuscripts followed by our recent publications (Lu et al., 2020; Wang et al., 2020; Wang et al., 2021a; Liu et al., 2021b; Wang et al., 2021b; He et al., 2021c; Chen et al., 2021; Yao et al., 2021).

RESULTS

General Structure Inferred From ADMIXTURE and PCA

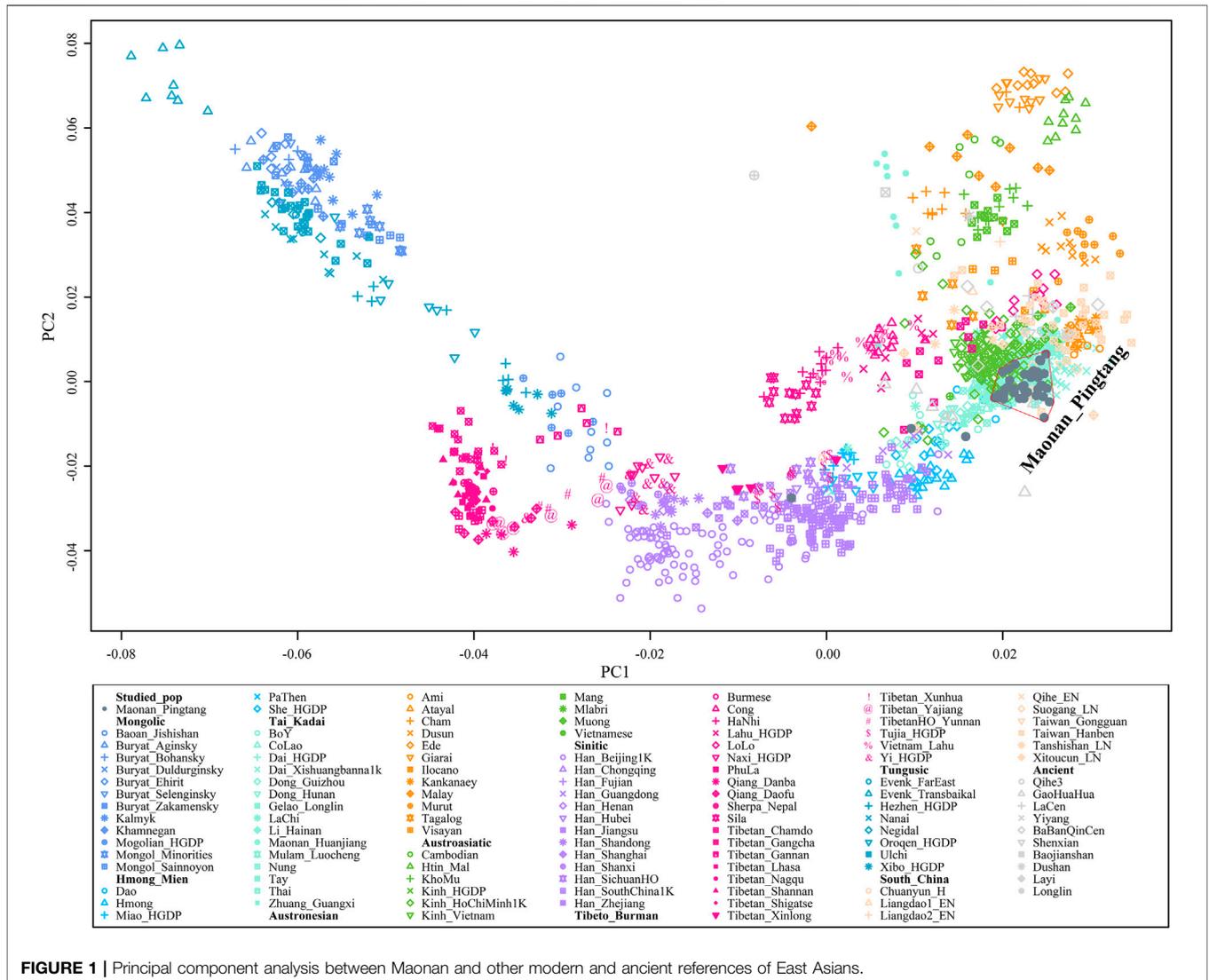
We generated genome-wide data in approximately 700,000 SNPs for 41 Maonan individuals from Guizhou Province, Southwest China. We first merged our data with modern and ancient

published populations from the Human Origins dataset. Then we carried out a principal component analysis (PCA) to understand the general patterns of relatedness between Guizhou Maonan and reference populations (Figure 1). We observed three major genetic clines: the northern cline consisting of Mongolic- and Tungusic-speaking populations, the southern cline comprising Hmong–Mien-, Tai–Kadai-, Austroasiatic-, and Austronesian-speaking populations, and the Sino-Tibetans comprising Sinitic- and Tibeto–Burman-speaking populations, which was located at an intermediate position between the northern cline and the southern cline. We projected publicly available data of ancient individuals from China into modern PC plots. Our studied Tai–Kadai-speaking Maonan population all overlapped with modern Tai–Kadai populations. To gain further insight into the genetic architecture of Guizhou Maonan, we further focused on the genetic backgrounds of Tai–Kadai and other southern modern and ancient East Asians (Supplementary Figure S2). We observed that our studied group partially overlapped with previously published Austroasiatic- and Hmong–Mien-speaking populations.

We carried out the model-based ADMIXTURE clustering analysis to dissect ancestral components and genetic similarity of our studied group with geographically close ancient and present-day populations. We used cross-validation to identify an “optimal” number of clusters ($K = 6$) (Figure 2). At optimal $K = 6$, we observed four specific ancestral components in our studied Pingtang Maonan in Guizhou Province: the ancestry maximizing in this cluster is ubiquitous in modern Sinitic-speaking populations (dark green), with the second component maximized in Austroasiatic-speaking populations (dark blue). The remaining ancestry component was maximized in Hmong–Mien- (dark purple) and Austronesian-speaking populations (dark pink). Hmong–Mien-related ancestry component was maximized in historical GaoHuaHua individuals and found at the highest proportions in Hmong. Austronesian-related ancestry component was maximized in ancient and modern Austronesian Taiwanese with a high proportion in earlier Fujian Neolithic individuals (Taiwan_Hanben/Taiwan_Gongguan) and found at the highest proportions in Atayal. We found that our studied Pingtang Maonan is genetically like the other Tai–Kadai-speaking populations, in which both harbored similar patterns of ancestry components.

