Population Genetic Analysis of Modern and Ancient DNA Variations Yields New Insights Into the Formation, Genetic Structure, and Phylogenetic Relationship of Northern Han Chinese

Pengyu Chen[1,2†], Jian Wu[1,2†], Li Luo[1,2†], Hongyan Gao[1,2], Mengge Wang[3], Xing Zou[3], Yingxiang Li[4], Gang Chen[4], Haibo Luo[2], Limei Yu[5], Yanyan Han[6], Fuquan Jia[7]* and Guanglin He[3]*

[1] Center of Forensic Expertise, Affiliated Hospital of Zunyi Medical University, Zunyi, China, [2] Department of Forensic Medicine, Zunyi Medical University, Zunyi, China, [3] Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, China, [4] Department of Bioinformatics, WeGene, Shenzhen, China, [5] Key Laboratory of Cell Engineering in Guizhou Province, Affiliated Hospital of Zunyi Medical University, Zunyi, China, [6] Department of Nutrition and Food Hygiene, School of Public Health, Zunyi Medical University, Zunyi, China, [7] Department of Forensic Medicine, Inner Mongolia Medical University, Hohhot, China

Modern East Asians derived from the admixture of aborigines and incoming farmers expanding from Yellow and Yangtze River Basins. Distinct genetic differentiation and subsequent admixture between Northeast Asians and Southeast Asians subsequently evidenced by the mitochondrial DNA, Y-chromosomal variations, and autosomal SNPs. Recently, population geneticists have paid more attention to the genetic polymorphisms and background of southern-Han Chinese and southern native populations. The genetic legacy of northern-Han remains uncharacterized. Thus, we performed this comprehensive population genetic analyses of modern and ancient genetic variations aiming to yield new insight into the formation of modern Han, and the genetic ancestry and phylogenetic relationship of the northern-Han Chinese population. We first genotyped 25 forensic associated markers in 3,089 northern-Han Chinese individuals using the new-generation of the Huaxia Platinum System. And then we performed the first meta-analysis focused on the genetic affinity between Asian Neolithic~Iron Age ancients and modern northern-Han Chinese by combining mitochondrial variations in 417 ancient individuals from 13 different archeological sites and 812 modern individuals, as well as Y-chromosomal variations in 114 ancient individuals from 12 Neolithic~Iron Age sites and 2,810 modern subjects. We finally genotyped 643,897 genome-wide nucleotide polymorphisms (SNPs) in 20 Shanxi Han individuals and combined with 1,927 modern humans and 40 Eurasian ancient genomes to explore the genetic structure and admixture of northern-Han Chinese. We addressed genetic legacy, population structure and phylogenetic relationship of northern-Han Chinese *via* various analyses. Our population genetic results from five different reference datasets indicated that Shanxi Han shares a closer phylogenetic relationship with

northern-neighbors and southern ethnically close groups than with Uyghur and Tibetan. Genome-wide variations revealed that modern northern-Han derived their ancestry from Yakut-related population (25.2%) and She-related population (74.8%). Summarily, the genetic mixing that led to the emergence of a Han Chinese ethnicity occurred at a very early period, probably in Neolithic times, and this mixing involved an ancient Tibeto-Burman population and a local pre-Sinitic population, which may have been linguistically Altaic.

INTRODUCTION

Han Chinese, with a total population size circa 1.4 billion, is the world's largest ethnic group and dominant ethnicity in China and Singapore. The origin of the Han Chinese population, genetic relationship with adjacent groups and past migratory pattern and admixture history have gained considerable attention from scientists working in the anthropology, linguistics, history, population and forensic genetics (Zhao et al., 2011; Gao et al., 2015; Zhao et al., 2015b; Li et al., 2017; Nothnagel et al., 2017; Zhang et al., 2017b; Chiang et al., 2018). Archaeological and anthropological evidence showed that human occupation in East Asia has experienced archaic hominin extinction, genetic introgression between early anatomically modern human and Denisovan or Neanderthals, the transformation from hunting–gathering to agriculture, massive admixture and migratory history with ethnolinguistically diverse populations in the past 50–100 thousand years (Nielsen et al., 2017). Expansions of the maternally-inherited mitochondrial DNA (mtDNA) and paternally-inherited Y-chromosome haplogroup lineages indicated that ethnically different East Asians derived from southeastern groups and experienced south-to-north migrations driven by a variety of evolutionary mechanisms (Su et al., 1999; Yao et al., 2002). Besides, social practices, including subsistence strategies, residence patterns, and agricultural expansion, play an indispensable role in shaping the patterns of Chinese populations (Nielsen et al., 2017). Ancient mitochondrial and Y-chromosomal DNA studies in East Asian Neolithic~Iron Age populations have drastically increased in past decades (Cui et al., 2010; Li et al., 2010; Li et al., 2011; Zhao et al., 2011; Wang et al., 2012; Cui et al., 2013; Zhao et al., 2014; Dong et al., 2015; Gao et al., 2015; Li et al., 2015; Zhao et al., 2015b; Li et al., 2017; Zhang et al., 2017b; Li et al., 2018), however, how the peopling and settlement history of Neolithic populations influence the origin, expansion, and migration of the Han Chinese population is still unclear.

Physical anthropological investigation of somatometric and nonmetric features revealed that a significant difference exists between northern-Han Chinese and southern-Han Chinese (Sanchez-Mazas et al., 2011). Subsequently, Chu et al. (1998) found genetic evidence to support the distinction between southern and northern populations. Phylogeographic or genetic differentiation between northern-Han and southern-Han have been also evidenced by Yao et al. (2002) using mitochondrial DNA, Wen et al. (2004) using combined testing Y-chromosome and mitochondrial DNA variations, and Chen et al. (2009)

and Xu et al. (2009) using high-density genotype data. Our previous study has investigated the genetic polymorphisms, forensic features and genetic relationship of currently widely-used autosomal short tandem repeats (STRs) in the southern-Han Chinese residing in the Pearl River Delta (He et al., 2018d). Thus, reconstructing the forensic reference database, estimating the forensic allele frequency and parameters and dissecting the genetic relationship of this genetically diverse northern-Han Chinese population are very necessary and urgent.

STR, also called as microsatellite, is one of the extraordinary mutated genetic markers, is widely existed human autosomal, X-chromosome and Y-chromosome genomes (Ge et al., 2014). This length polymorphism marker is generated by the slippage synthesis of simple sequence (2–8bp) (Schlotterer and Tautz, 1992). STRs located on the no-recombining region of Y-chromosome are the best candidates for forensic pedigree searches and identifying the perpetrator in the sexual crime or rape cases, and X-chromosomal STRs are best suitable for applications in the deficiency and incestuous cases (He et al., 2017d; Chen et al., 2018). Autosomal STR genotyping is the gold standard in the routine forensic cases. Nowadays, all organizations or countries optimized their accepted STR panels to improve the international collaboration, such as the expanded CODIS core loci, extended European standard set (ESS-extended). Huaxia Platinum System (Thermo Fisher Scientific) was integrated all twenty expanded CODIS core loci, additional STRs included in the Chinese National Database and two gender determination loci (He et al., 2018b). Single nucleotide polymorphisms (SNPs), with the number over 84.7 million in the human genome, are the best candidate to explore the detailed processes of human origin, migration, evolution, adaptation and admixture (Genomes Project et al., 2015).

Although Y-chromosomal and X-chromosomal variations of the northern-Han Chinese population have been investigated and reported (He et al., 2017d; Chen et al., 2018). Autosomal STR allele distribution of this new-generation of the Huaxia Platinum Amplification System with regard to forensic statistical features has not previously been investigated. Besides, the population genetic structure and admixture history of northern-Han *via* high-density genetic markers are unclear. Thus, we genotyped and analyzed 23 autosomal STRs in 3,089 unrelated Han Chinese individuals and 643,897 SNPs in 20 Hans residing in Shanxi Province. Shanxi Province is between 34°34′-40°44′ north latitude and 110°14′-114°33′ east longitude, which stretches about a total area of 156,700 km² from the Yellow River in the west and south to the Taihang Mountain in the east and the Great Wall in the north.

This area is bounded by the Shaanxi, Inner Mongolia, Hebei, Henan, and other Provinces. Archaeological, anthropological and genetic evidence from Hengbei site consistently considered that Han Chinese is originated from Shanxi and neighboring regions, also called the Central Plain (Zhao et al., 2015b). And then Han Chinese population migrated southward with the Han-associated culture (Demic diffusion) and admixed with southern Chinese natives and formed the current patterns of genetic diversity distribution (Wen et al., 2004). In addition to the estimation of forensic characterization of autosomal STRs in northern-Han, we evaluated three different population comparisons to gain a comprehensive genetic overview of the northern-Han Chinese population and nationwide and worldwide reference populations on the basis of the genetic variations of STRs (23-STRs genotype-based data set among 12 Chinese populations, 20-STRs frequency-based dataset among 53 worldwide populations and 19-STRs frequency-based dataset among 61 nationwide populations). Finally, we also collected the present available mitochondrial and Y-chromosomal genetic variations of Han Chinese populations and merged them with previously published uniparental marker variations, as well as combined whole-genome SNPs of modern and Eurasian ancient peoples, to explore the genetic legacy and phylogenetic relationship between northern Han Chinese and ancient populations (Cui et al., 2010; Li et al., 2010; Li et al., 2011; Zhao et al., 2011; Wang et al., 2012; Cui et al., 2013; Zhao et al., 2014; Dong et al., 2015; Gao et al., 2015; Li et al., 2015; Zhao et al., 2015b; Li et al., 2017; Zhang et al., 2017b; Li et al., 2018).

MATERIALS AND METHODS

Sample Collection, Ethics Statements and DNA Preparation

This project, including study design and experimental design, was conducted in conformity with the ethical research principles for medical research involving human subjects recommended by the World Medical Association Declaration of Helsinki and approved by the Ethics Committee of Zunyi Medical University. Blood samples were collected using the FTA card with the written informed consent from 3,089 unrelated healthy Han Chinese individuals (2,009 males and 1,080 females) from Yuncheng country in the Shanxi Province, northern China. A total of 20 EDTA anti-coagulated peripheral blood samples were collected for whole-genome SNPs genotyping. All self-declared Han Chinese indigenous subjects are needed to be no intermarriage with people of other ethnic groups and resided in here at least three generations. Human genomic DNA from male samples was extracted using the DNeasy blood and tissue kit (Qiagen) and measured utilizing the NanoDrop-2000 (Thermo Scientific, USA) following the manufacturer's instruction, and DNA from female samples was isolated using the Chelex-100 method. All our data used throughout this study are submitted in the **Supplementary Materials**.

PCR Amplification and STR Genotyping

We employed Huaxia Platinum Amplification System for multiallelic STR genotyping including 23-autosomal-STRs, two sex-linked inherited Y-InDel (rs2032678) and Amelogenin in a ProFlex PCR System (Thermo Fisher Scientific) in accordance with the manufacturer's recommendation. A 10 μL reaction volume, which contains 4 μL of primer set (the concentration of each locus is different), 4 μL of master mix, 0.4 ng of template DNA as well as Prep-N-Go buffer was employed with the standard thermal cycling conditions which comprise an initial step at 95°C for 1 min; followed by 27 cycles of denaturation at 94°C for 3 s, anneal at 59°C for 16 s, and extension at 65°C for 29 s; and following a final extension at 60°C for 5 min and preservation at 10°C. Applied Biosystems 3500 Genetic Analyzer (Thermo Fisher Scientific) was utilized to separate and detect the PCR products using the 36 cm capillary array and POP-4 polymer following the corresponding instruction. A 9.5 μL of deionized Hi-Di formamide and a 0.5 of μL GeneScan 600 LIZ Size Standard v2.0 (Thermo Fisher Scientific) were utilized to mix with the amplified products or allelic ladder before the capillary electrophoresis. Finally, we used the GeneMapper ID-X v.1.4 software to identify and analyze the corresponding allele. We employed the typical control DNA of 9947A human cell line sample (Thermo Fisher Scientific) as positive, and the ddH$_2$O as negative in each batch of PCR amplification and electrophoresis. Infinium Global Screening Array BeadChip of Illumina WeGene V2 Arrays (WeGene, Shenzhen, China) covering approximately 700k SNPs was used to genotype autosomal SNPs based on the manufacturers' introduction.

