



Population Genetic Diversity and Phylogenetic Characteristics for High-Altitude Adaptive Kham Tibetan Revealed by DNATyper™ 19 Amplification System

Xing Zou^{1†}, Zheng Wang^{1†}, Guanglin He^{1†}, Mengge Wang¹, Yongdong Su², Jing Liu¹, Pengyu Chen^{3,4}, Shouyu Wang¹, Bo Gao⁵, Zhao Li⁶ and Yiping Hou^{1*}

¹ Institute of Forensic Medicine, West China School of Basic Medical Sciences and Forensic Medicine, Sichuan University, Chengdu, China, ² Forensic Identification Center, Public Security Bureau of Tibet Autonomous Region, Lhasa, China, ³ Center of Forensic Expertise, Affiliated Hospital of Zunyi Medical University, Zunyi, China, ⁴ School of Forensic Medicine, Zunyi Medical University, Zunyi, China, ⁵ Institute of Forensic Science, Yili Public Security Bureau of Xinjiang, Kuytun, China, ⁶ Department of Criminal Investigation, Mianyang Public Security Bureau, Mianyang, China

OPEN ACCESS

Edited by:

Fulvio Cruciani,
Sapienza University of Rome, Italy

Reviewed by:

Antonio González-Martín,
Complutense University of Madrid,
Spain
Chuanchao Wang,
Xiamen University, China

*Correspondence:

Yiping Hou
forensic@scu.edu.cn

[†] These authors have contributed
equally to this work as co-first authors

Specialty section:

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

Received: 21 September 2018

Accepted: 26 November 2018

Published: 17 December 2018

Citation:

Zou X, Wang Z, He G, Wang M,
Su Y, Liu J, Chen P, Wang S, Gao B,
Li Z and Hou Y (2018) Population
Genetic Diversity and Phylogenetic
Characteristics for High-Altitude
Adaptive Kham Tibetan Revealed by
DNATyper™ 19 Amplification System.
Front. Genet. 9:630.
doi: 10.3389/fgene.2018.00630

Tibetans residing in the high-altitude inhospitable environment have undergone significant natural selection of their genetic architecture. Recently, highly mutational autosomal short tandem repeats were widely used not only in the anthropology and population genetics to investigate the genetic structure and relationships, but also in the medical genetics to explore the pathogenesis of multiple genetic diseases and in the forensic science to identify individual and parentage relatedness. However, genetic variants and forensic efficiency of DNATyper™ 19 amplification system and genetic background of Kham Tibetan remain uncharacterized. Thus, we genotyped 19 forensic genetic markers in 11,402 Kham Tibetans to gain insight into the genetic diversity of Chinese high-altitude adaptive population. Highly discriminating and polymorphic forensic measures were observed, which indicated that this new-developed DNATyper 19 PCR amplification is suitable for routine forensic identification purposes and Chinese national DNA database establishment. Pairwise genetic distances among the comprehensive population comparisons suggested that this high-altitude adaptive Kham Tibetan has genetically closer relationships with lowlanders of Tibeto-Burman-speaking populations (Chengdu Tibetan, Liangshan Tibetan, and Liangshan Yi). Genetic substructure analyses via phylogenetic reconstruction, principal component analysis, and multidimensional scaling analysis in both nationwide and worldwide contexts suggested that the genetic proximity exists along the linguistic, ethnic, and continental geographical boundary. Further studies with whole-genome sequencing of modern or archaic Kham Tibetans would be useful in reconstructing the Tibetan population history.

Keywords: Tibetan, genetic polymorphism, short tandem repeat, population relationship, forensic genetics

INTRODUCTION

Short tandem repeats (STRs), also referred to as microsatellites, are mainly scattered in the non-coding regions of the whole human genome (Gymrek et al., 2017; Willems et al., 2017). This most variable genetic marker in the eukaryotic genomes comprises tandem repeat motif of 2–6 base pairs. The *de novo* mutation rate of STRs is larger several orders of magnitude (approximately