Phylogenetic Relationship Sharing Alleles and Sharing Haplotypes

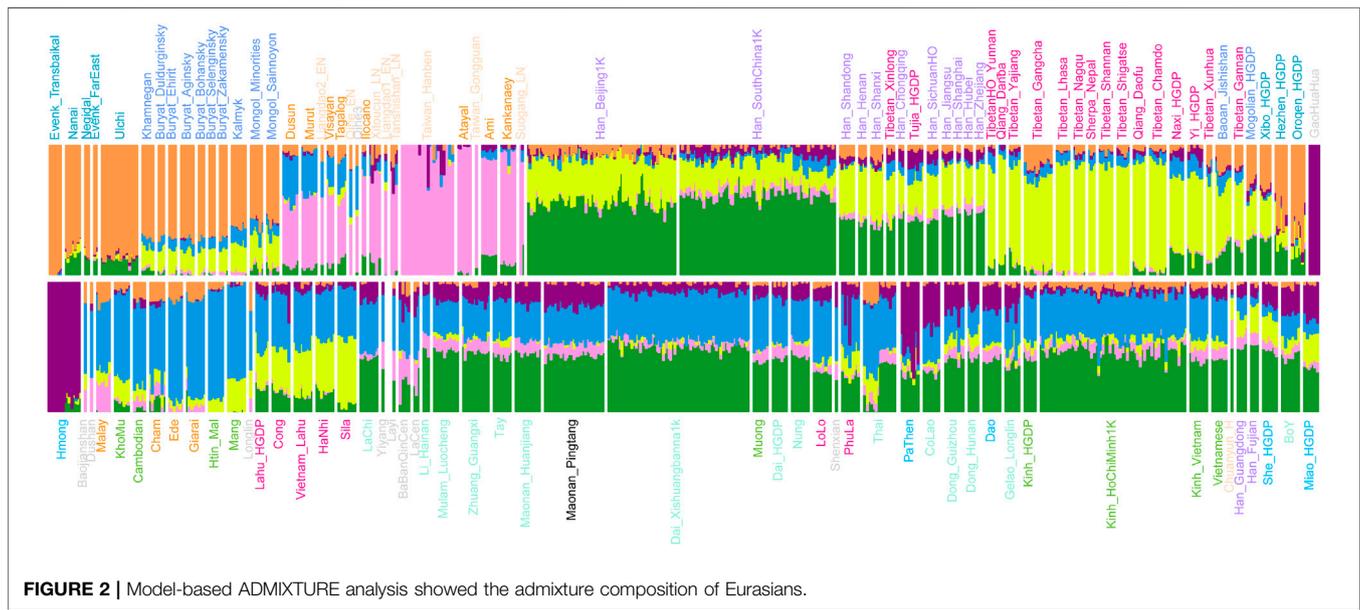
The estimated pairwise F_{st} genetic distances showed the close genetic relationship between Maonan and geographically close Tai–Kadai-speaking populations (Sui and Maonan people in Guangxi), which is consistent with the identified genetic affinity based on the outgroup- f_3 values. Based on the Illumina Array dataset, we further explored the phylogenetic relationship with admixture among 16 Chinese populations belonging to Sinitic, Tungusic, Mongolic, and Tai–Kadai people (Supplementary Figure S3). We found that Pingtang Maonan clustered with the Sandu Sui people and following closed with



Manchu and Mongolian people in Guizhou Province. We did not identify gene flow events into Maonan or from Maonan influx to other reference populations. In addition, to explore the genetic relationship with more reference populations in the maximum-likelihood-based TreeMix tree, we also reconstructed a tree among 39 East Asians (Figure 3). We found that the Northeast Asians clustered closely with each other, including Tungusic-, Mongolic-, and Sino-Tibetan-speaking populations. Southeast Asians also clustered with each other, including Tai-Kadai and Austronesian people. The genetic affinity between Maonan people and other Tai-Kadai-speaking populations was once again confirmed here, including Dai, Li, Zhuang, and Mulam, and in other Austronesian-speaking populations (Ami and Atayal). This identified genetic phylogenetic relationship further confirmed the close genetic relationship between Tai-Kadai and Austronesian people, which was consistent with recent linguistic similarities.

We further explored the fine-scale population structure between Maonan people and other East Asians based on the

patterns of the sharing haplotypes. PCA based on the coancestry matrix showed that Han Chinese clustered together and separated from Pingtang Maonan people and Guizhou Manchu and Mongolic people, which showed that the Maonan people had a different genetic history compared with Han Chinese populations (Figure 4A-C). Model-based ADMIXTURE results with three predefined ancestral sources further confirmed the genetic differentiation among Han Chinese, Manchu, and Maonan, who shared their unique ancestry component in Figure 4D. We also explored the genetic relationship among these populations based on the individual-level and population-level pairwise coincidence matrix estimated from the coancestry data (Figure 4E-F). We found that Maonan people clustered as a separated clade and then clustered with Manchu people. We finally explored the genetic heterozygosity and homogeneity among Tai-Kadai, Hmong-Mien, and Sinitic people using pairwise qpWave analysis. We found the statistically non-significant values ($p_{\text{rank}0} > 0.05$) between Maonan and other Tai-Kadai populations, as well as the geographically



close Hmong–Mien people. These observed patterns suggested Maonan had a relatively close relationship with other Guizhou populations when a distant outgroup was used here (**Supplementary Table S1**). Indeed, we found that Maonan, Hmong–Mien, and Tai–Kadai people shared similar patterns of the distribution of the p_rank0 values, and they clustered together and formed one clade in **Figure 5**.

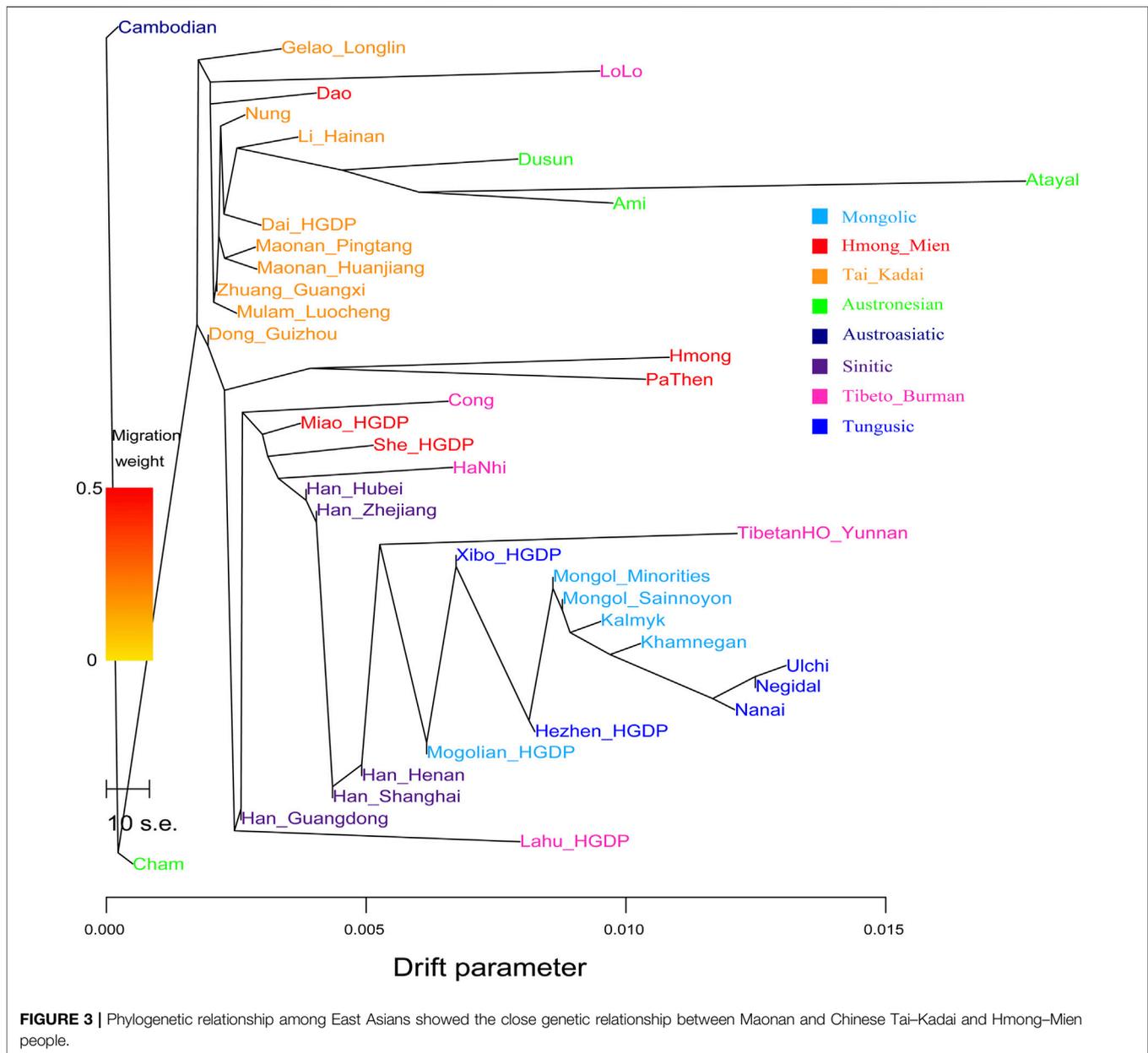
Shared Genetic Drift Inferred From F -Statistics