Datasets for Population Comparison and Statistical Analysis

To comprehensively dissect the genetic background of Shanxi northern-Han Chinese, we first integrated our raw genotyping data (23 STRs) with previously published genotype data from 11 Chinese populations from five Chinese ethnic groups (Han, Hui, Tibetan, Yi, and Uyghur) (He et al., 2018b; Wang et al., 2018a; Chen et al., 2019; Liu et al., 2019). To further explore the genetic relationship of Shanxi Han in the context of the genetic variations from the worldwide or nationwide populations, we subsequently combined our allele frequency of 20 STRs with publicly obtained data from 52 worldwide populations (Westen et al., 2012; Gaviria et al., 2013; Park et al., 2013; Fujii et al., 2014; Almeida et al., 2015; Parolin et al., 2015; Aguilar-Velazquez et al., 2016; Hossain et al., 2016; Ng et al., 2016; Park et al., 2016; Ramos-Gonzalez et al., 2016; Ristow et al., 2016; Vullo et al., 2016; Wang, 2016; Zhang et al., 2016a; Zhang et al., 2016b; Choi et al., 2017; Guerreiro et al., 2017; Jin et al., 2017; Liu et al., 2017; Moyses et al., 2017; Ossowski et al., 2017; Singh and Nandineni, 2017; Taylor et al., 2017; Wu et al., 2017; Yang et al., 2017a; He et al., 2018b; He et al., 2018e; Wang et al., 2018a; Liu et al., 2019) and allele frequency of 19 STRs with previously investigated the allele frequency distribution from 60 Chinese populations (Zhang et al., 2011; Zhang, 2012; Liu et al., 2013; Shen, 2013; Zhang, 2013; Wang, 2014; Wang et al., 2014; Xie, 2014; Hu et al., 2015; Li, 2015; Ruan, 2015; Shen et al., 2015; Wang, 2015; Yin, 2015; Zhang and Chen, 2015; Zhao et al., 2015a; Huang, 2016; Meng, 2016; Wang, 2016; Xiang, 2016; Xiao et al., 2016; Zhao, 2016; Chen et al., 2017; Fu et al., 2017; He et al., 2017a; He et al., 2017c; Jin et al., 2017;

Liu, 2017; Lu et al., 2017; Yao et al., 2017; Zhang, 2017a; Zhang, 2017b; Zhang, 2017c; Zhang et al., 2017a; Zou et al., 2017; He et al., 2018b; He et al., 2018e; Wang et al., 2018a). Subsequently, to explore the genetic affinity between northern Han Chinese and ancient Asian populations, we combined mitochondrial DNA variations of 812 modern Han Chinese individuals from seven geographical different populations and 417 ancient individuals in 13 different archeological sites, and then combined Y-chromosome variations of 2,810 modern subjects from 26 Chinese populations and 114 ancient individuals in 12 neolithic sites (Cui et al., 2010; Li et al., 2010; Li et al., 2011; Zhao et al., 2011; Wang et al., 2012; Cui et al., 2013; Zhao et al., 2014; Dong et al., 2015; Gao et al., 2015; Li et al., 2015; Zhao et al., 2015b; Li et al., 2017; Zhang et al., 2017b; Li et al., 2018). Finally, we merged our 20 whole-genome SNPs data with previously published 1,924 individuals from the Human Genome Diversity Project-Centre d'Etude du Polymorphisme Humain (HGDP-CEPH) panel and International HapMap Project Phase 3 and 40 ancient humans from Eurasia (Li et al., 2008; Jeong et al., 2016; Lipson et al., 2018).

STR Analysis for Forensics (STRAF) online software (Gouy and Zieger, 2017) was utilized to evaluate the allelic frequency and forensic statistical parameters of 23 STRs. Population genetic parameters based on the raw genotype data, including pairwise Fst genetic distance, locus-specific Fst, and Fis among 12 Chinese populations, were also calculated using the STRAF. Linkage and Hardy-Weinberg equilibrium analyses, as well as evaluation of the heterozygosity indexes, were performed using the Arlequin software (version3.5) (Excoffier and Lischer, 2010). Pairwise standard genetic distances (Nei, 1978; Reynolds et al., 1983; Kalinowski, 2002) between Shanxi Han and other reference populations were calculated using the Phylogeny Inference Packages version 3.5 (PHYLIP) (Excoffier and Lischer, 2010). Principal component analysis (PCA) based on the raw data was carried out using the STRAF, and PCAs on the basis of allele frequency distribution were conducted using the Multivariate Statistical Package (MVSP) version 3.22 software (Kovach, 2007). Genetic similarities and differences revealed by the genetic distances were visualized via the heatmap using the pheatmap program in R software v3.3. Genetic relationships between Shanxi Han and other three different reference population panels were subsequently explored and reconstructed via multidimensional scaling plots (MDS) using the IBM SPSS Statistics 21 (Hansen, 2005) and neighbor-joining tree using the Molecular Evolutionary Genetics Analysis Version 7.0 (Mega 7.0) (Kumar et al., 2016). Under the 'correlated allele frequencies' and 'LOCPRIOR' models, we dissected the ancestry component among 12 Chinese populations using the STRUCTURE version 2.3.4.21(Evanno et al., 2005) with predefined populations ranging from 2~6 with ten replications.

For genome-wide-based population genetic analyses, we used the plink v1.90 to conduct the PCA and ADMIXTURE 1.30 to perform the model-based analysis using the pruned data (-indep-pairwise 200 25 0.4). We employed 10-fold cross-validation with the predefined ancestry populations varying from 2 to 19. We conducted the *admixture-f₃(Source1, Source2; Shanxi* H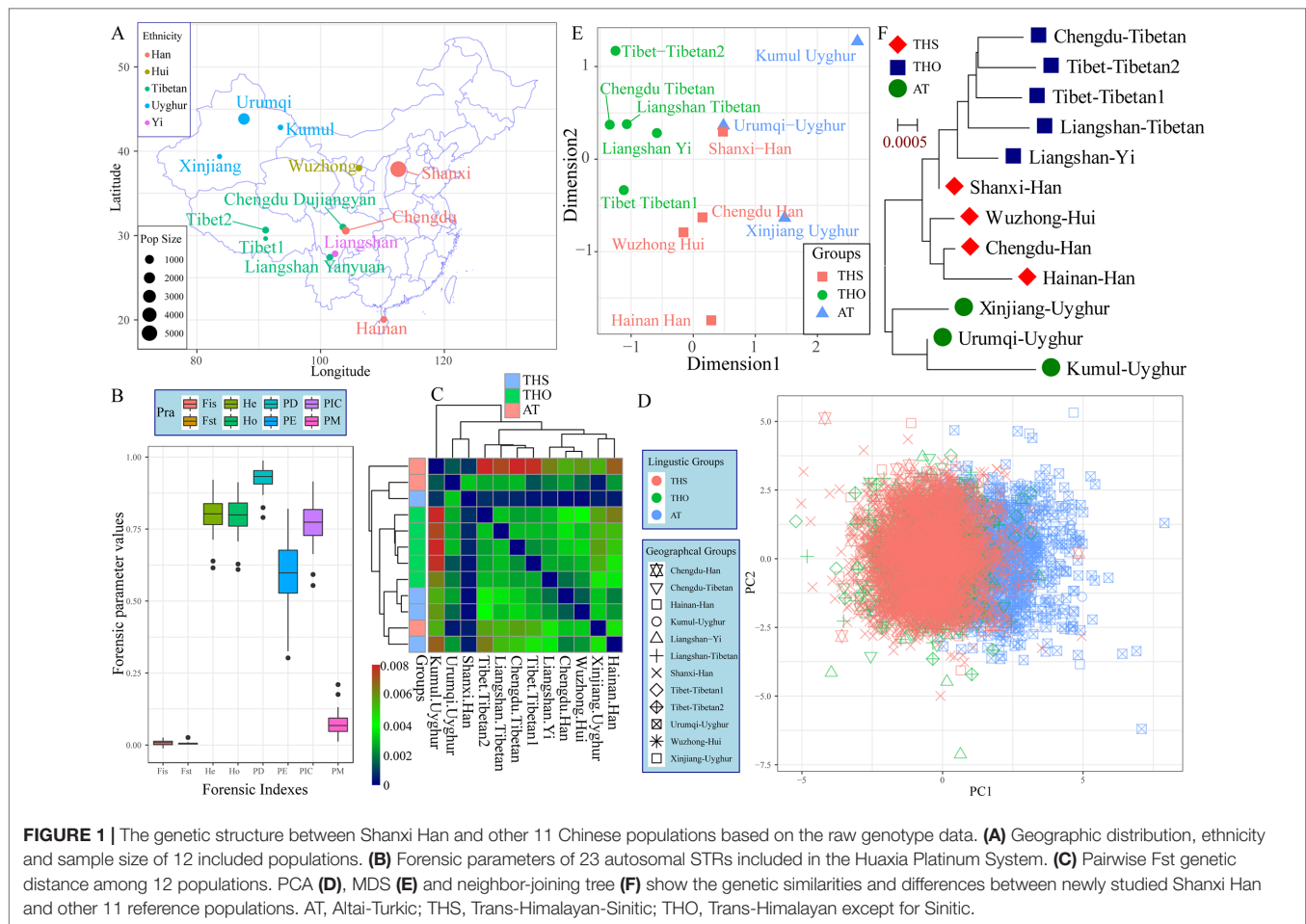an), outgroup-f_3 (X, Shanxi Han; Yoruba), D (X, Y; Shanxi Han, Yoruba), *qpWave* and *qpAdm* to explore the population history of northern Han under the genetic variations of modern and ancient population using the ADMIXTOOLS (Patterson et al., 2012). We finally explored the population splits and mixtures using unsupervised clustering analysis of TreeMix (Pickrell and Pritchard, 2012).

RESULTS

Genetic Structure and Population Genetic Features of Northern-Han Chinese by Raw Microsatellite Data

The 23 autosomal-STRs included in the Huaxia Platinum System were amplified from 3,089 northern-Han Chinese individuals residing in the Shanxi Province, which is considered as the origin place of the Han ethnic group (**Figure 1A**). The p values of Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) in 23 STRs were listed in **Table S1**. After correction using the Bonferroni standard, no deviations from HWE or LD were observed. Allelic frequency distribution and forensic statistical parameters are provided in **Figure 1B** and **Table S2**. A total of 342 alleles were identified with the corresponding allele frequency ranging from 0.0002 to 0.5391. The combined power of discrimination and the combined probability of exclusion in Shanxi Han were 0.999999999999999999999999994 and 0.0.9999999995, respectively. Forensic discrimination powers estimated in the northern-Han Chinese were consistent with forensic measures in the southern-Han populations, such as Zhujiang Han and Sichuan Han. All of the combined results showed that Huaxia Platinum System was informative and polymorphic in Shanxi Han population and could be considered to be a useful tool for forensic kinship identification and individual identification, and Chinese reference database establishment.

To explore the genetic background of the Han Chinese population, we conducted the first population genetic analysis employing a genotype dataset consisting of 6,060 individuals from two language families (Trans-Himalayan and Altaic). Pairwise Fst genetic distances between Shanxi Han and other eleven Chinese populations were listed in **Table S3** and **Figure 1C**. Chengdu Han showed the smallest genetic relationship with Shanxi Han (0.0002) and Urumqi Uyghur showed the largest genetic relationship with Shanxi Han (0.0032). Pairwise Fst results showed that Shanxi Han had a closer genetic relationship with all East Asians compared to the relationships of pairs of the other Chinese populations under the same reference panel. PCA based on the first two components could explain 1.07% variance (PC1: 0.58% and PC2: 0.49%). PC1 could partly distinguish the Altaic-speaking populations and other Chinese references. Both components could not separate any ethnic group from others except for Uyghur (**Figure 1D**). MDS and neighbor-joining tree were constructed on the basis of the pairwise Fst genetic distance matrixes. As shown in **Figure 1E**, five Tibeto-Burman-speaking populations grouped together and localized in the upper left part and Sinitic-speaking populations tended to be allocated in the center. Xinjiang Uyghur and Kumul Uyghur located in the right

FIGURE 1 | The genetic structure between Shanxi Han and other 11 Chinese populations based on the raw genotype data. **(A)** Geographic distribution, ethnicity and sample size of 12 included populations. **(B)** Forensic parameters of 23 autosomal STRs included in the Huaxia Platinum System. **(C)** Pairwise Fst genetic distance among 12 populations. PCA **(D)**, MDS **(E)** and neighbor-joining tree **(F)** show the genetic similarities and differences between newly studied Shanxi Han and other 11 reference populations. AT, Altai-Turkic; THS, Trans-Himalayan-Sinitic; THO, Trans-Himalayan except for Sinitic.

part except for Urumqi Uygur. It is strange to find a closer genetic relationship between Shanxi Han and Urumqi Uyghur, which might be caused by the metropolitan or the provincial center with the massively genetic admixture with adjoining populations in the historical time. As we expected, three obvious genetical affinity clusters were observed which corresponding to linguistic classifications, including the Altaic-Turkic (AT) cluster, Trans-Himalayan-Sinitic (THS) cluster and other Trans-Himalayan populations (THO) cluster, and Shanxi Han was located between the THO and THS, which first grouped with the Liangshan Yi group and then grouped with Han and Hui populations (**Figure 1F**). To directly visualize the genetic component and the corresponding proportion, we conducted Structure analysis assuming 2–6 predefined ancestry populations (**Figure S1**). At k = 2, AT populations showed different ancestry components with other Chinese populations. At k = 3, THO populations showed their specific ancestry component. Geographically different ancestries gradually appeared within the same language family and shared varying proportions of each predefined ancestry when the k values continually increased. Shanxi Han consistently kept a unique ancestry component and harbored a closer genetic relationship with Chengdu Han and Wuzhong Hui than with Hainan Han.
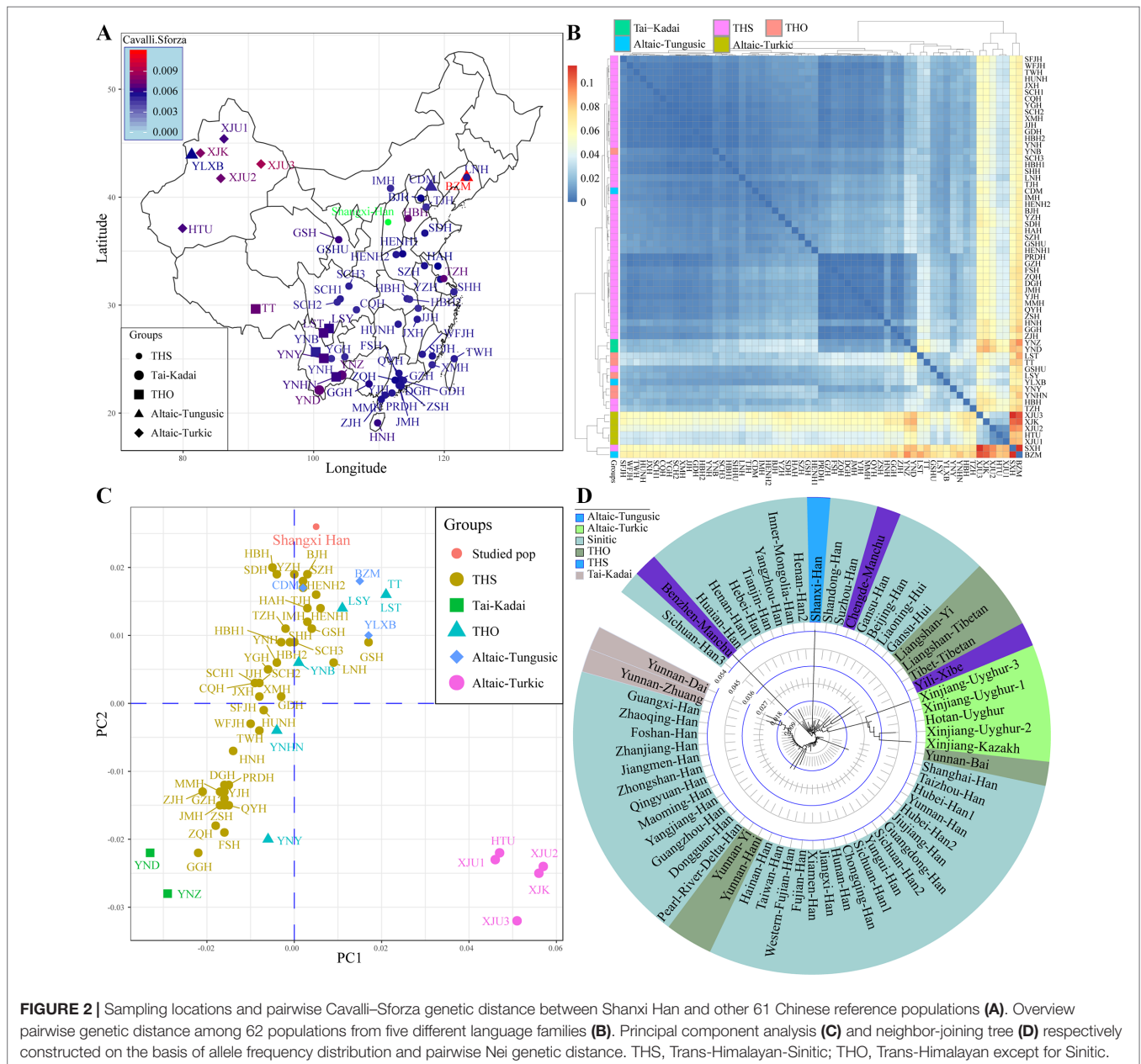
Genetic Heterogeneity and Phylogenetic Relationships Among Worldwide and Nationwide Populations