10^{-3} – 10^{-4}) when compared with the binary genetic markers (approximately 10^{-8} – 10^{-9}) of single nucleotide polymorphisms (SNPs) and Insertion/Deletion (InDel) (Willems et al., 2014). STR mutations are generally generated through the molecular mechanism of replication slippage and stepwise mutation model, which can add or subtract one repeat unit (such as the motif of TATC in the D13S317 locus). With the advent of Polymerase Chain Reaction (PCR) in the late 1980s and subsequently tremendous progresses of capillary electrophoresis (CE) and whole-genome sequencing, STRs are broadly used in the disease pathogenesis, genetic diversity, population differentiation, and forensic identification (Willems et al., 2014; Gymrek, 2017). Human population genetic scientists hold the opinions that a large number of factors, such as inbreeding and geographical isolation, migration, gene flow, genetic admixture and population fragmentation, contribute to the genetic diversity of the human genome (Kayser and de Knijff, 2011; Sun et al., 2012). Tishkoff et al. (2009) used 848 microsatellites in over 2,500 individuals to characterize the genetic diversity and dissect the population structure across linguistically, geographically and ethnically diverse African populations, as well as reconstruct the complex human evolutionary history.

In forensic science, multiplex STRs genotyping by the fluorescent labeled PCR amplification combined with the CE approach is recognized as the current gold standard in the personal identification, kinship testing and missing person's identification (Kayser and de Knijff, 2011). Since second generation multiplex (SGM) including six STRs was used in the establishment of the National DNA Database by the Forensic Science Service (FSS) in England in 1996 (Werrett, 1997), a variety of commercial kits containing 15–25 loci selected from the combined DNA index system (CODIS), expanded CODIS, UK core loci (UCL), German core loci (GCL) and Australia national DNA database (NCIDD), International Criminal Police Organization (INTERPOL) standard set of loci (ISSL) and extended European standard set (ESS-extended) were subsequently developed, validated and applied in the forensic cases (Gill et al., 2006; Hares, 2012). GlobalFiler Express PCR Amplification Kit and Huaxia Platinum System (Thermo Fisher Scientific) are typical systems to increase discrimination power, improve international compatibility and reduce the likelihood of adventitious matches (Wang et al., 2015; He et al., 2018b,e). More recently, a new PCR amplification system, DNATyper™ 19 kit, was developed and validated by the Institute of Forensic Science in the Ministry of Public Security (Beijing, China), which can co-amplify 18 autosomal STRs and one sex-determination marker of Amelogenin focused on Chinese populations.

The Tibetan Plateau is generally considered to have been covered by the ice sheet during the last glacial maximum. Until recently, there has been no consensus view about when colonization began, how Tibetans got there and how they occupied and adapted this cold, arid, hypobaric, and hypoxic environment. Archeological evidence from the Heimaha, Jiangxigou sites suggested that the gradual expansion of foragers' occupation of Tibet began from 40–25 thousand years ago (kya) (Madsen et al., 2006). Abundant evidence from

genetic perspectives documented and reconstructed the Tibetan population history and high-altitude adaptation evolutionary history (Zhao et al., 2009; Qi et al., 2013; Huerta-Sanchez et al., 2014). Zhao et al. (2009) suggested that the matrilineal genetic relics and genetic continuity exist between the Late Paleolithic Tibet inhabitants and modern Tibetans. Genetic analyses of simultaneous testing paternal Y chromosome, maternal mitochondrial DNA and autosomal variations documented the upper Paleolithic occupation and at least one Neolithic expansion (Qi et al., 2013). Additionally, many whole-genome genetic studies have identified that the genetic basis of variations in EPAS1 and EGLN seems to be involved in high altitude adaptation of Tibetans and the corresponding adaptive haplotypes (AGGAA) in the EPAS1 gene are obtained by introgression from Denisovan archaic hominin (Huerta-Sanchez et al., 2014). There are also too many other genetic, linguistic and archeological studies which isolated or combined to reconstruct the complex genetic origin, admixture, divergence with the surrounding populations (Lu et al., 2016; Hu et al., 2017; Zhang et al., 2017; He et al., 2018a,c,f). However, existing genetic data are not sufficient to explore the genetic variations and features of the forensic related markers of Tibetans with different origins and cultural background (Ü-Tsang, Kham and Ando Tibetans).

Thus, we conducted and reported the first large-scale autosomal STRs study in this unique high-altitude adaptive Tibetan population based on a new-generation DNATyper™ 19 PCR amplification system and explored the detailed genetic variants, genetic diversity and forensic efficiency of STRs in the Kham Tibetans in this study. Furthermore, we performed two comprehensive population comparisons (nationwide population relationship investigation among 64 groups and worldwide genetic affinity exploration among 53 groups) to dissect the genetic differentiation between the Kham Tibetans and reference populations and simultaneously provide some new insights for patterns of global or local population substructure based on autosomal genetic variability.