To explore the relationship between the investigated population and reference populations, we measured allele sharing and admixture signals *via* outgroup- f_3 and admixture- f_3 -statistics. We performed outgroup- f_3 statistics in the form of $f_3(X, Maonan_Pingtang; Mbuti)$, and found that Guizhou Maonan shared more genetic drift with southern Chinese populations, especially for Hmong–Mien- (e.g., Yao, Gejia, and Dongjia) and Tai–Kadai-speaking populations (e.g., Mulam, Zhuang, and Dong) (**Supplementary Table S2A**). Then we used admixture- f_3 statistics of the form $f_3(Source\ 1, Source\ two; Maonan_Pingtang)$ to model possible sources for Maonan_Pingtang people (**Supplementary Table S2B**). However, we did not observe admixture signals (Z -scores less than -3) significantly in the Maonan_Pingtang when we used different East Asian and Southeast Asian ancestral source candidates.

To further explore the differentiation between the Maonan_Pingtang and other East Asian populations, we performed the $f_4(Reference\ population\ 1, Studied\ population; Reference\ population\ 2, Mbuti)$. The identified statistically non-significant f_4 -values (absolute Z -scores less than 3) in

$f_4(Maonan_Huanjiang, Maonan_Pingtang; East\ Asians, Mbuti)$ indicated that Maonan_Pingtang and Maonan_Huanjiang have a close genetic relationship (**Supplementary Table S3A**) compared with other reference populations. Focused on other Tai–Kadai-speaking populations, we observed significantly negative f_4 -values in $f_4(Dai/Zhuang_Guangxi, Maonan_Pingtang; East\ Asians, Mbuti)$, which suggested that Guizhou indigenous populations and Hmong–Mien-speaking populations shared more alleles with Maonan_Pingtang than other Tai–Kadai-speaking populations (**Supplementary Table S3B**). Among Austronesian-speaking populations, the observed significant negative f_4 statistics in the form $f_4(Atayal/Ami, Maonan_Pingtang; East\ Asians\ of\ Hmong-Mien- and\ Tai-Kadai-speaking\ population, Mbuti)$ in **Supplementary Table S3C** showed that Pingtang Maonan shared more alleles with Hmong–Mien- and Tai–Kadai-speaking populations compared with modern Austronesian people. Further comparative ancient DNA evidence demonstrated that Maonan_Pingtang also harbored more ancestry related to Hmong–Mien- and Sinitic-speaking populations when we used Neolithic to Iron Age from Fujian and Taiwan as reference population 1 in the form $f_4(ancient\ southeastern\ East\ Asians, Maonan_Pingtang; East\ Asians, Mbuti)$ (**Supplementary Table S3D**).

To directly compare the genetic relationship between Maonan- and Hmong–Mien-speaking populations, we used Hmong–Mien-speaking populations in Vietnam and Guizhou as reference population one in the form $f_4(Hmong-Mien-speaking\ populations, Maonan_Pingtang; East\ Asians, Mbuti)$. We observed significantly negative f_4 -values here, indicating that Maonan_Pingtang shared more alleles with Tai–Kadai- and Austronesian-speaking populations (**Supplementary Table S3E,F**) than with Hmong–Mien-speaking populations. To



explore the genetic relationship between the studied population and the ancient populations in Guangxi, we used GaoHuahua, Longlin, Baojianshan, and BaBanQinCen as reference population 1. We observed that Guizhou Maonan shared more alleles with Hmong-Mien- and Sino-Tibetan-speaking populations than GaoHuahua, Longlin, and Baojianshan, as we observed significantly negative f_4 -values in the form of $f_4(\text{GaoHuahua}/\text{Longlin}/\text{Baojianshan}, \text{Maonan_Pingtang}; \text{East Asians}, \text{Mbuti})$, which suggests studied population was influenced by gene flow from the north (**Supplementary Tables S3G-I**). We have not observed significant negative f_4 -values in $f_4(\text{BaBanQinCen}, \text{Maonan_Pingtang}; \text{East Asians}, \text{Mbuti})$, indicating that

Maonan_Pingtang shared more alleles with northern populations than BaBanQinCen in Guangxi. We observed BaBanQinCen shared more alleles with other ancient populations than Maonan_Pingtang via significant positive f_4 -values in the form of $f_4(\text{BaBanQinCen}, \text{Maonan_Pingtang}; \text{ancient East Asians}, \text{Mbuti})$ (**Supplementary Table S3J**). We observed significant negative f_4 -values in $f_4(\text{Austroasiatic-speaking populations}, \text{Maonan_Pingtang}; \text{East Asians}, \text{Mbuti})$, which showed that Sinitic-, Tai-Kadai-, and Hmong-Mien-speaking populations from southern China shared more alleles with Maonan_Pingtang than Austroasiatic-speaking populations (**Supplementary Table S3K**). Significant negative values were