To characterize the general patterns of genetic similarities and differences between northern-Han Chinese and more reference populations across the whole world, we combined our newly obtained allele frequency data with previously published data from 52 worldwide populations. We calculated two pairwise genetic distances, Nei and Reynolds, among the 53 populations (**Table S4** and **Figure S2A**). Central Chinese Han showed the smallest genetic distances with Shanxi Han (Reynolds: 0.004 and Nei: 0.0013), followed by the southern-Han Chinese populations (Guangdong Han, Sichuan Han, Xiamen, and Guizhou Han), Tibeto-Burman speakers, Turkic speakers, and other continental groups in order. Heatmap furtherly illustrated the genetic affinity existing within East Asian populations (**Figure S2B**). PCA was conducted based on the allele frequency distribution of 20 STRs in 53 populations (**Figures S3** and **4**). Top ten components could extract approximately 84.40% variance (PC1: 32.126%; PC2: 15.892%; PC3: 10.083%; PC4: 8.077%; PC5: 6.523%; PC6: 3.432%; PC7: 2.55%; PC8: 2.104%; PC9: 1.857% and PC10: 1.745%). PC1 displayed a clear separation between East Asians, Central Asians, and others, and PC2 showed differences between Americans and

Europeans with some exceptions. Genetic relationships were also constructed using MDS plots. The observed patterns of relationships were consistent with the aforementioned findings *via* the first three PCA components. The patterns of the genetic affinity of East Asian populations were in accordance with the linguistic classification. Southern-Han Chinese populations were well-positioned at the end of the tree and then clustered with Central-Han Chinese population, followed by clustering with northern-Han Chinese populations (Shanxi Han), finally followed by Japanese, Korean and other language family or continental populations (**Figure S3D**).

Subsequently, focused on the genetic variations in China, we collected previously investigated allele frequency data of 19 STRs from 61 populations and combined with our data. Results from pairwise genetic distances showed that Shanxi Han had a close genetic relationship with Tianjin Han (Cavalli–Sforza: 0.0048, Nei: 0.0514) and other geographically adjacent populations (**Figures 2A**, **B**). The largest genetic distances between Shanxi Han were observed with Benzhen Manchu (Cavalli–Sforza: 0.0118 and Nei: 0.1142), followed by Uyghur and Kazakh populations. MDS was constructed based on the pairwise Nei standard genetic distance matrixes (**Figure 2C**). Five AT-speaking populations were clustered together and allocated in the right-lower position and other Chinese minority groups had a scattered distribution and near to the Han Chinese populations. A significant south-to-north cline was observed and

FIGURE 2 | Sampling locations and pairwise Cavalli–Sforza genetic distance between Shanxi Han and other 61 Chinese reference populations **(A)**. Overview pairwise genetic distance among 62 populations from five different language families **(B)**. Principal component analysis **(C)** and neighbor-joining tree **(D)** respectively constructed on the basis of allele frequency distribution and pairwise Nei genetic distance. THS, Trans-Himalayan-Sinitic; THO, Trans-Himalayan except for Sinitic.

Shanxi Han was clustered closely with northern-Han populations. Phylogenetic relationships within Chinese populations were subsequently constructed (**Figure 2D**). Southern-Han Chinese and southern native minorities formed the first branch, and the AT- and THO-speaking populations formed the second branch. Northern-Han Chinese and northern Manchu groups formed the third branch. Shanxi Han was localized between the neighboring northern populations, such as Yanzhou, Inner Mongolia, Henan2, Shandong and Suzhou Han Chinese populations. The results from different standard genetic distances, PCA, MDS, and phylogenetic relationship reconstruction consistently indicated that Shanxi Han carried a higher resemblance to Beijing Han Chinese than with southern-Han Chinese or other THO-, AT-, Hmong-Mien- and Tai-Kadai-speaking populations.

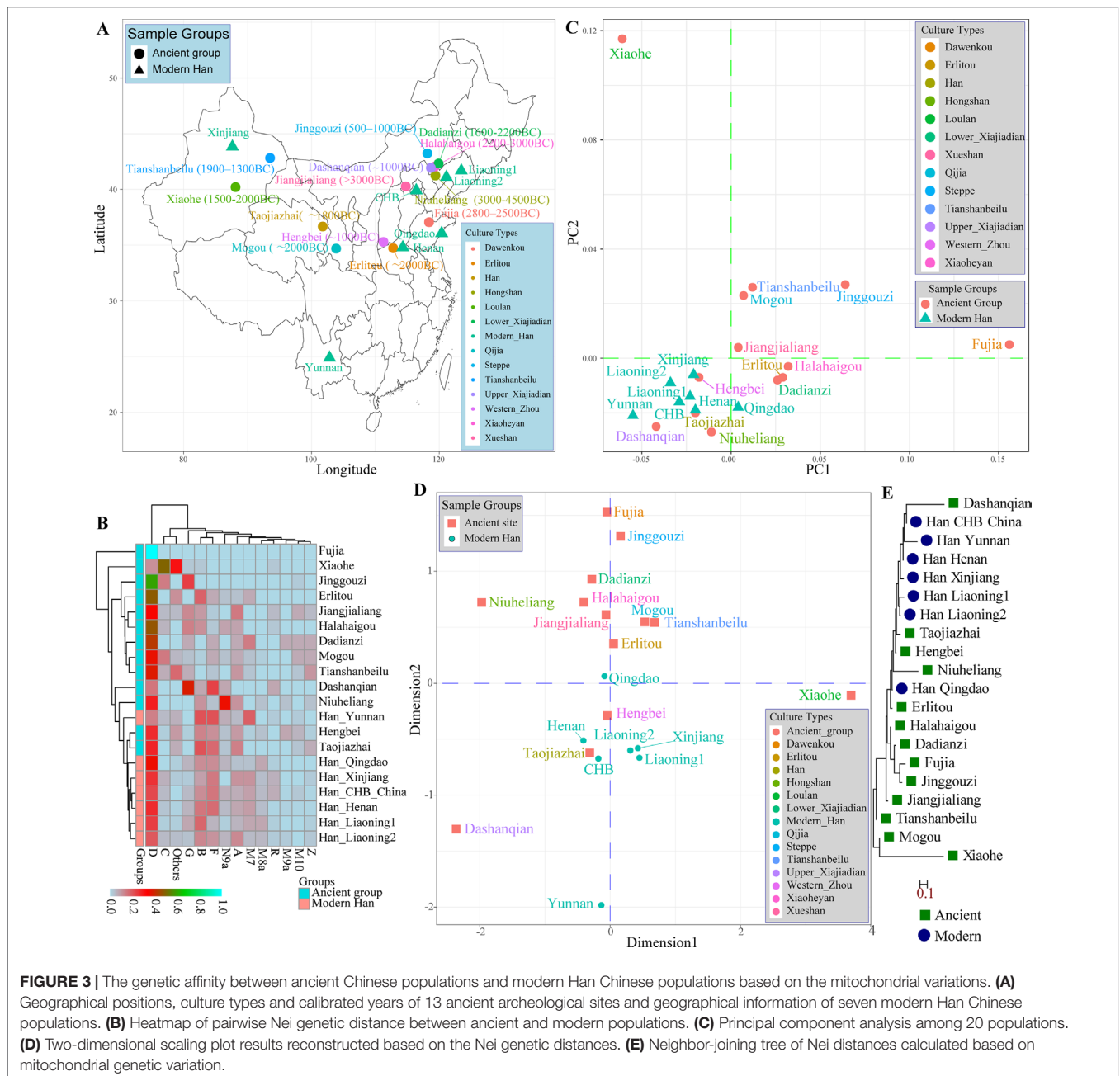## Haplogroup-Based Meta-Analysis of Ancient and Modern Human Populations

A large number of archeological and historical evidence supported that modern Han Chinese population was derived from the central plain area, near to Shanxi Province, and subsequently admixed with neighboring minority groups (Zhao et al., 2014; Zhao et al., 2015b). Fortunately, many genetic studies focused on the genetic architecture of Chinese ancient populations from Neolithic, Bronze, and Iron Age using the uniparental genetic markers (MtDNA and Y-chromosome) had been performed. Therefore, we comprehensively assessed the genetic difference and phylogenetic relationship between modern and ancient Chinese populations using the publicly available prehistoric Chinese archeological samples. We first meta-analyzed mitochondrial variations using 417 ancient samples assembling from 13 archeological sites (Jiangjialiang, Niuheliang, Halahaigou, Dadianzi, Dashanqian, Erlitou, Hengbei, Taojiazhai, Mogou, Xiaohe, Tianshanbeilu, Fuji, and Jinggouzi) and 770 modern northern-Hans from Beijing Han in the 1,000 genomes project and other genetic studies (Henan, Liaoning, Qingdao, and Xinjiang) and 42 southern-Han samples from Yunnan Province. Geographical position and corresponding calibrated ages of archeological sites were presented in **Figure 3A**. We assessed the haplogroup frequency distribution and measured the genetic distances between northern-Han Chinese and adjacent ancient populations. Significant mitochondrial haplogroup frequency differences between modern and ancient groups except for Hengbei and Taojiazhai populations were observed (**Figure 3B**). The relatively smaller genetic distances between ancient and modern populations were observed between Erlitou, Hengbei, Taojiazhai, and other modern Han Chinese populations. The top two components revealed 69.322% variance from 25 populations. Xiaohe population was isolated from other modern and ancient populations, which was consistent with their origin from admixture between European and East Asian (**Figure 3C**) (Li et al., 2015). Hengbei, Taojiazhai, Niuheliang, and Dashanqian populations had a close relationship with modern Han populations. Similar patterns of population distribution model were also reported in the MDS results (**Figure 3D**). Southern Han Chinese from Yunnan kept a separate relationship from the northern-Han Chinese. The phylogenetic relationship between

modern and ancient populations in **Figure 3E** showed a close relationship between Dashanqian ancient population and Beijing Han Chinese and a close relationship between aforementioned three archeological sites (Hengbei, Taojiazhai, and Niuheliang). The Erlitou ancient population was close to Qingdao Han. These ancient populations might be the ancestry population of modern Han Chinese.

To further evaluate the genetic ancestry and genetic relationship between northern-Han and ancient populations from paternal inherited legacy, we collected data from 114 samples obtained from 12 ancient archeological sites (Jiangjialiang, Sanguan, Niuheliang, Halahaigou, Dadianzi, Dashanqian, Jinggouzi, Miaozigou, Hengbei, Taojiazhai, Mogou, and Tianshanbeilu) and then combined with modern data from 2,810 individuals. A total of 80.6% variance extracted from 38 populations showed population substructure existed among the modern Han Chinese population. And Hengbei, Jinggouzi, Dashanqian, Mogou, Taojiazhai, and Sanguan ancient populations showed a paternally close relationship with the modern Han Chinese populations (**Figure S5A**). Fst and corresponding p values between different populations were calculated and submitted in **Table S6**. Significant Y-chromosome haplogroup frequency differences were observed due to the statistically significant differences among 572 out of 703 population comparison pairs. MDS was constructed on the basis of the linearized Fst genetic distance matrix (**Figure S5B**). Nine ancient populations were localized in the upper right position and the other three groups were allocated in the left-center part. Modern Han Chinese populations were clustered in the center and right lower positions. We reconstructed phylogenetic neighbor-joining tree on the basis of the pairwise linearized standard distance, which showed two different clusters: one mainly consisted of modern Han Chinese populations and the other comprised three Taiwan populations, one China Han Chinese, and all ancient populations with the exception of the Hengbei ancient population (**Figure S5C**). The Hengbei ancient population was first clustered with Shanxi and other northern Han Chinese populations (Shaanxi and Heilongjiang).