MATERIALS AND METHODS

DNA Sample Collections and Ethics Statements

This project and corresponding protocol were considered and approved by the Ethics Committee of the Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University (Approval Number: K2015008). Our participants are needed to be the indigenous Tibetans and no intermarriage or long-distance migration at least three generations. Our subjects have signed written informed consent and analyzed anonymously. A total of 11,402 unrelated healthy individuals (4,846 females and 6,556 males) were collected from the east of Tibet Kham Tibetan autonomous region (Chengdu country), Aba and Muli city in Sichuan province. To insure the included the donors which meet the aforesaid requirement, we followed the following criteria: (1) both parents and grandparents being Tibetans; (2) the language used first is Tibeto-Burman language; (3) all participants residing in the same village or

owning the same family names are need to check with relative relationships with previous included subjects to avoid included close relatives; (4) in the past three generations, there is no documented ancestors from other ethnic groups. Besides, to avoid the potential included close relatives, we employed a large sample size to dilute the sample collection bias. Blood samples are collected using FTA cards or cotton swab. All datasets generated and analyzed for this study are included in the **Supplementary Material**.

DNA Amplification

Nineteen forensic genetic markers labeled with multi-fluorescent dyes (vWA, TPOX, TH01, Penta E, FGA, D8S1179, D7S820, D6S1043, D5S818, D3S1358, D2S1338, D21S11, D19S433, D18S51, D16S539, D13S317, D12S391, CSF1PO and Amelogenin) were amplified simultaneously using the DNATyper™ 19 PCR amplification system on a GeneAmp PCR System 9700 Thermal Cycler (Applied Biosystems, Foster City, CA, United States) on the basis of the manufacturer's instruction. We employed the following PCR amplification conditions: decomposition at 72°C for 20 min and denaturation at 95°C for 11 min, and then amplification for 26 cycles of denaturation for 30 s at 94°C, anneal for 2 min at 59°C and extension for 1 min at 72°C, following a final extension at 60°C for 60 min, and holding at 25°C. PCR products are mixed with the deionized Formamide and Typer500, and then isolated using the capillary electrophoresis on an ABI 3500 XL Genetic Analyzer (Applied Biosystems, Foster City, CA, United States). Electrophoresis results were visualized and checked using the GeneMapper ID-X Software v1.5 (Applied Biosystems, Foster City, CA, United States).

Statistical Analysis

The exact tests using a Markov chain of linkage disequilibrium and Hardy–Weinberg equilibrium among 18 forensic autosomal genetic markers, as well as estimation of the observed heterozygosity (H_o) and expected heterozygosity (H_e), were carried out using the Arlequin version 3.5.2.2¹ (Excoffier and Lischer, 2010). Online tool of the STRAF (STR Analysis for Forensics) (Gouy and Zieger, 2017) was used to calculate the allelic frequencies and statistical parameters of forensic interest, which included the power of exclusion (PE), probability of matching (PM), polymorphism information content (PIC), and power of discrimination (PD). Population genetic differentiation analyses were conducted in two distinctive reference population panels: nationwide panel and worldwide panel. Pairwise Reynolds genetic distances between the Kham Tibetan and reference populations were calculated using the Phylogeny Inference Package (PHYLIP) version 3.6.7² (Cummings, 2004). Principal component analyses (PCA) on the basis of the allelic frequency distribution of the 18 autosomal STRs among 64 nationwide populations and the 16 autosomal STRs among 53 worldwide populations were carried out using a Multivariate Statistical

Package (MVSP) for Windows, version 3.1³ (Kovach, 2007). Multidimensional scaling (MDS) plots based on the two pairwise Reynolds genetic distance matrixes were conducted using the IBM SPSS® software⁴ (Hansen, 2005). Finally, two phylogenetic relationships were constructed using the neighbor-joining method in the Molecular Evolutionary Genetics Analysis (MEGA) Version 7.0 (Kumar et al., 2016).

Quality Control

This study was in accordance with the recommendations of scientific standards for studies in forensic genetics proposed and advocated by the International Society for Forensic Genetics (ISFG) (Schneider, 2007). The experiment was conducted in an ISO 17025 accredited laboratory, which simultaneously passed and accredited by the China National Accreditation Service for Conformity Assessment (CNAS). Laboratory internal standard and manufacturer's instruction were strictly followed to minimize errors. Negative control (H₂O) and positive control (9947A) were genotyped along with each batch of samples.