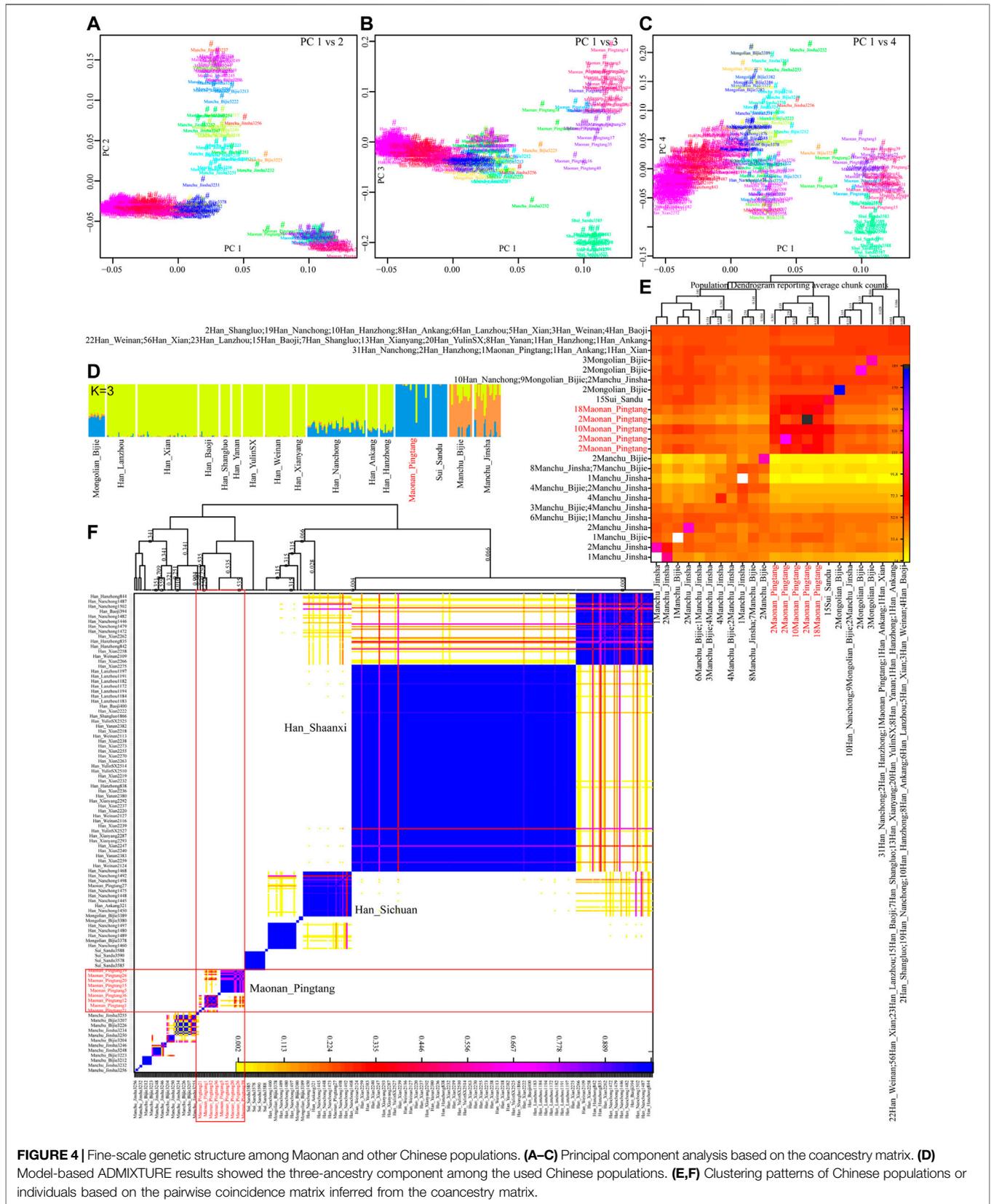


FIGURE 4 | Fine-scale genetic structure among Maonan and other Chinese populations. **(A–C)** Principal component analysis based on the coancestry matrix. **(D)** Model-based ADMIXTURE results showed the three-ancestry component among the used Chinese populations. **(E,F)** Clustering patterns of Chinese populations or individuals based on the pairwise coincidence matrix inferred from the coancestry matrix.