High-Density Genetic Variations of Modern and Ancient Genomes Show Fine-Scale Population Structure of Northern Han

We additionally investigated the fine-scale genetic structure of Shanxi Han by determining the genetic relationships under the context of 65 worldwide populations (**Figure 4A**). PCA of worldwide populations allocated Shanxi Han at the end of the Eurasian–American genetic cline (**Figure 4B**). And Shanxi Han clustered closely with Beijing Han in the finer scale of variations from East Asia (**Figure 4C**). The observed patterns of genetic affinity were subsequently supported by the results from ADMIXTURE analysis (**Figure 5**). East Asians were a homogeneous population when the predefined ancestry populations are less than eight. Genetic component kept similar between Shanxi Han and Beijing Han, Denver Chinese, Han, and northern Han. The pairwise Fst genetic distances estimated using SNP data were presented in **Table S7**. The smallest genetic

FIGURE 3 | The genetic affinity between ancient Chinese populations and modern Han Chinese populations based on the mitochondrial variations. **(A)** Geographical positions, culture types and calibrated years of 13 ancient archeological sites and geographical information of seven modern Han Chinese populations. **(B)** Heatmap of pairwise Nei genetic distance between ancient and modern populations. **(C)** Principal component analysis among 20 populations. **(D)** Two-dimensional scaling plot results reconstructed based on the Nei genetic distances. **(E)** Neighbor-joining tree of Nei distances calculated based on mitochondrial genetic variation.

distance was identified between Shanxi Han and Beijing Han (Fst = 0.0006), followed by Tujia and another northern-Han (**Figure 6A**). *Outgroup-f₃* in the form of f_3 *(X, Shanxi Han; Yoruba)* showed that the greater genetic affinity identified between Shanxi Han and Han Chinese populations, subsequently followed by the southern Tai-Kadai and Hmong-Mien speakers, and Western Trans-Himalayan and northern Altaic speakers (**Figures 6B**, **C**). We subsequently estimated the *D-statistics* in the form of D *(X, Y; Shanxi Han, Yoruba)*, where X represents the worldwide populations and Y denotes the Chinese populations from different language families, to explore the status of allele sharing. Our results provided supporting evidence for more

shared genetic drift between Shanxi Han and northern-Han or neighboring minorities (**Figure 7** and **S6** and **S7**). We used the admixture f_3 (Source1, Source2; Shanxi Han) to find the potentially admixed ancestral populations. Two hundred sixty-four out of 2,016 pairs were observed with significant negative values (**Figure 8**). The potential ancestry populations revealed by the *admixture-f₃* indicated that the ancestral populations of Shanxi-Han derived their ancestry from southern Chinese-related population (Ancestral Southeast Asians) and East-Siberian-related population (Ancestral Northeast Asians), like the two ancestry populations observed in Indian (Ancestral North Indians and Ancestral South Indians) (Reich et al., 2009).

FIGURE 4 | Geographic position **(A)** and PCA results **(B** and **C)** among 65 worldwide populations.
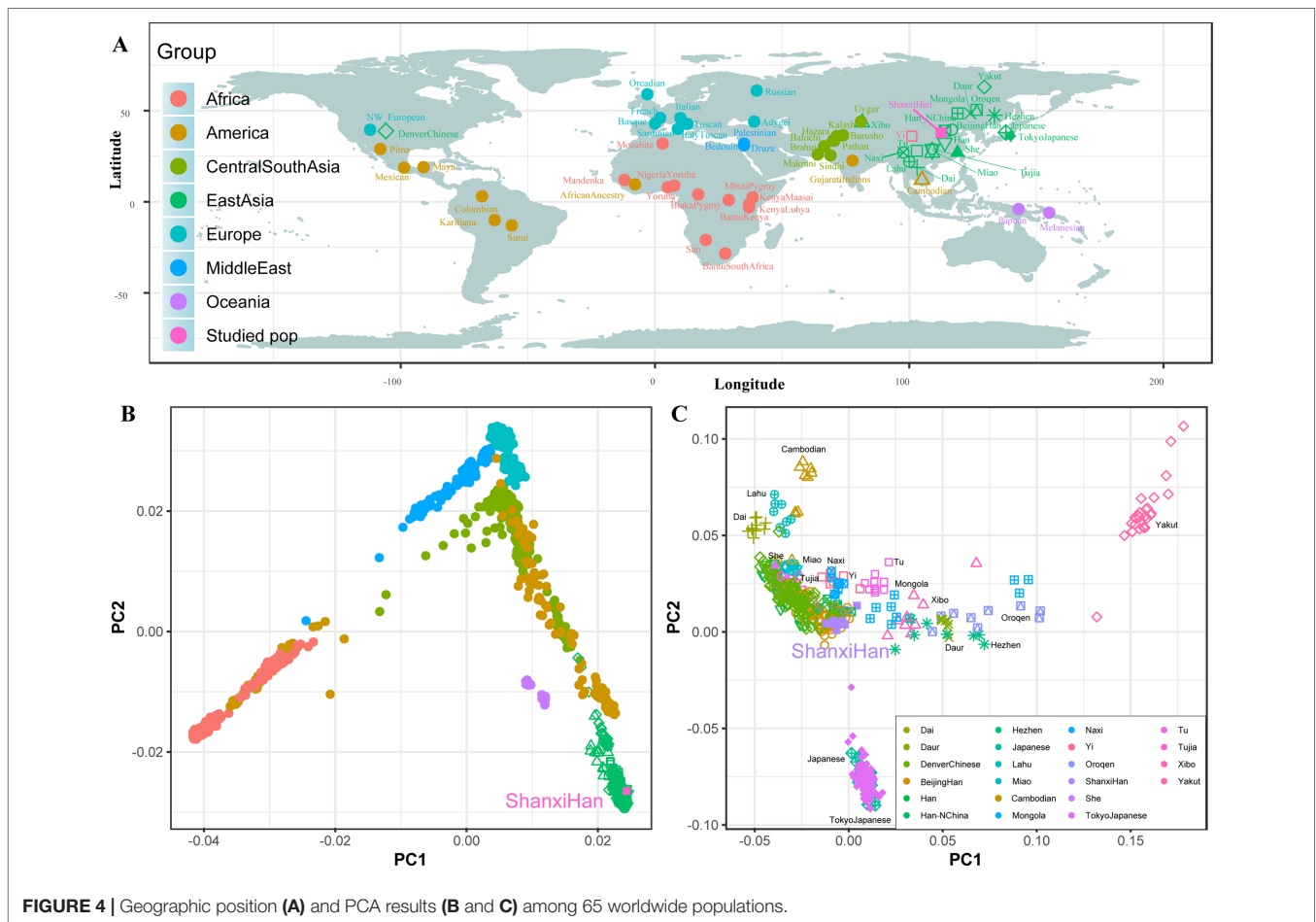
To further validate the minimum streams of ancestry populations and evaluate the corresponding admixture proportion, we first performed the TreeMix analyses among 65 worldwide populations (**Figures 9** and **S8**) and 25 Asian populations (**Figure S9**). A larger number of recent admixtures or migrations were observed in our TreeMix model. Considering the statistical significance ($f_3 = -0.0018$ and $Z = -6.662$) was observed in the form of $f_3(She, Yakut; Shanxi Han)$, we then conducted the *qpWave* and *qpAdm* using the She and Yakut as source populations and using Yoruba, San, Papuan and Melanesian as outgroup populations. *QpWave* results (p = 0.052 for rank1) indicated that Shanxi Han was derived from two ancestral populations. The *qpAdm* analysis further suggested that Shanxi Han has derived 25.2% Yakut-related ancestry and 74.8% She-related ancestry.

Nineteen ancient populations from Eurasian were employed to explore the genetic admixture history between Shanxi Han. The shared genetic history of all pairs was presented in **Figure S10**. Denisovan and Vindija Neanderthal shared smaller genetic components with others. We found that Shanxi Han kept a distant genetic relationship with two archaic human populations and Italy Iceman, and shared more alleles with DevilsGate (0.2039) and other ancient populations from Southeast Asia and Nepal (**Figure 10**). DevilsGate, Oakaie showed significantly negative f_3

value in the form of $f_3(A, B; Shanxi Han)$ (**Table S8**). We finally used *qpWave* to find that the minimum ancient ancestry streams modern northern-Han were 2 (rank1: p = 0.118). *QpAdm* further showed that DevilsGate Hunter-Gatherer-related population contributed 45% ancestry and Oakaie-related ancient population contributed 55% ancestry to modern northern-Han Chinese.

DISCUSSION

East Asia is occupied by anatomically modern human 50 kya when they migrated out of Africa. These regions are populated by the hunter–gatherer over 40 kya in the Paleolithic time (Nielsen et al., 2017). Under the natural selections from different environments, substance strategies and disease pathogens, the different ethnic group formed their specific genetic structure with a different culture, appearance, and language. In the Neolithic time, agriculture originated from the Yellow, Yangtze and Zhujiang River Basins, may be also included in Liaohe, promoted the process of population genetic structure change with different cultures (Piao et al., 2010). For China in East Asia, the world's largest ethnic group of the Han Chinese, 55 officially recognized and several unrecognized ethnic groups are subsequently formed with their specific cultural background.
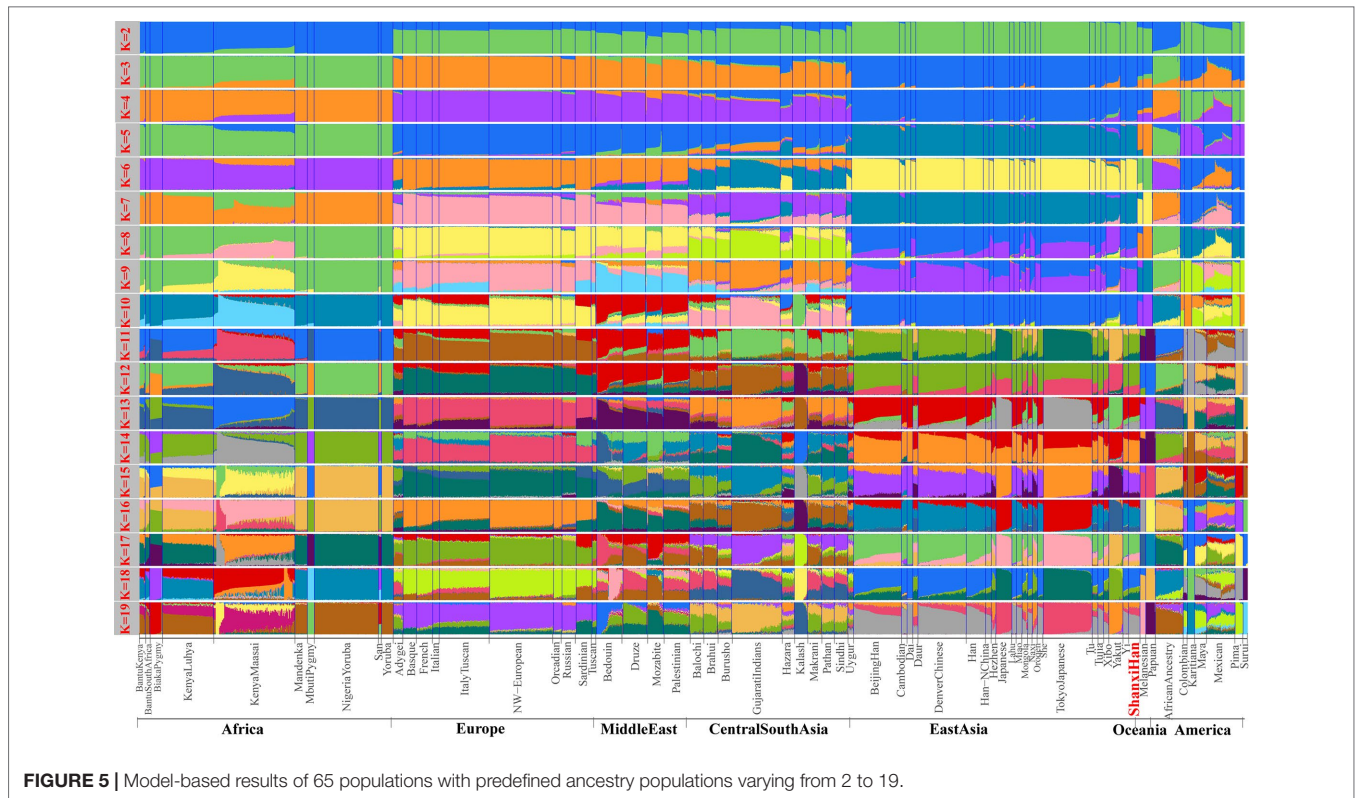
FIGURE 5 | Model-based results of 65 populations with predefined ancestry populations varying from 2 to 19.

The languages they used in this region include over nine language families: Indo-European, Altaic (also called "Trans-Eurasian," including Turkic, Mongolic, and Tungusic language groups) mainly distrused in the north; Tai-Kadai, Hmong-Mien, Austronesian, Austroasiatic, and Trans-Himalayan language families in the south. Population substructures among Chinese modern populations revealed by our autosomal STR panel have supported the patterns of population relationship found by the X and Y-chromosome markers, as well as ancestry-informative single nucleotide polymorphisms (He et al., 2017b; He et al., 2018a; He et al., 2018c). Here, our study presented the first comprehensive genetic analysis, including autosomal STR, a meta-analysis of mitochondrial and Y-chromosomal haplogroup distribution, and autosomal SNP data of northern-Han Chinese residing in Shanxi Province. We performed the comprehensive population comparison to investigate the origin, genetic legacy and detailed genetic relationship of modern Han Chinese population, especially for the northern Han Chinese. Our results showed that Altaic and Trans-Himalayan speakers except for Sinitic speakers harbored considerable genetic differences with Han Chinese populations. However, no apparent genetic differentiation between Hmong-Mien-, Tai-Kadai and neighboring Han Chinese populations is revealed in our present study (**Figure 1**). Analysis from the haplogroup distribution of Neolithic Chinese populations showed that the significant association of genetic continuity between ancient populations from Yellow River Valley sites (Mogou, Taojiazhai, and Hengbei) and modern northern-Han Chinese (**Figures 2** and **S5**). Whole-genome high-density SNP data illustrate that Shanxi Han Chinese

inherited 25.2% their ancestry from Yakut-related population and 74.8% from She-related population. Ancient autosomal genetic variation subsequently shows a two-way admixture from ancient North East Asian (45% ancestry from DevilsGate Hunter-Gatherer-related population) and ancient South Asian (55% ancestry Oakaie-related ancient population). These results consistently showed a more complex and ancient population admixture history of northern Han Chinese. Han Chinese may be originated from the admixture between the ancient Tibeto-Burman population and a local pre-Sinitic population which may have been linguistically Altaic in the Neolithic time when agriculture emerged in Yangtze and Yellow River Basins.