RESULTS

Hardy–Weinberg Equilibrium and Linkage Disequilibrium

A total of 11,402 Kham Tibetan subjects were successfully genotyped using the DNATyper™ 19 amplification system (**Supplementary Table S1**). As shown in **Table 1**, We observed no significant deviation from the Hardy–Weinberg equilibrium (HWE) for the 18 autosomal STRs in Chinese Kham Tibetan after applying the Bonferroni correction for multiple tests ($p < 0.05/18 = 0.0028$). Simultaneously, pairwise Linkage Disequilibrium (LD) among 153 locus pairs was conducted, and we identified 34 pairs existing linkage or associated inheritance in the Kham Tibetan (**Supplementary Table S2**). To authorize whether population stratifications exist in this Tibetan group, we first test the genetic heterogeneity or homogeneity of the Kham Tibetan via principal component analysis (PCA). As shown in **Supplementary Figure S1**, 2.23% genetic variations extracted from Kham Tibetan demonstrated that Kham Tibetan is a homogeneous population. To further validate the genetic homogeneity and initially explore genetic similarities with neighboring populations, we conducted a PCA, F_{st} genetic distance calculation and phylogenetic relationship reconstruction on the basis of raw genotype data of 18 autosomal STRs from 18,499 individuals from 12 populations. As shown in **Supplementary Table S3** and **Figure 1**, a total of 1.73% genetic variations can be extracted by the first three PCs. We identified light population stratifications among geographically and genetically different populations due to most individual plots are overlapped in the PCA analyses. But we can also observe genetic affinity among populations belongs to the same language family (Sinitic, Tibeto-Burman, and Turkic). Generally, population comparisons between the meta-Tibetan

¹<http://cmpg.unibe.ch/software/arlequin35/>

²<http://evolution.genetics.washington.edu/phylip.html>

³<https://www.kovcomp.co.uk/mvsp/index.html>

⁴<https://www.ibm.com/analytics/spss-statistics-software>

TABLE 1 | Forensic statistical parameters of 18 forensic autosomal genetic markers in 11,402 unrelated Kham Tibetans residing Tibet autonomous region.

| Locus | CSFIPO | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D21S11 | D2S1338 | D3S1358 | D5S818 | D6S1043 | D7S820 | D8S1179 | FGA | PentaE | TH01 | TPOX | vWA |
|-------|--------|---------|---------|---------|--------|---------|--------|---------|---------|--------|---------|--------|---------|--------|--------|--------|--------|--------|
| Nall | 10 | 16 | 9 | 9 | 19 | 16 | 19 | 13 | 9 | 9 | 20 | 11 | 10 | 23 | 21 | 7 | 8 | 9 |
| PIC | 0.6824 | 0.8004 | 0.7908 | 0.7355 | 0.8249 | 0.8052 | 0.8222 | 0.8369 | 0.6570 | 0.7071 | 0.8566 | 0.7560 | 0.8059 | 0.8443 | 0.9122 | 0.5751 | 0.5301 | 0.7660 |
| PM | 0.1188 | 0.0536 | 0.0589 | 0.0878 | 0.0421 | 0.0514 | 0.0439 | 0.0380 | 0.1358 | 0.1031 | 0.0301 | 0.0766 | 0.0518 | 0.0344 | 0.0125 | 0.1884 | 0.2267 | 0.0720 |
| PD | 0.8812 | 0.9464 | 0.9411 | 0.9122 | 0.9579 | 0.9486 | 0.9561 | 0.9620 | 0.8642 | 0.8969 | 0.9699 | 0.9234 | 0.9482 | 0.9656 | 0.9875 | 0.8116 | 0.7733 | 0.9280 |
| PE | 0.4661 | 0.6232 | 0.6137 | 0.5322 | 0.6635 | 0.6443 | 0.6714 | 0.6947 | 0.4323 | 0.4932 | 0.7214 | 0.5600 | 0.6483 | 0.6904 | 0.8064 | 0.3095 | 0.2677 | 0.5797 |
| TPI | 1.8104 | 2.6725 | 2.6020 | 2.1091 | 3.0116 | 2.8406 | 3.0866 | 3.3320 | 1.6812 | 1.9247 | 3.6592 | 2.2587 | 2.8749 | 3.2840 | 5.2836 | 1.2992 | 1.1909 | 2.3754 |
| Ho | 0.7238 | 0.8129 | 0.8078 | 0.7629 | 0.8340 | 0.8240 | 0.8380 | 0.8499 | 0.7026 | 0.7402 | 0.8634 | 0.7786 | 0.8261 | 0.8477 | 0.9054 | 0.6152 | 0.5802 | 0.7895 |
| He | 0.7258 | 0.8227 | 0.8166 | 0.7710 | 0.8417 | 0.8266 | 0.8401 | 0.8633 | 0.7096 | 0.7467 | 0.8698 | 0.7876 | 0.8285 | 0.8593 | 0.9181 | 0.6258 | 0.5847 | 0.7969 |
| p | 0.1753 | 0.0200 | 0.0029 | 0.2376 | 0.0031 | 0.0565 | 0.0043 | 0.0981 | 0.3540 | 0.8699 | 0.0470 | 0.0073 | 0.0631 | 0.0648 | 0.0087 | 0.0332 | 0.8125 | 0.4315 |