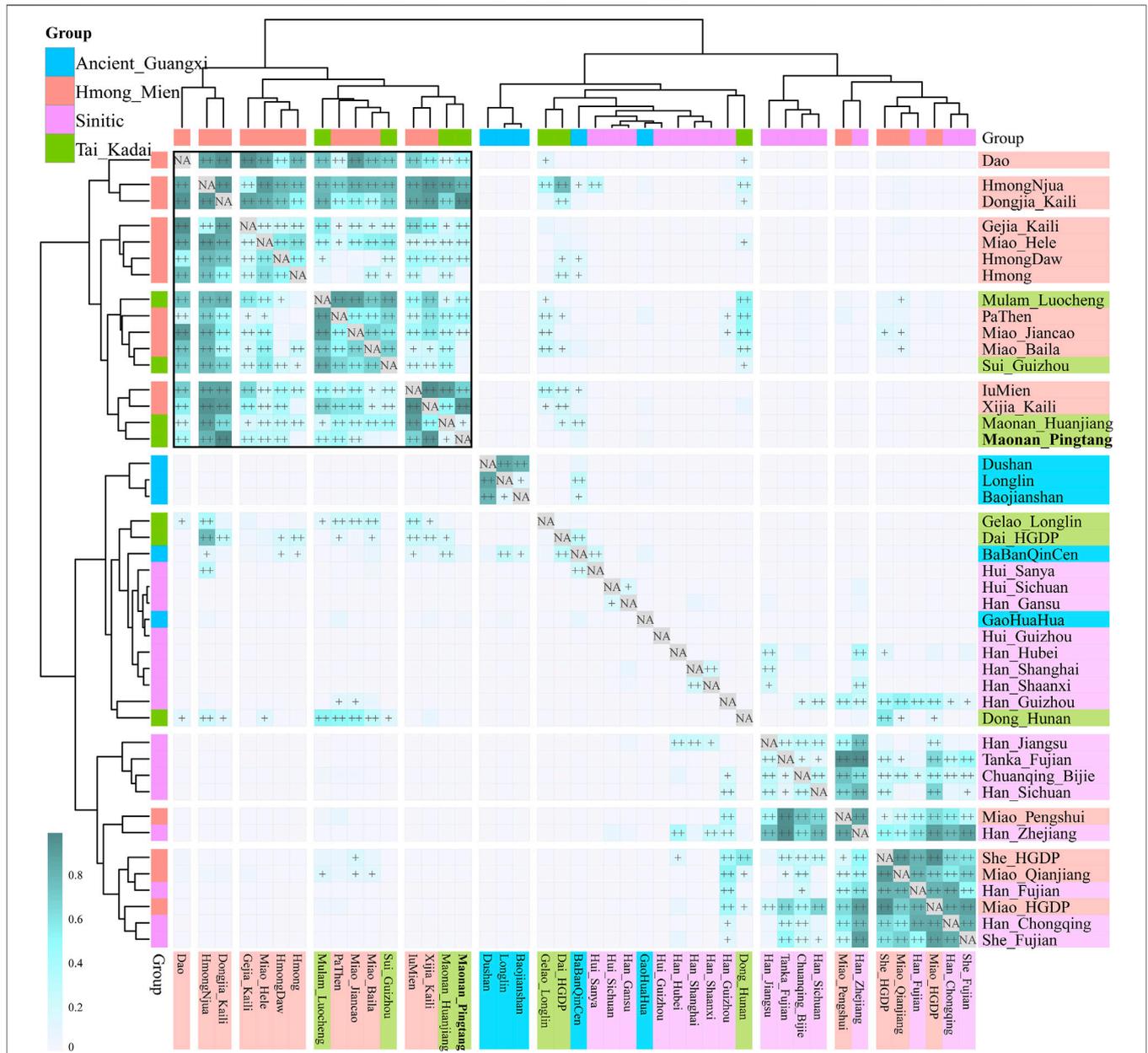


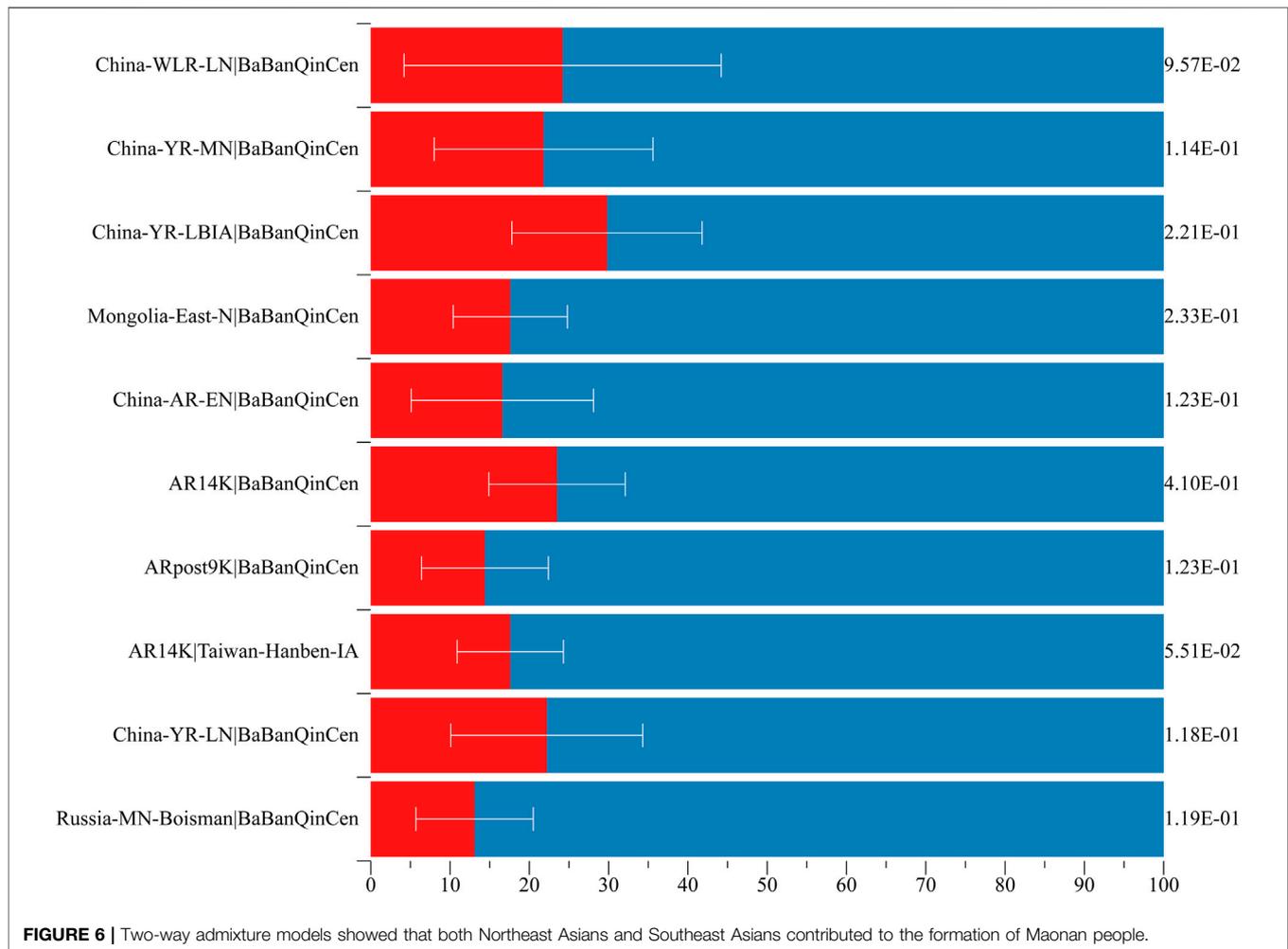
FIGURE 5 | Pairwise qpWave results showed the genetic homogeneity between Maonan and Tai–Kadai and Hmong–Mien people.

observed in the $f_4(\text{southern Tibeto-Burman-speaking population, Maonan_Pingtang; East Asians, Mbuti})$, it was shown that Tai–Kadai-, Austronesian-, and Hmong–Mien-speaking populations shared more alleles with studied population than southern Tibeto–Burman-speaking populations (Supplementary Table S3L). We found that Maonan_Pingtang harbored more Southeast Asian–related ancestry than ancient Yellow River farmers *via* significant negative f_4 -values in $f_4(\text{ancient Yellow$

River farmers, Maonan_Pingtang; Southeast Asians, Mbuti) (Supplementary Table S3M).

Admixture Modeling and Estimates of Admixture Times

To explore the genomic formation of Guizhou Maonan people, we applied *qpWave/qpAdm* methods to model the minimum



number of ancestry sources and evaluated the corresponding ancestry coefficients. We used ancient northern populations (Russia_MN_Boisman, China_YR_LN, AR14K, Chokhopani, ARpost9K, AR14K, China_AR_EN, Mongolia_East_N, China_YR_LBIA, China_YR_MN, and China_WLR_LN) as the northern ancestral sources and ancient southern populations (Taiwan_Hanben_IA, Guangxi_BaBanQinCen) as the southern ancestral sources to estimate the admixture proportions. Here, BaBanQinCen was the representative ancestral source in Guangxi Province, where many Tai-Kadai people lived. When we used ancient Yellow River farmers as the northern source, Southeast Asian ancestry related to the BaBanQinCen in Maonan people spanned from 70.2 to 86.9% (Figure 6, Supplementary Table S4). Date estimates with southern and northern East Asians further revealed that these identified north-to-south admixture events occurred in different historical times (Supplementary Table S5), which is consistent with genetically attested and historically documented southward

population movements (Yang et al., 2020; Wang et al., 2021c; Wang C. C. et al., 2021).

Uniparental Admixture History of Maonan People

Southern China was the originated center of many dominant uniparental lineages of southern Chinese indigenous populations and Southeast Asians (Ke et al., 2001; Wen et al., 2004; He et al., 2020; Sun et al., 2021). Sun et al. provided a higher-resolution phylogeny of O1a-M119 and found this founding lineage widely existed in modern Austronesian-, Tai-Kadai-speaking populations, and southern Han Chinese, suggesting that O1a-M119 lineage was the common lineage of these populations, and the diversified sub-lineages of O1a-M119 were their unique downstream paternal lineages (Sun et al., 2021). Other population genetic studies of Southeast Asians also identified other paternal founding lineages, including O1b1a1a-M95 and O2-M122 (Wen et al., 2004). Similarly, the complex landscape of

maternal lineages in South China was also identified in the previous control-region or whole mitochondrial sequencing projects, including D4, B4, and M7 (Li et al., 2019; Mengge et al., 2020). Here, we also identified that northern and southern paternal and maternal lineages contributed to the uniparental gene pool of Guizhou Maonan people (**Supplementary Table S6**). We observed eight paternal Y-chromosome lineages (O1a1a2a1, O1b1a1a1a1a1a1b, O2a2b1a1a5a2, D1a1a2a2~, O1b1a1a1b, O1b1a1a1a1a1a1a1a1, O2a1c2, and O1b1a1a1a1a1a1a2-Z24131) with the frequencies ranging from 0.0435 to 0.6957 among 23 males, in which haplogroup of O1b1a1a1a1a1a2-Z24131 was observed in 16 individuals. Our results showed that Southeast Asian-dominant paternal lineage of O1b1a1a1a1a1a2 was the founding lineage of Tai–Kadai-speaking Maonan people. O1b1a1a* (O-M95*) was previously evidenced to contribute much to the paternal gene pool of populations from South China, Thailand, and Laos (Kutanan et al., 2019). O1b1a1a* was also evidenced has experienced significant population expansions in the Neolithic period in Tai–Kadai and Austroasiatic populations. Thus, combined with the identified unique population structure of Maonan people based on the autosomal SNPs and this identified specific lineage of O1b1a1a1a1a1a2 with high frequency, Maonan people could be treated as the most representative inland Tai–Kadai-speaking population of southern ancient indigenous populations with the relatively minor genetic influence of southward Han Chinese expansion. Our results also suggested that the genetically documented marriage pattern was consistent with that of the culturally documented customs of patrilocality, which is also evidenced *via* the marriage pattern among Tai–Kadai people in Southeast Asians (Kutanan et al., 2019). In addition, we assigned all 41 maternally inherited mtDNA lineages into 25 terminal lineages with frequencies ranging from 0.0244 to 0.0732 (B4a1e, B5a1c1, F1a, F1c, F3b, and B4), which were observed frequently in both northern and southern East Asians.