A recent large-scale whole-genome variation study covering 11,670 Han Chinese individuals from 24 out of 33 administrative regions was carried out to explore the Han Chinese population genetic structure and genetic ancestry (Chiang et al., 2018). Their valuable finding of east-west genetic distinction among Han Chinese is one indispensable previously unrecognized population structure, which is perfectly complemented the north-south differentiation previously found by Xu et al. (2009). Using high-density SNP typing data and other scientists using uniparental markers and classical markers (Xu et al., 2009; Stoneking and Delfin, 2010; Sanchez-Mazas et al., 2011). Our results in this study on the basis of the nationwide STR variations also provide the microsatellite evidence for north-south genetic cline but fail to reveal the east-to-west difference, which may be caused by the sample coverage. Lu et al. (2016) whole-genome sequenced 39 Han Chinese and 38 Tibetan individuals to investigate the gene pool of the Tibetan and Han
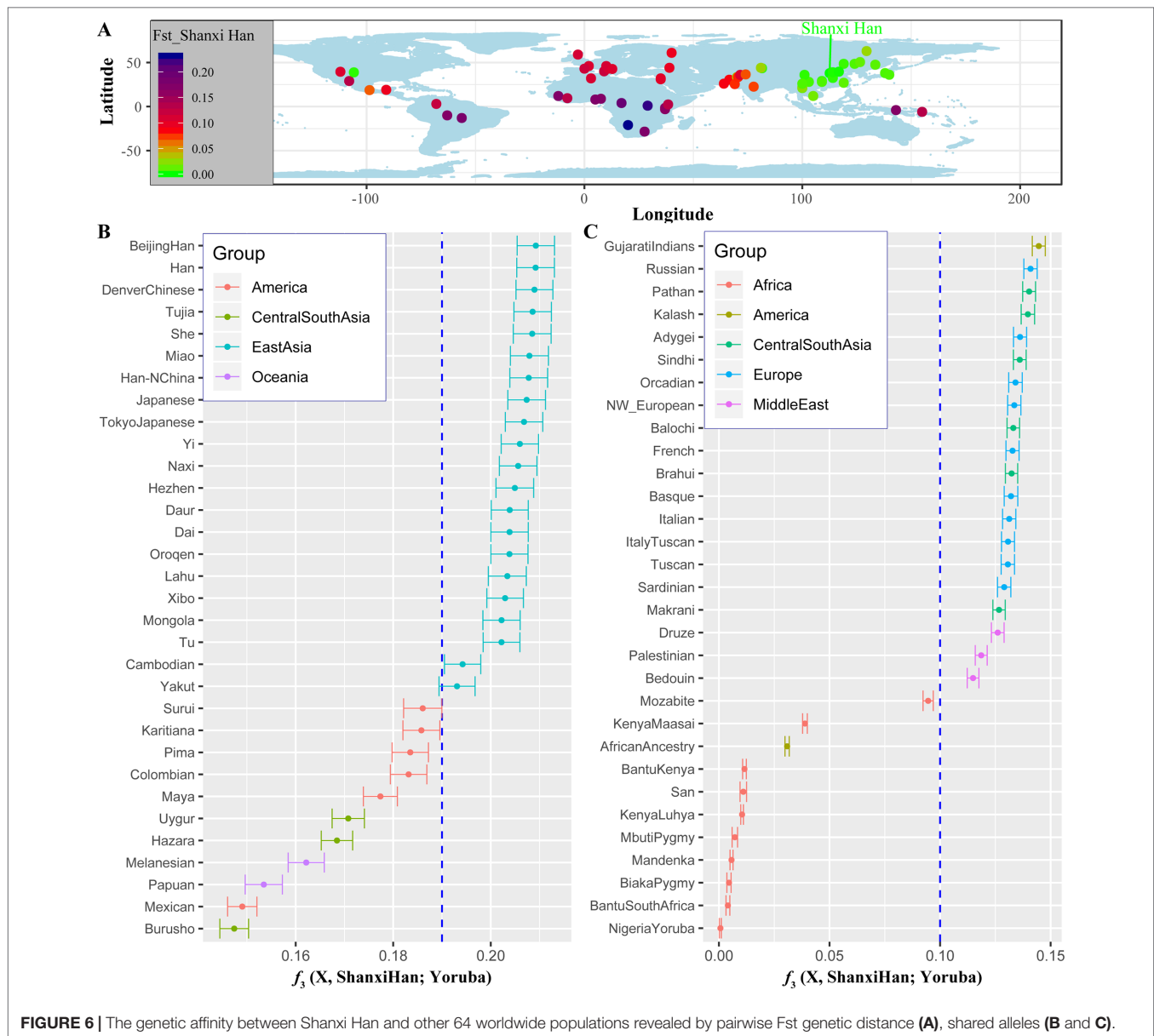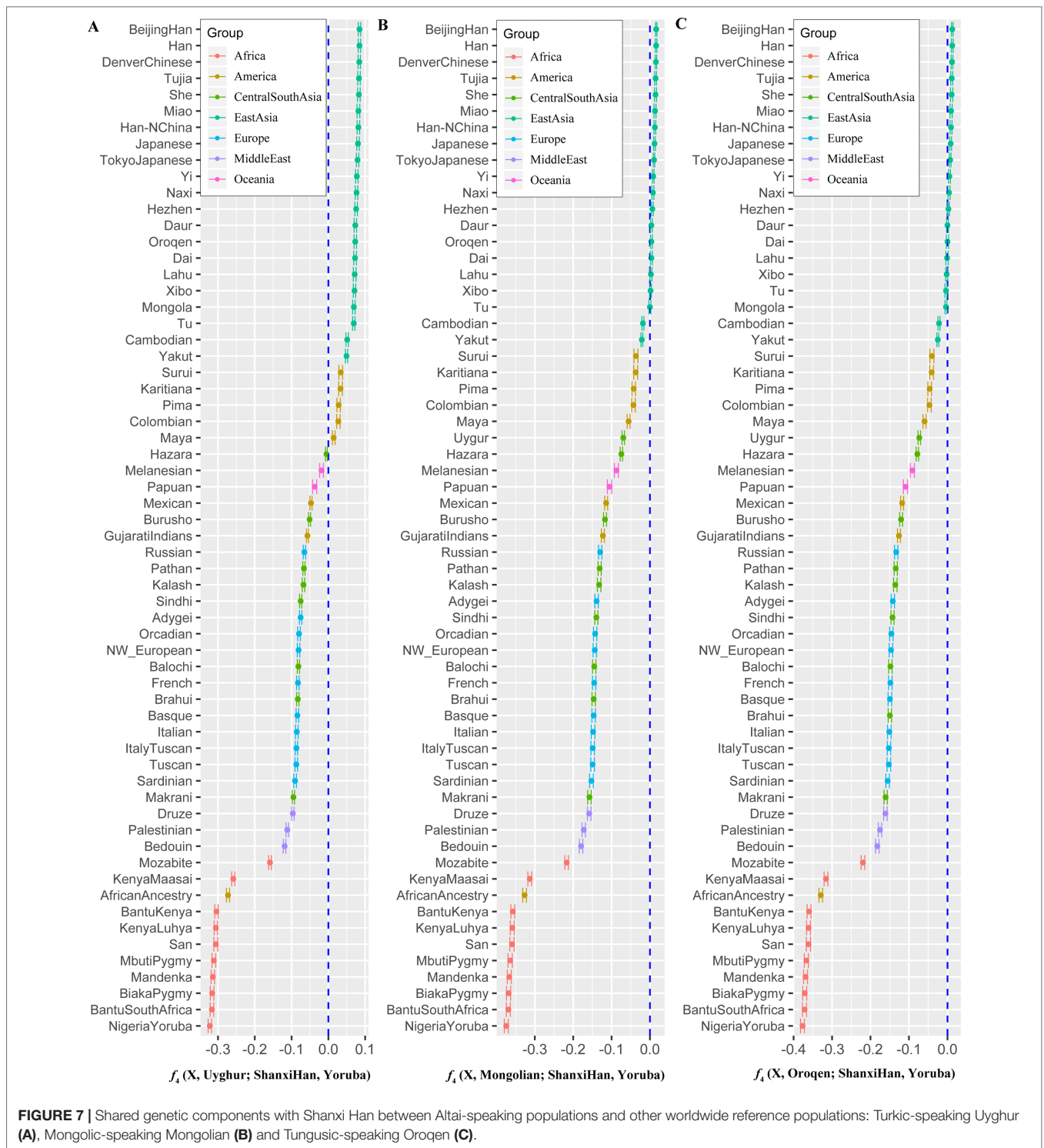
FIGURE 6 | The genetic affinity between Shanxi Han and other 64 worldwide populations revealed by pairwise Fst genetic distance **(A)**, shared alleles **(B and C)**.

group. They found that Tibetan and Han Chinese diverged from each other at 7,000–13,000BC during the last glacial maximum (Lu et al., 2016). Wang et al. (2018) recently also tried to investigate and elucidate the precise divergence time, genetic structure and admixture history between Han Chinese and neighboring country populations (Japanese and Korean). Their results suggested that Han Chinese and other focused two populations split approximately 1,000–1,600 BC in the Shang dynasty in Chinese history and subsequently substantial genetic admixture between them and other adjacent populations have occurred (Wang et al., 2018b).

The processes of ancient whole-genome DNA studies with the technological innovations of DNA hybridization enrichment and next-generation sequencing has revolutionized the phylogenetic relationship and population history reconstruction

in the European, American, Oceanian and even southeast Asians (Nielsen et al., 2017). In East Asia, just two projects respectively focused on one 40,000-year-old individual from Tianyuan cave and two hunter–gatherers from Devil's Gate have been performed. Yang et al. (2017) sequenced the whole-genome of Tianyuan ancient people (40,000BP) and discovered a strong genetic affinity between these ancient people and present populations, which indicated that there is a genetic continuity or population turnover in the East Asian continent (Yang et al., 2017b). Siska et al. (2017) genome-wide analyzed two Devil's Neolithic individuals (~7,700BP) near to the Amur basin and also detected the genetic continuity in northeast Asia. If ancient people from Paleolithic, Neolithic, Bronze, and Iron Ages in East Asia are all sequenced and conducted corresponding population history reconstruction combined with the historical, cultural,

FIGURE 7 | Shared genetic components with Shanxi Han between Altai-speaking populations and other worldwide reference populations: Turkic-speaking Uyghur **(A)**, Mongolic-speaking Mongolian **(B)** and Tungusic-speaking Oroqen **(C)**.

linguistic and archeological findings, a complete genetic landscape of the East Asians will be obtained. However, a number of ancient people excavated from different archeological sites in China have so far received little attention. Fortunately, there still some exploratory projects focused on the genetic variations of the uniparental markers (mtDNA and Y-chromosome)

and Neolithic or historical ancient people been carried out. Thus, we can perform the first meta-analysis to investigate the phylogenetic relationship between the ancient population and modern northern-Han Chinese population.

Our present meta-analysis results from the Neolithic ancient people and modern Han Chinese on the basis of the

FIGURE 8 | Admixture f_3 results with significant negative f_3 values.
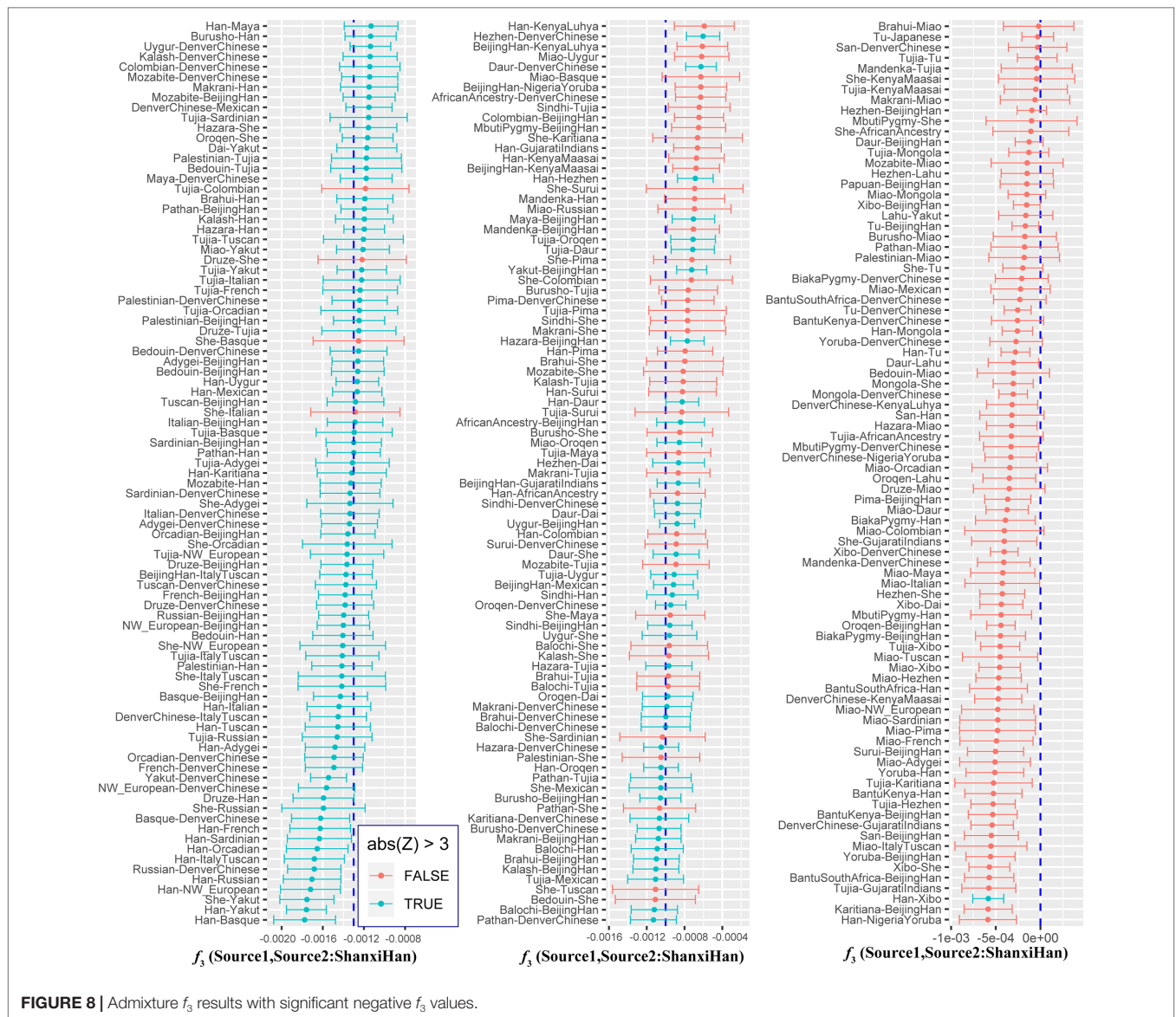
combined genetic variations of mtDNA and Y-chromosome first showed that the ancient populations from West Liao River Valley sites (Dasanqian and Niuheliang) and Yellow River Valley sites (Hengbei, Taojiazhai) share considerable similar mitochondrial haplogroup with the modern northern-Han Chinese populations. For Y-chromosome variations, ancient people from the Hengbei site shared the more significant genetic similarity with modern northern-Han Chinese from Shanxi and Heilongjiang provinces, and Dashanqian and Mogou ancient people bear a similar genetic assemblage with modern Taiwan Han people. Mogou site in the Ganqing region adjacent to the central plain is the hometown of Di-Qiang people who are thought as the direct ancestral population of Han Chinese, which is genetically close to the Han Chinese population. Our results reveal a close genetic relationship among Hengbei, Mogou and modern northern-Han Chinese

populations. Our findings combined with the archeological, historical and linguistic evidence consistently supported the admixed genetic origin of modern Han Chinese.

PERSPECTIVE

In summary, we genotyped 23-autosomal-STRs in 3,089 Shanxi northern-Han Chinese individuals and provided the first batch of allele frequency, forensic and population genetic parameters of northern Han Chinese. Comprehensive worldwide and nationwide population comparisons not only showed that Shanxi harbored a strong similar genetic assemblage with adjacent Han populations but also illustrated that there were apparent genetic distinctions between north-to-south Han Chinese as well as genetic differentiation
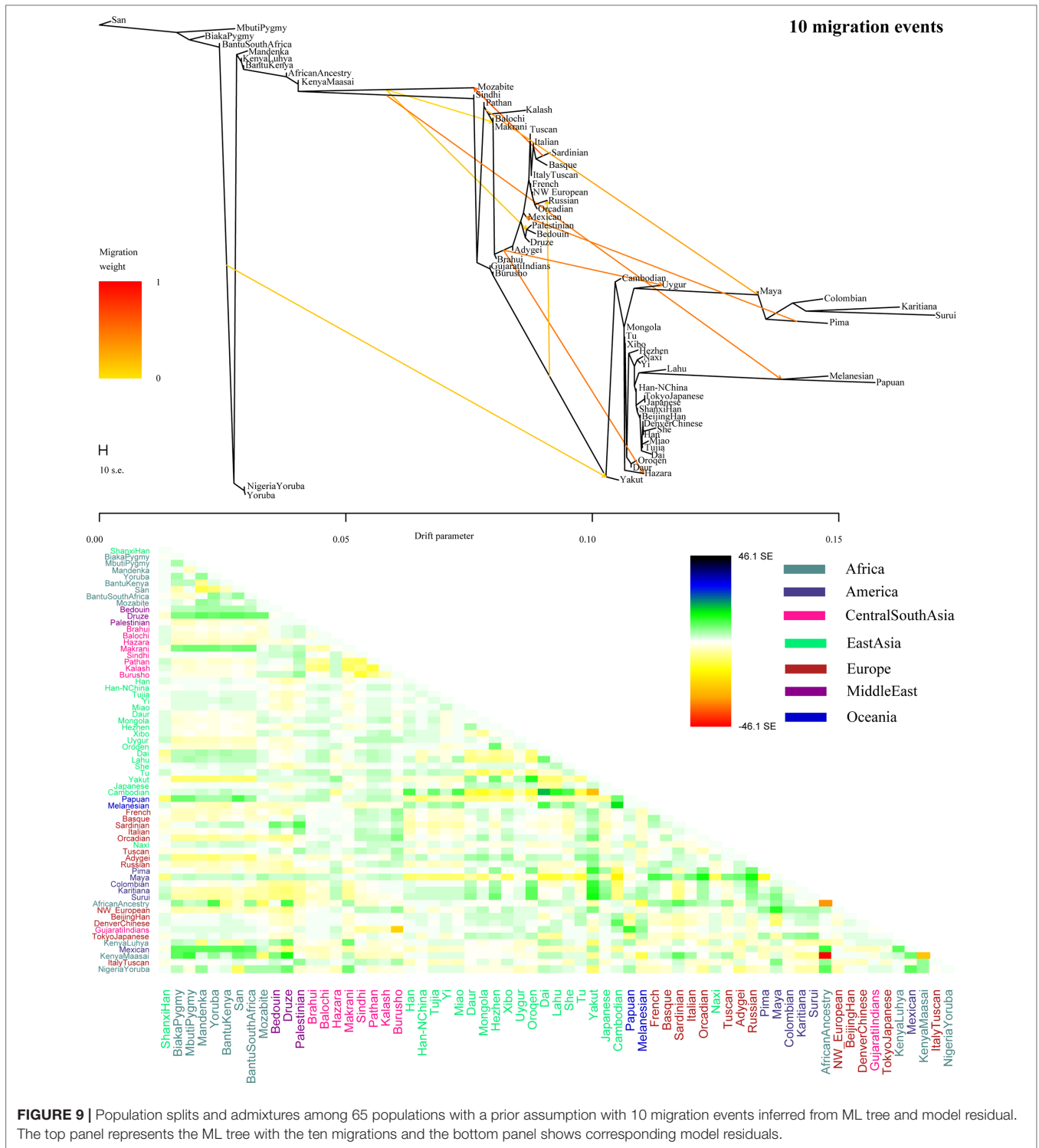
FIGURE 9 | Population splits and admixtures among 65 populations with a prior assumption with 10 migration events inferred from ML tree and model residual. The top panel represents the ML tree with the ten migrations and the bottom panel shows corresponding model residuals.

between populations belonging to different language families, obviously differences observed between Tibetan, Uyghur, and others here. The first meta-analysis based on the mitochondrial and Y-chromosomal genetic variations among ancient and modern Asian populations showed a genetic affinity and genetic continuity between Mogou, Hengbei ancient population and present-day northern-Han Chinese. We also found Neolithic agriculture expansion related Dashanqian and Niuheliang ancient populations are genetically close to modern northern Han. The qpWave/qpAdm modeling further
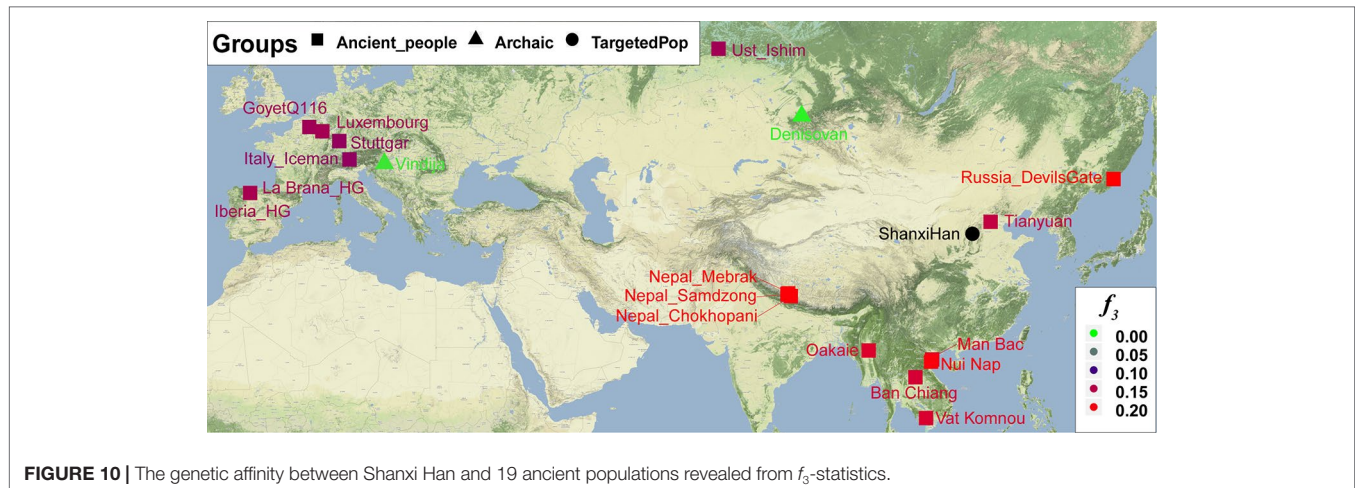
FIGURE 10 | The genetic affinity between Shanxi Han and 19 ancient populations revealed from f_3-statistics.

revealed that modern northern Han Chinese carried 74.8% She-related ancestry and 25.2% Yakut-related ancestry. Both Hengbei-associated and Tibetan-related uniparental lineage (D haplogroup) were observed in modern Northern Han Chinese. Besides, approximately 45% DevilsGate-like ancestry, one Tungusic-affiliated Neolithic population, was modeled *via* ancient DNA. Summarily, consistent with previous linguistic and archaeological evidence, the genetic mixing that led to the emergence of a Han Chinese ethnicity occurred at a very early period, probably in Neolithic times, and this mixing involved an ancient Tibeto-Burman population and a local pre-Sinitic population, which may have been linguistically Altaic. Fine-scale population history reconstruction of north Han from modern and ancient genomes consistently model their ancestral populations deriving from ancestral North East Asian and ancestral South East Asian.

ETHICS STATEMENT

This study was approved by the Ethics Committee of Zunyi Medical University and corresponding experiments have followed the recommendations of the World Medical Association

Declaration of Helsinki. Informed consent was obtained before the sample collection from the participants.

AUTHOR CNTRIBUTIONS

GH, PC, and FJ conceived the idea for the study. GH, PC, and FJ performed or supervised wet laboratory work. PC, JW, LL, HG, XZ, HL, LY, GH, MW, YH, GC, and YL analyzed the data. GH wrote and edited the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01045/full#supplementary-material

REFERENCES

Aguilar-Velazquez, J. A., Martinez-Cortes, G., Inclan-Sanchez, A., Romero-Renteria, O., Diaz-Navarro, X. X., and Rangel-Villalobos, H. (2016). Population data of 23 STR loci (PowerPlex(R) Fusion System) in Mexican Mestizos from the West Region. *Int. J. Legal Med.* 130, 1489–1491. doi: 10.1007/s00414-016-1361-0

Almeida, C., Ribeiro, T., Oliveira, A. R., Porto, M. J., Costa Santos, J., Dias, D., et al. (2015). Population data of the GlobalFiler ((R) Express loci in South Portuguese population. *Forensic Sci. Int. Genet.* 19, 39–41. doi: 10.1016/j.fsigen.2015.06.001

Chen, J., Zheng, H., Bei, J. X., Sun, L., Jia, W. H., Li, T., et al. (2009). Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *Am. J. Hum. Genet.* 85, 775–785. doi: 10.1016/j.ajhg.2009.10.016

Chen, P., He, G., Zou, X., Wang, M., Jia, F., Bai, H., et al. (2018). 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., Wang, B., Gao, B., and He, G. (2019). Forensic features and genetic structure of 23 autosomal STRs in Artux Turkic-speaking population residing in southwestern Xinjiang Uyghur Autonomous Region. *Int. J. Legal Med.* 133, 1393–1395. doi: 10.1007/s00414-019-02072-7

Chen, P., Wang, B., He, G., Song, F., Luo, H., Zhang, L., et al. (2017). Population genetic analyses of 20 autosomal STR loci in Chinese Han population from Tianjin in Northern China. *Forensic Sci. Int. Genet.* 27, 184–185. doi: 10.1016/j.fsigen.2016.11.010

Chiang, C. W. K., Mangul, S., Robles, C., and Sankararaman, S. (2018). A Comprehensive map of genetic variation in the World's Largest Ethnic Group-Han Chinese. *Mol. Biol. Evol.* 35, 2736–2750. doi: 10.1093/molbev/msy170

Choi, E.-J., Park, K.-W., Lee, Y.-H., Nam, Y.-H., Suren, G., Ganbold, U., et al. (2017). Forensic and population genetic analyses of the GlobalFiler STR loci

in the Mongolian population. *Genes Genomics* 39, 423–431. doi: 10.1007/s13258-016-0511-6

Chu, J. Y., Huang, W., Kuang, S. Q., Wang, J. M., Xu, J. J., Chu, Z. T., et al. (1998). Genetic relationship of populations in China. *Proc. Natl. Acad. Sci. U. S. A.* 95, 11763–11768. doi: 10.1073/pnas.95.20.11763

Cui, Y., Li, C., Gao, S., Xie, C., and Zhou, H. (2010). Early Eurasian migration traces in the Tarim Basin revealed by mtDNA polymorphisms. *Am. J. Phys. Anthropol.* 142, 558–564. doi: 10.1002/ajpa.21257