DISCUSSION

Guizhou Province is rich in ethnolinguistic and cultural diversity (Wang et al., 2021a). Previous genetic studies have investigated the general landscape of genetic variations of Guizhou populations based on the autosomal, X/Y-chromosomal short tandem repeats (STRs), and ancestry-informative SNPs (Chen et al., 2018a; Chen et al., 2018b; Chen et al., 2018c; Chen et al., 2018d; Chen et al., 2019a; Chen et al., 2019b; He et al., 2021b). Archaeologically attested Daxi, Qujialing, and Shijiahe people were occupied in what is now Guizhou Province, and the present Guizhou region was occupied by Sino-Tibetan-speaking (Han, Yi, and others), Tai–Kadai-speaking (Gelao, Bouyei, Dong et al.), and Hmong–Mien-speaking populations (Miao, She, and Yao). Recent ancient DNA studies from the Yellow River Basin in Northeast Asia, Fujian, and Guangxi provinces from Southeast Asia also found that the bi-directional north-to-south population movements have shaped the genetic landscape of East Asians (Ning et al., 2020; Yang et al., 2020; Wang et al., 2021c; Mao et al., 2021; Wang C. C. et al., 2021). Demographic modeling of central Chinese populations, including Han and Tujia, also showed the

genetic influences from both the northern millet farmers and southern rice farmers (He et al., 2021a). The Tai–Kadai language family and their corresponding people widely existed in Guizhou and surrounding regions; however, the fine-scale genetic structure of this linguistically specific population still needs to be further explored.

We genotyped the genome-wide SNP data from 41 Tai–Kadai-speaking Maonan people and explored their genetic origin, admixture history, and phylogenetic relationship with surrounding populations. Descriptive analysis based on the PCA and ADMIXTURE analyses showed that the Guizhou Maonan people and Guangxi Maonan people had the closest genetic relationship and shared the most genetic affinity, suggesting their common origin and admixture history. The genetic affinity within the population among Tai–Kadai-speaking populations was also evidenced by the Dong and Bouyei based on the Affymetrix-based array data (Wang et al., 2021a). In addition, population genetic analysis based on the forensic genetic markers (STR and Indels) also revealed the genetic affinity between geographically different Tai–Kadai-speaking populations rather than populations from other language families (Chen et al., 2018c; He et al., 2019a; He et al., 2019b). Interestingly, we also identified the genetic affinity between Maonan and Hmong–Mien-speaking populations in Guizhou Province among the non-Tai–Kadai-speaking populations based on the observed non-statistically significant f_4 -statistics in the form $f_4(\text{Maonan}, \text{Hmong–Mien-speaking Gejia/Dongjia/Xijia}; \text{other reference Asians populations}, \text{Mbuti})$ and statistically negative f_4 -values in $f_4(\text{Asian reference populations}, \text{Maonan}; \text{Hmong–Mien}, \text{Mbuti})$. Our results were consistent with previous reports based on the forensic markers. He et al. explored the genetic diversity and forensic features of Guizhou Tai–Kadai-speaking Gelao people and identified the population interplay between Gelao and neighboring Hmong–Mien-speaking populations (He et al., 2019b). Our identified extensive genetic admixture between Hmong–Mien and Tai–Kadai people suggested that there was no clear genetic barrier between geographically close but linguistically different ethnic groups, which suggested that they have experienced extensive population interaction although initially they were of independent origin.

Linguistic interaction between Tai–Kadai and Sino-Tibetan languages was widely documented (Edmondson, 1988; Edmondson, 1997; Lu, 2008). Population interaction between Maonan- and Sino-Tibetan-speaking populations was also identified in our genetic study, which is consistent with other recently published population genetic investigations. He et al. recently explored the fine-scale genetic structure of four Guizhou Han populations and found their extensive admixture with Guizhou indigenes (Wang et al., 2021b). In addition, Wang et al. studied the genetic admixture of Guizhou culturally unique Hui people and found their connection between indigenous Han people (Wang et al., 2020). Other genome-wide SNP-based genetic analyses focused on Guizhou officially unrecognized Chuanqing people also found their genetic affinity with geographically close Han populations (Lu et al., 2020). Additionally, we found a close genetic relationship between Maonan and Guizhou Hans in the f_3/f_4 -statistics and the TreeMix-based maximum-likelihood-based phylogenetic tree, which suggested their recent admixture process. The shared ancestry between Guizhou Maonan and

Han people in ADMIXTURE and their qualitative indices was consistent with the shared cultural background between present-day Guizhou Han and Maonan people.

More and more ancient genomes in the surrounding regions of Guizhou Province were reported recently, especially important ancestry sources of the possible ancestor of Tai–Kadai-speaking populations (BaBanQinCen) and a possible ancestor of Hmong–Mien-speaking populations (GaoHuaHua). BaBanQinCen was one meta-population from four archeological sites that lived in Guangxi Province 2000 years ago, and GaoHuaHua was also a meta-population from three archeological sites that lived in Guangxi around 1,500 years ago (Wang et al., 2021c). Here, we found that Maonan people shared the most genetic affinity with ancient Guangxi historic BaBanQinCen, which was recently genetically attested as the direct ancestor of Guizhou Tai–Kadai people (Wang et al., 2021c). Outgroup- f_3 -statistics and shared ancestry inferred from f_4 -statistics further confirmed the closest genetic connection between Maonan and BaBanQinCen compared with other Guizhou historic people (GaoHuaHua, Layi et al.) and prehistoric Longlin, Dushan, and Baojianshan people. Among all reported ancient Northeast Asians, including the inland and coastal Neolithic Northeast Asians from Shandong, Henan, Shaanxi, Gansu, Inner Mongolia in the Yellow River Basin, and Neolithic Siberian, we found a close genetic connection between Maonan and Bronze Age to Iron Age people from Henan Province, which suggested Maonan people might have obtained gene influxes from them. Indeed, we obtained statistically negative f_4 -values in the f_4 (Guangxi historic people, Maonan; Yellow River millet farmers, Mbuti), which directly evidenced that compared with the genetically attested ancestors of Tai–Kadai-speaking populations from the Guangxi region, Maonan people shared additional genetic materials from Northeast Asians. This identified admixture process was further confirmed *via* the qpAdm-based two-way admixture models with one source from Guangxi historic people and the other sources from Northeast Asians. We also found that two-way admixture models of Hanben–Northeast Asians can also well-fit the genetic composition of studied Maonan, which suggests the genetic influence from the southeastern coastal Fujian ancient people in the gene pool of inland Tai–Kadai-speaking Maonan people. As we know, Tai–Kadai-speaking populations were widely distributed in South China, including Guizhou, Guangxi, and Hainan. Previous genetic analysis from Hainan Province also reported a relatively isolated genetic structure of the Hainan Tai–Kadai-speaking Hlai people (He et al., 2020). Thus, dense sampling of Tai–Kadai-speaking populations and obtaining their whole-genome sequencing data would help to characterize the complete genetic admixture landscape of Chinese Tai–Kadai-speaking populations.

CONCLUSION

We reported the first batch of genome-wide SNP data of Tai–Kadai-speaking Maonan people from Guizhou Province and comprehensively explored genetic structure, origin, and

admixture processes based on the descriptive analyses (PCA and ADMIXTURE) and qualitative measures (f -statistics, qpAdm). Results from PCA and ADMIXTURE showed a close genetic relationship between Maonan and other geographically different Tai–Kadai-speaking populations, especially for the closest relationship between Guizhou Maonan and Guangxi Maonan. No-admixture signatures were identified *via* admixture- f_3 statistics showed the unique genetic structure of Maonan people compared with geographically close Han people. Further analysis based on the outgroup- f_3 statistics and f_4 -based analysis showed a close relationship between Maonan and Guizhou Sino-Tibetan and Hmong-speaking populations, as well as a close connection between Guangxi historic people and Guizhou Tai–Kadai-speaking populations, suggesting their admixture history with the sources from surrounding regions. The well-fitted two-way admixture models with ancient northern and southern East Asians demonstrated that Tai–Kadai-speaking populations derived primary ancestry related to 1500-year-old Guangxi BaBanQinCen people and additional genes from Northeast Asia.