Cui, Y., Li, H., Ning, C., Zhang, Y., Chen, L., Zhao, X., et al. (2013). Y Chromosome analysis of prehistoric human populations in the West Liao River Valley, Northeast China. *BMC Evol. Biol.* 13, 216. doi: 10.1186/1471-2148-13-216

Dong, Y., Li, C., Luan, F., Li, Z., Li, H., Cui, Y., et al. (2015). Low Mitochondrial DNA diversity in an ancient population from china: insight into social organization at the Fujia site. *Hum Biol* 87, 71–84. doi: 10.13110/humanbiology.87.1.0071

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x

Excoffier, L., and Lischer, H. E. (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, X., He, J., Guo, J., Liu, Y., Wang, Z., Lv, J., et al. (2017). Genetic distribution of 21 STR loci in the Chinese Han population from Hainan Island using Goldeneye DNA ID 22NC Kit. *Forensic Sci. Int. Genet.* 29, e19–e20. doi: 10.1016/j.fsigen.2017.04.014

Fujii, K., Iwashima, Y., Kitayama, T., Nakahara, H., Mizuno, N., and Sekiguchi, K. (2014). Allele frequencies for 22 autosomal short tandem repeat loci obtained by PowerPlex Fusion in a sample of 1501 individuals from the Japanese population. *Leg. Med. (Tokyo)* 16, 234–237. doi: 10.1016/j.legalmed.2014.03.007

Gao, S. Z., Zhang, Y., Wei, D., Li, H. J., Zhao, Y. B., Cui, Y. Q., et al. (2015). Ancient DNA reveals a migration of the ancient Di-qiang populations into Xinjiang as early as the early Bronze Age. *Am. J. Phys. Anthropol.* 157, 71–80. doi: 10.1002/ajpa.22690

Gaviria, A., Zambrano, A. K., Morejon, G., Galarza, J., Aguirre, V., Vela, M., et al. (2013). Twenty two autosomal microsatellite data from Ecuador (Powerplex Fusion). *Forensic Sci. Int. Genet. Suppl. Ser.* 4, e330–e333. doi: 10.1016/j.fsigss.2013.10.169

Ge, J., Sun, H., Li, H., Liu, C., Yan, J., and Budowle, B. (2014). Future directions of forensic DNA databases. *Croat. Med. J.* 55, 163–166. doi: 10.3325/cmj.2014.55.163

Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393

Gouy, A., and Zieger, M. (2017). STRAF-A convenient online tool for STR data evaluation in forensic genetics. *Forensic Sci. Int. Genet.* 30, 148–151. doi: 10.1016/j.fsigen.2017.07.007

Guerreiro, S., Ribeiro, T., Porto, M. J., Carneiro De Sousa, M. J., and Dario, P. (2017). Characterization of GlobalFiler loci in Angolan and Guinean populations inhabiting Southern Portugal. *Int. J. Legal Med.* 131, 365–368. doi: 10.1007/s00414-016-1497-y

Hansen, J. (2005). Using SPSS for windows and macintosh: analyzing and understanding data. *Am. Statistician* 59, 113–113. doi: 10.1198/tas.2005.s139

He, G.-L., Chen, P.-Y., Gao, B., Han, Y.-Y., Zou, X., Song, F., et al. (2017a). Phylogenetic analysis among 27 Chinese populations and genetic polymorphisms of 20 autosomal STR loci in a Chinese Uyghur ethnic minority group. *Aust. J. Forensic Sci.* 50 (5), 493–502. doi: 10.1080/00450618.2016.1275793

He, G., Chen, P., Zou, X., Chen, X., Song, F., Yan, J., et al. (2017b). Genetic polymorphism investigation of the Chinese Yi minority using PowerPlex(R) Y23 STR amplification system. *Int. J. Legal Med.* 131, 663–666. doi: 10.1007/s00414-017-1537-2

He, G., Li, Y., Wang, Z., Liang, W., Luo, H., Liao, M., et al. (2017c). Genetic diversity of 21 autosomal STR loci in the Han population from Sichuan province, Southwest China. *Forensic Sci. Int. Genet.* 31, e33–e35. doi: 10.1016/j.fsigen.2017.07.006

He, G., Li, Y., Zou, X., Zhang, Y., Li, H., Wang, M., et al. (2018a). X-chromosomal STR-based genetic structure of Sichuan Tibetan minority ethnicity group and its relationships to various groups. *Int. J. Legal Med.* 132, 409–413. doi: 10.1007/s00414-017-1672-9

He, G., Wang, M., Liu, J., Hou, Y., and Wang, Z. (2018b). Forensic features and phylogenetic analyses of Sichuan Han population *via* 23 autosomal STR loci included in the Huaxia Platinum System. *Int. J. Legal Med.* 132, 1079–1082. doi: 10.1007/s00414-017-1679-2

He, G., Wang, Z., Wang, M., Luo, T., Liu, J., Zhou, Y., et al. (2018c). Forensic ancestry analysis in two Chinese minority populations using massively parallel sequencing of 165 ancestry-informative SNPs. *Electrophoresis* 39, 2732–2742. doi: 10.1002/elps.201800019

He, G., Wang, Z., Wang, M., Zou, X., Liu, J., Wang, S., et al. (2018d). Genetic variations and forensic characteristics of Han Chinese population residing in the Pearl River Delta revealed by 23 autosomal STRs. *Mol. Biol. Rep.* 45, 1125–1133. doi: 10.1007/s11033-018-4264-y

He, G., Wang, Z., Wang, M., Zou, X., Liu, J., Wang, S., et al. (2018e). Genetic variations and forensic characteristics of Han Chinese population residing in the Pearl River Delta revealed by 23 autosomal STRs. *Mol. Biol. Rep.* 45 (5), 1125–1133. doi: 10.1007/s11033-018-4264-y

He, G., Wang, Z., Yan, J., and Hou, Y. (2017d). Chinese population genetic substructure using 23 Y-chromosomal STRs. *Forensic Sci. Int. Genet. Suppl. Ser.* 6, e110–e111. doi: 10.1016/j.fsigss.2017.09.038

Hossain, T., Hasan, M., Mazumder, A. K., Momtaz, P., Sufian, A., Khandaker, J. A., et al. (2016). Genetic polymorphism studies on 22 autosomal STR loci of the PowerPlex fusion system in Bangladeshi population. *Leg. Med. (Tokyo)* 23, 44–46. doi: 10.1016/j.legalmed.2016.09.005

Hu, S., Zhao, W., and Wang, T. (2015). Genetic polymorphisms of 19 STR loci in Gansu Hui. *Chin. J. Forensic Med.* 30, 620–621. doi: 10.13618/j.issn.1001-5728.2015.06.021

Huang, L., Chen, X., Xiang, C., Su, S., Fan, Y., and Xu, B. (2016). Genetic Polymorphisms of 19 STR Loci in Yunnan Han population. *Chin. J. Forensic Med.* 31 (02), 187–191. doi: 10.13618/j.issn.1001-5728.2016.02.023

Jeong, C., Ozga, A. T., Witonsky, D. B., Malmstrom, 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, X., Wei, Y., Chen, J., Kong, T., Mu, Y., Guo, Y., et al. (2017). Phylogenic analysis and forensic genetic characterization of Chinese Uyghur group *via* autosomal multi STR markers. *Oncotarget* 8, 73837–73845. doi: 10.18632/oncotarget.17992

Kalinowski, S. T. (2002). Evolutionary and statistical properties of three genetic distances. *Mol. Ecol.* 11, 1263–1273. doi: 10.1046/j.1365-294X.2002.01520.x

Kovach, W. L. (2007). *MVSP-A MultiVariate Statistical Package for Windows, ver. 3.1.* (Wales, U.K: Kovach Computing Services, Pentraeth).

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

Li, C., Li, H., Cui, Y., Xie, C., Cai, D., Li, W., et al. (2010). Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biol.* 8, 15. doi: 10.1186/1741-7007-8-15

Li, C., Ning, C., Hagelberg, E., Li, H., Zhao, Y., Li, W., et al. (2015). Analysis of ancient human mitochondrial DNA from the Xiaohe cemetery: insights into prehistoric population movements in the Tarim Basin, China. *BMC Genet.* 16, 78. doi: 10.1186/s12863-015-0237-5

Li, H., Zhao, X., Zhao, Y., Li, C., Si, D., Zhou, H., et al. (2011). Genetic characteristics and migration history of a bronze culture population in the West Liao-River valley revealed by ancient DNA. *J. Hum. Genet.* 56, 815–822. doi: 10.1038/jhg.2011.102

Li, J., Zeng, W., Zhang, Y., Ko, A. M., Li, C., Zhu, H., et al. (2017). Ancient DNA reveals genetic connections between early Di-Qiang and Han Chinese. *BMC Evol. Biol.* 17, 239. doi: 10.1186/s12862-017-1082-0

Li, J., Zhang, Y., Zhao, Y., Chen, Y., Ochir, A., Sarenbilige, et al. (2018). The genome of an ancient Rouran individual reveals an important paternal lineage in the Donghu population. *Am. J. Phys. Anthropol.* 166, 895–905. doi: 10.1002/ajpa.23491

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100–1104. doi: 10.1126/science.1153717

Li, Y. (2015). Allele frequency of 19 autosomal STR loci in the Bai population from the southwestern region of mainland China. *Electrophoresis* 36, 2498–2503. doi: 10.1002/elps.201500129

Lipson, M., Cheronet, O., Mallick, S., Rohland, N., Oxenham, M., Pietrusewsky, 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, J., Guo, L., Qi, R., Li, S.-Y., Yin, J.-Y., Zhang, W., et al. (2013). Allele frequencies of 19 autosomal STR loci in Manchu population of China with phylogenetic structure among worldwide populations. *Gene* 529, 282–287. doi: 10.1016/j.gene.2013.07.033

Liu, J., Wang, Z., He, G., Wang, M., and Hou, Y. (2019). Genetic polymorphism and phylogenetic differentiation of the Huaxia Platinum System in three Chinese minority ethnicities. *Sci. Rep.* 9, 3371. doi: 10.1038/s41598-019-39794-y

Liu, Q. (2017). Population data and mutation rates of 19 STR loci in seven provinces from China based on Goldeneye DNA ID System 20A. *Int. J. Legal Med.* 131, 653–656. doi: 10.1007/s00414-016-1441-1

Liu, Y., Guo, L., Jin, H., Li, Z., Bai, R., Shi, M., et al. (2017). Developmental validation of a 6-dye typing system with 27 loci and application in Han population of China. *Sci. Rep.* 7, 4706. doi: 10.1038/s41598-017-04548-1

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

Lu, Y., Song, P., Huang, J., and Wu, X. (2017). Genetic polymorphisms of 20 autosomal STR loci in 5141 individuals from the Han population of Xiamen. *Southeast China. Forensic Sci. Int. Genet.* 29, e31–e32. doi: 10.1016/j.fsigen.2017.03.023

Meng, H. (2016). Autosomal-STR based genetic structure of Chinese Xibe ethnic group and its relationships to various groups. *Int. J. Legal Med.* 130, 1501–1503. doi: 10.1007/s00414-016-1369-5

Moyses, C. B., Tsutsumida, W. M., Raimann, P. E., Da Motta, C. H., Nogueira, T. L., Dos Santos, O. C., et al. (2017). Population data of the 21 autosomal STRs included in the GlobalFiler((R)) kits in population samples from five Brazilian regions. *Forensic Sci. Int. Genet.* 26, e28–e30. doi: 10.1016/j.fsigen.2016.10.017

Nei, M. (1978). The theory of genetic distance and evolution of human races. *Jinrui Idengaku Zasshi* 23, 341–369. doi: 10.1007/BF01908190

Ng, J., Oldt, R. F., Mcculloh, K. L., Weise, J. A., Viray, J., Budowle, B., et al. (2016). Native American population data based on the Globalfiler((R)) autosomal STR loci. *Forensic Sci. Int. Genet.* 24, e12–e13. doi: 10.1016/j.fsigen.2016.06.014

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., and Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature* 541, 302–310. doi: 10.1038/nature21347

Nothnagel, M., Fan, G., Guo, F., He, Y., Hou, Y., Hu, S., et al. (2017). Revisiting the male genetic landscape of China: a multi-center study of almost 38,000 Y-STR haplotypes. *Hum. Genet.* 136, 485–497. doi: 10.1007/s00439-017-1759-x

Ossowski, A., Diepenbroek, M., Szargut, M., Zielinska, G., Jedrzejczyk, M., Berent, J., et al. (2017). Population analysis and forensic evaluation of 21 autosomal loci included in GlobalFiler PCR Kit in Poland. *Forensic Sci. Int. Genet.* 29, e38–e39. doi: 10.1016/j.fsigen.2017.05.003

Park, H. C., Kim, K., Nam, Y., Park, J., Lee, J., Lee, H., et al. (2016). Population genetic study for 24 STR loci and Y indel (GlobalFiler PCR Amplification kit and PowerPlex(R) Fusion system) in 1000 Korean individuals. *Leg. Med. (Tokyo)* 21, 53–57. doi: 10.1016/j.legalmed.2016.06.003

Park, J. H., Hong, S. B., Kim, J. Y., Chong, Y., Han, S., Jeon, C. H., et al. (2013). Genetic variation of 23 autosomal STR loci in Korean population. *Forensic Sci. Int. Genet.* 7, e76–e77. doi: 10.1016/j.fsigen.2012.10.005

Parolin, M. L., Real, L. E., Martinazzo, L. B., and Basso, N. G. (2015). Population genetic analyses of the Powerplex((R)) Fusion kit in a cosmopolitan sample of Chubut Province (Patagonia Argentina). *Forensic Sci. Int. Genet.* 19, 221–222. doi: 10.1016/j.fsigen.2015.07.020

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

Piao, S., Ciais, P., Huang, Y., Shen, Z., Peng, S., Li, J., et al. (2010). The impacts of climate change on water resources and agriculture in China. *Nature* 467, 43–51. doi: 10.1038/nature09364

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

Ramos-Gonzalez, B., Aguilar-Velazquez, J. A., Chavez-Briones Mde, L., Delgado-Chavarria, J. R., Alfaro-Lopez, E., and Rangel-Villalobos, H. (2016). Population data of 24 STRs in Mexican-Mestizo population from Monterrey, Nuevo Leon (Northeast, Mexico) based on Powerplex((R)) Fusion and GlobalFiler((R)) kits. *Forensic Sci. Int. Genet.* 21, e15–e17. doi: 10.1016/j.fsigen.2015.12.004

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

Reynolds, J., Weir, B. S., and Cockerham, C. C. (1983). Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105, 767–779.

Ristow, P. G., Cloete, K. W., and D'amato, M. E. (2016). GlobalFiler((R)) Express DNA amplification kit in South Africa: extracting the past from the present. *Forensic Sci. Int. Genet.* 24, 194–201. doi: 10.1016/j.fsigen.2016.07.007

Ruan, X., Wang, W. N., Yang, Y. R., Xie, B. B., Chen, J., Liu, Y. C., et al. (2015). Genetic variability and phylogenetic analysis of 39 short tandem repeat loci in Beijing Han population. *Yi Chuan Hereditas* 37 (7), 683–691. doi: 10.16288/j.yczz.15-025

Sanchez-Mazas, A., Di, D., and Riccio, M. E. (2011). A genetic focus on the peopling history of East Asia: critical views. *Rice* 4, 159–169. doi: 10.1007/s12284-011-9066-y

Schlotterer, C., and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20, 211–215. doi: 10.1093/nar/20.2.211

Shen, H., Yu, S., Guo, F., Yu, J., Jin, P., Sun, Z., et al. (2013). Genetic polymorphisms of 19 STR loci in Hui population in Liaoning province. *Chin. J. Forensic Med.* 28 (06), 504–505. doi: 10.13618/j.issn.1001-5728.2013.06.018

Shen, J., Kang, H., Dong, F., Guo, J., and Wang, R. (2015). Research of genetic polymorphism on 19 STR loci in Han nationality in Shanxi province and forensic medicine application. *J. Shanxi Police Acad.* 23, 63–66.

Singh, M., and Nandineni, M. R. (2017). Population genetic analyses and evaluation of 22 autosomal STRs in Indian populations. *Int. J. Legal Med.* 131, 971–973. doi: 10.1007/s00414-016-1525-y

Siska, V., Jones, E. R., Jeon, S., Bhak, Y., Kim, H. M., Cho, Y. S., et al. (2017). Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci. Adv.* 3, e1601877. doi: 10.1126/sciadv.1601877

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

Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J., et al. (1999). Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.* 65, 1718–1724. doi: 10.1086/302680

Taylor, D., Bright, J. A., Mcgovern, C., Neville, S., and Grover, D. (2017). Allele frequency database for GlobalFiler STR loci in Australian and New Zealand populations. *Forensic Sci. Int. Genet.* 28, e38–e40. doi: 10.1016/j.fsigen.2017.02.012

Vullo, C., Rotondo, M., Daguerre, N., Quiroz, A., Velazquez, V., Machado, P., et al. (2016). Genetic polymorphism of 22 autosomal STR markers in Paraguay. *Forensic Sci. Int. Genet.* 25, e16–e17. doi: 10.1016/j.fsigen.2016.08.002

Wang, H. (2014). Genetic polymorphisms of 20 short tandem repeat loci from the Han population in Henan, China. *Electrophoresis* 35, 1509–1514. doi: 10.1002/elps.201300597

Wang, H., Chen, L., Ge, B., Zhang, Y., Zhu, H., and Zhou, H. (2012). Genetic data suggests that the Jinggouzi people are associated with the Donghu, an ancient nomadic group of North China. *Hum Biol* 84, 365–378. doi: 10.3378/027.084.0402

Wang, J., Zhang, Q., Zhai, Z., Wang, H., Zhang, J., and Wang, Y. (2015). Genetic Polymorphisms of 19 STR loci in the Inner Mongolia Han population. *Chin. J. Forensic Med.* 30, 82–84. doi: 10.13618/j.issn.1001-5728.2015.01.026

Wang, M., Wang, Z., He, G., Jia, Z., Liu, J., and Hou, Y. (2018a). Genetic characteristics and phylogenetic analysis of three Chinese ethnic groups using the Huaxia Platinum System. *Sci. Rep.* 8, 2429. doi: 10.1038/s41598-018-20871-7

Wang, Y., Lin, B., Chen, X., Cai, J., and Zhang, Q. (2014). Genetic polymorphisms of 20 loci in Han population from Southern Fujian. *Chin. J. Forensic Med.* 29, 264–266. doi: 10.13618/j.cnki.1001-5728.2014.03.020

Wang, Y., Lu, D., Chung, Y. J., and Xu, S. (2018b). Genetic structure, divergence and admixture of Han Chinese, Japanese and Korean populations. *Hereditas* 155, 19. doi: 10.1186/s41065-018-0057-5

Wang, Z., Zhou, D., Jia, Z., Li, L., Wu, W., Li, C., et al. (2016). Developmental validation of the Huaxia platinum system and application in 3 main ethnic groups of China. *Sci. Rep.* 6, 31075. doi: 10.1038/srep31075

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

Westen, A. A., Haned, H., Grol, L. J., Harteveld, J., Van Der Gaag, K. J., De Knijff, P., et al. (2012). Combining results of forensic STR kits: HDplex validation including allelic association and linkage testing with NGM and Identifiler loci. *Int. J. Legal Med.* 126, 781–789. doi: 10.1007/s00414-012-0724-4

Wu, F., Chen, M., Chao, C., and Pu, C. (2015). *Study on STR markers of DNA for improving the power of identification on cases in Taiwan.* International Carnahan Conference on Security Technology (ICCST). IEEE, 403–408. doi: 10.1109/CCST.2015.7389718

Wu, L., Pei, B., Ran, P., and Song, X. (2017). Population genetic analysis of Xiamen Han population on 21 short tandem repeat loci. *Leg. Med. (Tokyo)* 26, 41–44. doi: 10.1016/j.legalmed.2017.03.002

Xiang, C., Huang, L., Su, S., Fan, Y., Yu, J., Huang, Y., and Xu, B. (2016). Genetic variation of the 20 STRs in Han population from Yunnan-Kweichow plateau of China. *Forensic Sci. Int. Genet.* 23, e10–e11. doi: 10.1016/j.fsigen.2016.03.003

Xiao, C., Zhang, W., Wei, T., Pan, C., and Huang, D. (2016). Population data of 21 autosomal STR loci in Chinese Han population from Hubei province in Central China. *Forensic Sci. Int. Genet.* 20, e13–e14. doi: 10.1016/j.fsigen.2015.11.002

Xie, J., Shao, C., Zhou, Y., Zhu, W., Xu, H., Liu, Z., et al. (2014). Genetic distribution on 20 STR loci from the Han population in Shanghai, China. *Forensic Sci. Int. Genet.* 9, e30–e31. doi: 10.1016/j.fsigen.2013.08.007

Xu, S., Yin, X., Li, S., Jin, W., Lou, H., Yang, L., et al. (2009). Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am. J. Hum. Genet.* 85, 762–774. doi: 10.1016/j.ajhg.2009.10.015

Yang, M., Ren, Z., Ji, J., Zhou, H., Zhang, H., Dai, J., et al. (2017a). Population genetic data and mutations of 22 autosomal STR loci in Guizhou Han population. *Forensic Sci. Int. Genet.* 29, e29–e30. doi: 10.1016/j.fsigen.2017.03.008

Yang, M. A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., et al. (2017b). 40,000-Year-old individual from asia provides insight into early population structure in eurasia. *Curr. Biol.* 273202-3208, e3209. doi: 10.1016/j.cub.2017.09.030

Yao, J., Xiong, K., and Zhang, Y. (2017). Population data of 19 autosomal STR loci in the Chinese Han population from Jiujiang. *Southern China. Forensic Sci. Int. Genet.* 28, e47–e48. doi: 10.1016/j.fsigen.2017.03.007

Yao, Y. G., Kong, Q. P., Bandelt, H. J., Kivisild, T., and Zhang, Y. P. (2002). Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am. J. Hum. Genet.* 70, 635–651. doi: 10.1086/338999

Yin, C., Ji, Q., Li, K., Mu, H., Zhu, B., Yan, J., et al. (2015). Analysis of 19 STR loci reveals genetic characteristic of eastern Chinese Han population. *Forensic Sci. Int. Genet.* 14, 108–109. doi: 10.1016/j.fsigen.2014.09.015

Zhang, H., Xia, M., Qi, L., Dong, L., Song, S., Ma, T., et al. (2016a). Forensic and population genetic analysis of Xinjiang Uyghur population on 21 short tandem repeat loci of 6-dye GlobalFiler™ PCR Amplification kit. *Forensic Sci. Int. Genet.* 22, 22–24. doi: 10.1016/j.fsigen.2016.01.005

Zhang, H., Yang, S., Guo, W., Ren, B., Pu, L., Ma, T., et al. (2016b). Population genetic analysis of the GlobalFiler STR loci in 748 individuals from the Kazakh population of Xinjiang in northwest China. *Int. J. Legal Med.* 130, 1187–1189. doi: 10.1007/s00414-016-1319-2

Zhang, L., Xu, C., Chen, H., Zhou, Y., Li, F., and Chen, J. (2012). Genetic polymorphism of 19 STR loci in Xinjiang Barkol Kazakh population. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* 37 (9), 934–938. doi: 10.3969/j.issn.1672-7347.2012.09.014

Zhang, M. X., Han, S. Y., Gao, H. M., Sun, S. H., Xiao, D. J., Liu, Y., et al. (2013). Genetic polymorphisms of 19 STR loci in Shandong Han population. *J. Forensic Sci.* 29 (6), 440–443. doi: 10.3969/j.issn.1004-5619.2013.06.010

Zhang, X., Hu, L., Du, L., Nie, A., Rao, M., bo Pang, J., et al. (2017a). Genetic analysis of 20 autosomal STR loci in the Miao ethnic group from Yunnan

Province. Southwest China. *Forensic Sci. Int. Genet.* 28, e28–e29. doi: 10.1016/j.fsigen.2017.02.003

Zhang, X., Hu, L., Du, L., Nie, A., Rao, M., bo Pang, J., et al. (2017b). Genetic polymorphisms of 20 autosomal STR loci in the Vietnamese population from Yunnan Province, Southwest China. *Int. J. Legal Med.* 131 (3), 661–662. doi: 10.1007/s00414-016-1496-z

Zhang, X., Hu, L., Du, L., Zheng, H., Nie, A., Rao, M., et al. (2017c). Population data for 20 autosomal STR loci in the Yi ethnic minority from Yunnan Province Southwest China. *Forensic Sci. Int. Genet.* 28, e43–e44. doi: 10.1016/j.fsigen.2017.02.017

Zhang, X., Du, L., Huang, Z., Gu, T., Hu, L., and Nie, S. (2017a). Genetic variation of 20 autosomal STR loci in three ethnic groups (Zhuang, Dai and Hani) in the Yunnan province of Southwestern China. *Forensic Sci. Int. Genet.* 31, e41–e42. doi: 10.1016/j.fsigen.2017.06.005

Zhang, Y., Li, A., Zhao, L., Liu, J., and Song, J. (2011). Genetic polymorphisms of 19 autosomal short tandem repeats in the Hebei Han population. *Foren. Sci. Technol.* 2, 48–50. doi: 10.3969/j.issn.1008-3650.2011.02.023

Zhang, Y., Li, J., Zhao, Y., Wu, X., Li, H., Yao, L., et al. (2017b). Genetic diversity of two Neolithic populations provides evidence of farming expansions in North China. *J. Hum. Genet.* 62, 199–204. doi: 10.1038/jhg.2016.107

Zhang, Z., and Chen, T. (2015). Genetic polymorphisms of 20 STR loci in Han population from Western Fujian. *Chin. J. Forensic Med.* 30, 411–412. doi: 10.13618/j.issn.1001-5728.2015.01.025

Zhao, L., Wu, S., Gu, F., Zhang, Y., Wang, X., and Li, A. (2016). Genetic Polymorphisms of 19 STR Loci in Manchu population in Hebei Province. *Chin. J. Forensic Med.* 31 (01), 72–73. doi: 10.13618/j.issn.1001-5728.2016.01.020

Zhao, W., Hu, S., and Wang, T. (2015a). Genetic polymorphisms of 19 STR loci of Pingliang Han of Gansu province. *Chin. J. Forensic Med.* 30, 618–619. doi: CNKI:SUN:FUAN.0.2015-06-022

Zhao, Y. B., Li, H. J., Li, S. N., Yu, C. C., Gao, S. Z., Xu, Z., et al. (2011). Ancient DNA evidence supports the contribution of Di-Qiang people to the han Chinese gene pool. *Am. J. Phys. Anthropol.* 144, 258–268. doi: 10.1002/ajpa.21399

Zhao, Y. B., Zhang, Y., Li, H. J., Cui, Y. Q., Zhu, H., and Zhou, H. (2014). Ancient DNA evidence reveals that the Y chromosome haplogroup Q1a1 admixed into the Han Chinese 3,000 years ago. *Am. J. Hum. Biol.* 26, 813–821. doi: 10.1002/ajhb.22604

Zhao, Y. B., Zhang, Y., Zhang, Q. C., Li, H. J., Cui, Y. Q., Xu, Z., et al. (2015b). Ancient DNA reveals that the genetic structure of the northern Han Chinese was shaped prior to 3,000 years ago. *PLoS One* 10, e0125676. doi: 10.1371/journal.pone.0125676

Zou, X., Li, Y., Li, P., Nie, Q., Wang, T., Hu, Y., et al. (2017). Genetic polymorphisms for 19 autosomal STR loci of Chongqing Han ethnicity and phylogenetic structure exploration among 28 Chinese populations. *Int. J. Legal Med.* 131, 1539–1542. doi: 10.1007/s00414-017-1577-7