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 below: <https://zenodo.org/record/5701604>, 10.5281/zenodo.5701604.

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

C-CW and JH designed this study. JC, GH, and C-CW wrote and revised the manuscript. ZR, QW, YL, HZ, MY, HZ, JJ, and JH collected the samples. QW, ZR, HZ, JJ, YL, MY, JC, and JH conducted the experiment. GH, JZ, JG, JC, KZ, XY, RW, HM, LT, YL, QS, WY, and C-CW analyzed the data. All authors reviewed the manuscript.

FUNDING

This work was funded by the Guizhou Scientific Support Project, Qian Science Support (2021) General 448; Guizhou Province Education Department, Characteristic Region Project, Qian Education KY No. (2021) 065; Guizhou “Hundred” High-level Innovative Talent Project, Qian Science Platform Talents (2020) 6,012; Guizhou Scientific Support Project, Qian Science Support

(2020) 4Y057; Guizhou Science Project, Qian Science Foundation (2020) 1Y353; Guizhou Scientific Support Project, Qian Science Support (2019) 2,825; the Guizhou Scientific Cultivation Project, Qian Science Platform Talent (2018) 5779-X; and the Guizhou Engineering Technology Research Center Project, Qian High-Tech of Development and Reform Commission NO. (2016) 1,345; the National Natural Science Foundation of China (NSFC 31801040), the “Double First Class University Plan” key construction project of Xiamen University (the origin and evolution of East Asian populations and the spread of Chinese civilization, 0310/X2106027); Nanqiang Outstanding Young Talents Program of Xiamen University (X2123302); the Major Project of National Social Science Foundation of China

granted to C-CW (21 & ZD285) and Xiaohua Deng (20 & ZD248); the European Research Council (ERC) grant to Dan Xu (ERC-2019-ADG-883700-TRAM). We thank S. Fang and Z. Xu from Information and Network Center of Xiamen University for their help with high-performance computing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.815285/full#supplementary-material>

REFERENCES

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast Model-Based Estimation of Ancestry in Unrelated Individuals. *Genome Res.* 19, 1655–1664. doi:10.1101/gr.094052.109
- Bin, X., Wang, R., Huang, Y., Wei, R., Zhu, K., Yang, X., et al. (2021). Genomic Insight into the Population Structure and Admixture History of Tai-Kadai-Speaking Sui People in Southwest China. *Front. Genet.* 12, 735084. doi:10.3389/fgene.2021.735084
- Browning, B. L., and Browning, S. R. (2011). A Fast, Powerful Method for Detecting Identity by Descent. *Am. J. Hum. Genet.* 88, 173–182. doi:10.1016/j.ajhg.2011.01.010
- 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. *GigaSci* 4, 7. doi:10.1186/s13742-015-0047-8
- Chen, J., He, G., Ren, Z., Wang, Q., Liu, Y., Zhang, H., et al. (2021). Genomic Insights into the Admixture History of Mongolic- and Tungusic-Speaking Populations from Southwestern East Asia. *Front. Genet.* 12, 685285. doi:10.3389/fgene.2021.685285
- Chen, P., Han, Y., He, G., Luo, H., Gao, T., Song, F., et al. (2018a). Genetic Diversity and Phylogenetic Study of the Chinese Gelao Ethnic Minority via 23 Y-STR Loci. *Int. J. Leg. Med.* 132, 1093–1096. doi:10.1007/s00414-017-1743-y
- Chen, P., He, G., Xing, H., Gao, H., Wang, M., Zhao, M., et al. (2019a). Forensic Characteristics and Phylogenetic Analysis of 23 Y-STR Loci in the Miao Population from Guizhou Province, Southwest China. *Ann. Hum. Biol.* 46, 84–87. doi:10.1080/03014460.2019.1583374
- Chen, P., He, G., Zou, X., Wang, M., Jia, F., Bai, H., et al. (2018b). Forensic Characterization and Genetic Polymorphisms of 19 X-Chromosomal STRs in 1344 Han Chinese Individuals and Comprehensive Population Relationship Analyses Among 20 Chinese Groups. *PLoS One* 13, e0204286. doi:10.1371/journal.pone.0204286
- Chen, P., He, G., Zou, X., Wang, M., Luo, H., Yu, L., et al. (2018c). Genetic Structure and Polymorphisms of Gelao Ethnicity Residing in Southwest China Revealed by X-Chromosomal Genetic Markers. *Sci. Rep.* 8, 14585. doi:10.1038/s41598-018-32945-7
- Chen, P., He, G., Zou, X., Zhang, X., Li, J., Wang, Z., et al. (2018d). Genetic Diversities and Phylogenetic Analyses of Three Chinese Main Ethnic Groups in Southwest China: A Y-Chromosomal STR Study. *Sci. Rep.* 8, 15339. doi:10.1038/s41598-018-33751-x
- Chen, P., Luo, L., Gao, H., Wu, J., Wang, Y., He, G., et al. (2019b). Forensic Performance of 30 InDels Included in the Investigator DIPlex System in Miao Population and Comprehensive Genetic Relationship in China. *Int. J. Leg. Med.* 133, 1389–1392. doi:10.1007/s00414-019-02057-6
- Diamond, J., and Bellwood, P. (2003). Farmers and Their Languages: the First Expansions. *Science* 300, 597–603. doi:10.1126/science.1078208
- Edmondson, J. A. (1988). *Comparative Kadai: Linguistic Studies beyond Tai*. Dallas: Summer Institute of Linguistics, University of Texas at Arlington.
- Edmondson, J. A. (1997). *Comparative Kadai: The Tai branch*. Dallas: Summer Institute of Linguistics, University of Texas at Arlington.
- He, G., Li, Y. X., Wang, M. G., Zou, X., Yeh, H. Y., Yang, X. M., 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., Liu, J., Wang, M., Zou, X., Ming, T., Zhu, S., et al. (2021b). Massively Parallel Sequencing of 165 Ancestry-Informative SNPs and Forensic Biogeographical Ancestry Inference in Three Southern Chinese Sinitic/Tai-Kadai Populations. *Forensic Sci. Int. Genet.* 52, 102475. doi:10.1016/j.fsigen.2021.102475
- He, G., Ren, Z., Guo, J., Zhang, F., Zou, X., Zhang, H., et al. (2019a). Population Genetics, Diversity and Forensic Characteristics of Tai-Kadai-Speaking Bouyei Revealed by Insertion/deletions Markers. *Mol. Genet. Genomics* 294, 1343–1357. doi:10.1007/s00438-019-01584-6
- 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, Z., Guo, J., Wang, M., Zou, X., Tang, R., et al. (2020). Inferring the Population History of Tai-Kadai-Speaking People and Southernmost Han Chinese on Hainan Island by Genome-wide Array Genotyping. *Eur. J. Hum. Genet.* 28, 1111–1123. doi:10.1038/s41431-020-0599-7
- He, G., Wang, Z., Zou, X., Wang, M., Liu, J., Wang, S., et al. (2019b). Tai-Kadai-speaking Gelao Population: Forensic Features, Genetic Diversity and Population Structure. *Forensic Sci. Int. Genet.* 40, e231–e239. doi:10.1016/j.fsigen.2019.03.013
- 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
- Huang, J. (2016). An Analysis of the Origin of Rice-growing Culture in China. *Local Cult. Res.* 4, 40–57. (In Chinese).
- Huang, X., Xia, Z.-Y., Bin, X., He, G., Guo, J., Lin, C., et al. (2020). Genomic Insights into the Demographic History of Southern Chinese. *bioRxiv*. doi:10.1101/2020.11.08.373225
- Ke, Y., Su, B., Song, X., Lu, D., Chen, L., Li, H., et al. (2001). African Origin of Modern Humans in East Asia: a Tale of 12,000 Y Chromosomes. *Science* 292, 1151–1153. doi:10.1126/science.1060011
- Kutanan, W., Kampuansai, J., Srikumool, M., Brunelli, A., Ghirotto, S., Arias, L., et al. (2019). Contrasting Paternal and Maternal Genetic Histories of Thai and Lao Populations. *Mol. Biol. Evol.* 36, 1490–1506. doi:10.1093/molbev/msz083
- Kutanan, W., Liu, D., Kampuansai, J., Srikumool, 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
- Larena, M., Sanchez-Quinto, F., Sjödin, P., Mckenna, J., Ebeo, C., Reyes, R., et al. (2021). Multiple Migrations to the Philippines during the Last 50,000 Years. *Proc. Natl. Acad. Sci. USA* 118, e2026132118. doi:10.1073/pnas.2026132118
- Li, Y.-C., Ye, W.-J., Jiang, C.-G., Zeng, Z., Tian, J.-Y., Yang, L.-Q., et al. (2019). River Valleys Shaped the Maternal Genetic Landscape of Han Chinese. *Mol. Biol. Evol.* 36, 1643–1652. doi:10.1093/molbev/msz072

- Liang, M., and Zhang, J. (1996). *An Introduction to the Tai-Kadai Language Family (Chinese)* Beijing, China: Social Sciences Publishing House.
- Lipson, M., Cheronet, O., Mallick, S., Rohland, N., Oxenham, M., Pietruszewski, M., et al. (2018). Ancient Genomes Document Multiple Waves of Migration in Southeast Asian Prehistory. *Science* 361, 92–95. doi:10.1126/science.aat3188
- 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., Xie, J., Wang, M., Liu, C., Zhu, J., Zou, X., et al. (2021a). Genomic Insights into the Population History and Biological Adaptation of Southwestern Chinese Hmong-Mien People. *Front. Genet.* 12, 815160. doi:10.3389/fgene.2021.815160
- 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
- Loh, P.-R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., et al. (2013). Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics* 193, 1233–1254. doi:10.1534/genetics.112.147330
- Lu, J., Zhang, H., Ren, Z., Wang, Q., Liu, Y., Li, Y., et al. (2020). Genome-wide Analysis of Unrecognised Ethnic Group Chuanqing People Revealing a Close Affinity with Southern Han Chinese. *Ann. Hum. Biol.* 47, 465–471. doi:10.1080/03014460.2020.1782470
- Lu, T. Q. (2008). *A Grammar of Maonan*. Boca Raton, United States: Universal Publishers.
- Ma, T., Rolett, B. V., Zheng, Z., and Zong, Y. (2020). Holocene Coastal Evolution Preceded the Expansion of Paddy Field rice Farming. *Proc. Natl. Acad. Sci. USA* 117, 24138–24143. doi:10.1073/pnas.1919217117
- 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. doi:10.1016/j.cell.2021.04.040
- Mccoll, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J. V., et al. (2018). The Prehistoric Peopling of Southeast Asia. *Science* 361, 88–92. doi:10.1126/science.aat3628
- Mengege, W., Guanglin, H., Yongdong, S., Shouyu, W., Xing, Z., Jing, L., et al. (2020). Massively Parallel Sequencing of Mitogenome Sequences Reveals the Forensic Features and Maternal Diversity of Tai-Kadai-Speaking Hlai Islanders. *Forensic Sci. Int. Genet.* 47, 102303. doi:10.1016/j.fsigen.2020.102303
- Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., et al. (2020). Ancient Genomes from Northern China Suggest Links between Subsistence Changes and Human Migration. *Nat. Commun.* 11, 2700. doi:10.1038/s41467-020-16557-2
- 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
- Pickrell, J. K., and Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-wide Allele Frequency Data. *Plos Genet.* 8, e1002967. doi:10.1371/journal.pgen.1002967
- Stevens, C. J., and Fuller, D. Q. (2017). The Spread of Agriculture in Eastern Asia. *Lang. Dyn. Change* 7, 152–186. doi:10.1163/22105832-00702001
- Sun, J., Li, Y. X., Ma, P. C., Yan, S., Cheng, H. Z., Fan, Z. Q., et al. (2021). Shared Paternal Ancestry of Han, Tai-Kadai-speaking, and Austronesian-speaking Populations as Revealed by the High Resolution Phylogeny of O1a-M119 and Distribution of its Sub-lineages within China. *Am. J. Phys. Anthropol.* 174, 686–700. doi:10.1002/ajpa.24240
- Wang, C. C., Yeh, H.-Y., Popov, A. N., Zhang, H.-Q., Matsumura, H., Sirak, K., et al. (2021). Genomic Insights into the Formation of Human Populations in East Asia. *Nature* 591, 413–419. doi:10.1038/s41586-021-03336-2
- Wang, M., He, G., Zou, X., Chen, P., Wang, Z., Tang, R., et al. (2021a). Reconstructing The Genetic Admixture History of Tai-Kadai and Sinitic People: Insights From Genome-Wide SNP Data From South China. *J. Syst. Evol.* doi:10.1111/jse.12825
- Wang, M., Yuan, D., Zou, X., Wang, Z., Yeh, H.-Y., Liu, J., et al. (2021b). 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, Q., Zhao, J., Ren, Z., Sun, J., He, G., Guo, J., et al. (2020). Male-Dominated Migration and Massive Assimilation of Indigenous East Asians in the Formation of Muslim Hui People in Southwest China. *Front. Genet.* 11, 618614. doi:10.3389/fgene.2020.618614
- Wang, T., Wang, W., Xie, G., Li, Z., Fan, X., Yang, Q., et al. (2021c). Human Population History at the Crossroads of East and Southeast Asia since 11,000 Years Ago. *Cell* 184, 3829–3841. doi:10.1016/j.cell.2021.05.018
- Wang, W. G. (2004). A Comprehensive Study of the History of Baiyue Ethnic Group. *J. yunnan Univ.* (In Chinese).
- Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y., et al. (2004). Genetic Evidence Supports Demic Diffusion of Han Culture. *Nature* 431, 302–305. doi:10.1038/nature02878
- 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
- Zhang, H., He, G., Guo, J., Ren, Z., Zhang, H., Wang, Q., et al. (2019). Genetic Diversity, Structure and Forensic Characteristics of Hmong-Mien-speaking Miao Revealed by Autosomal Insertion/deletion Markers. *Mol. Genet. Genomics* 294, 1487–1498. doi:10.1007/s00438-019-01591-7
- Zhang, J. (2016). An Exploration of the Bronze Drums Culture and the Origin of the Belief System of Southern China. *Arts Explorat* 30 (4), 68–80. (In Chinese). doi:10.13574/j.cnki.artsexp.2016.04.008
- Zhang, X., Li, C., Zhou, Y., Huang, J., Yu, T., Liu, X., et al. (2020). A Matrilineal Genetic Perspective of Hanging Coffin Custom in Southern China and Northern Thailand. *iScience* 23, 101032. doi:10.1016/j.isci.2020.101032
- Zhao, Z. (2011). New Archaeobotanic Data for the Study of the Origins of Agriculture in China. *Curr. Anthropol.* 52, S295–S306. doi:10.1086/659308

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.

Copyright © 2022 Chen, He, Ren, Wang, Liu, Zhang, Yang, Zhang, Ji, Zhao, Guo, Chen, Zhu, Yang, Wang, Ma, Tao, Liu, Shen, Yang, Wang and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.