PM, random matching probability; PD, power of discrimination; PIC, polymorphism information content; PE, power of exclusion; TPI, typical paternity index; Ho, observed heterozygosity; He, expected heterozygosity; p, p-value in the Hardy-Weinberg.

and 11 previously investigated populations revealed that Kham Tibetan keep a close genetic relationship with other four Tibeto-Burman-speaking populations ($F_{st} = 0.0001$, **Supplementary Table S3**). Thus, we can establish one database of allele frequency distributions of Tibetan population for forensic routine applications.

Genetic Diversities and Forensic Efficiency Parameters

To explore more precise Tibetan-specific allele frequencies for likelihood estimation in the forensic parentage testing and comprehensively evaluate forensic efficiency of the DNATyper™ 19 amplification system in the forensic personal identification, we calculated the allele frequencies of 18 autosomal STRs and corresponding forensic efficiency parameters in this Kham Tibetan population. A total of 238 alleles with corresponding allelic frequencies spanning from 0.00004 to 0.58209 were observed (**Supplementary Table S4**). FGA was the locus with the most alleles (23) and TH01 had the least gene locus with 7 alleles. The expected heterozygosity, also named as genetic diversity, varied from 0.5847 at locus of TPOX to 0.9181 at locus of Penta E [Average ± Standard (SD): 0.7903 ± 0.0852]. Ho values spanned from 0.5802 (TPOX) to 0.9054 (Penta E) (Average ± SD: 0.7835 ± 0.0848). PIC varied from 0.5301 (TPOX) to 0.9122 (Penta E) (Average ± SD: 0.7616 ± 0.0992) and PM varied from 0.0125 (Penta E) to 0.2267 (TPOX) (Average ± SD: 0.0792 ± 0.0568). PD spanned from 0.7733 (TPOX) to 0.9875 (Penta E) (Average ± SD: 0.9208 ± 0.0568) and PE varied from 0.2677 (TPOX) to 0.8064 (Penta E) (Average ± SD: 0.5788 ± 0.1413). TPI spanned from 1.1909 (TPOX) to 5.2836 (Penta E) (Average ± SD: 2.6276 ± 0.9661).

Population Genetic Diversity Analysis Revealed by Pairwise Reynolds Genetic Distance

For population genetic relationship comparison, we first explored the genetic differentiation between the Kham Tibetan and other 63 Chinese nationwide populations on the basis of 18 overlapped STRs (CSFIPO, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D2S1338, D3S1358, D5S818, D6S1043, D7S820, D8S1179, FGA, Penta E, TH0, TPOX, and vWA). This reference population panel comprises of 43 geographically diverse Han Chinese populations, four Uyghur populations, two Manchu, Hui, Yi and Tibetan populations and one Kazakh, Bai, Vietnamese, Miao, Zhuang, Hani and Xibe (**Supplementary Figure S2** and **Supplementary Table S5**). The pairwise Reynolds genetic distances among 64 populations were calculated and presented in **Supplementary Table S6**. The smallest genetic distance was identified between Jiangxi Jiujiang Han and Yunnan Han (0.0001), followed by Pearl River Delta Han and Guangdong Guangzhou Han (0.0002). The largest genetic one was observed between Xinjiang Kazakh and Yunnan Miao (0.0465). Kham Tibetan had genetically closer relationships with Liangshan Tibetan (0.0035) and Yi (0.0040) and Yunnan Bai (0.0036). A heat map of this genetic matrix showed that the Yunnan Miao and Vietnamese, three

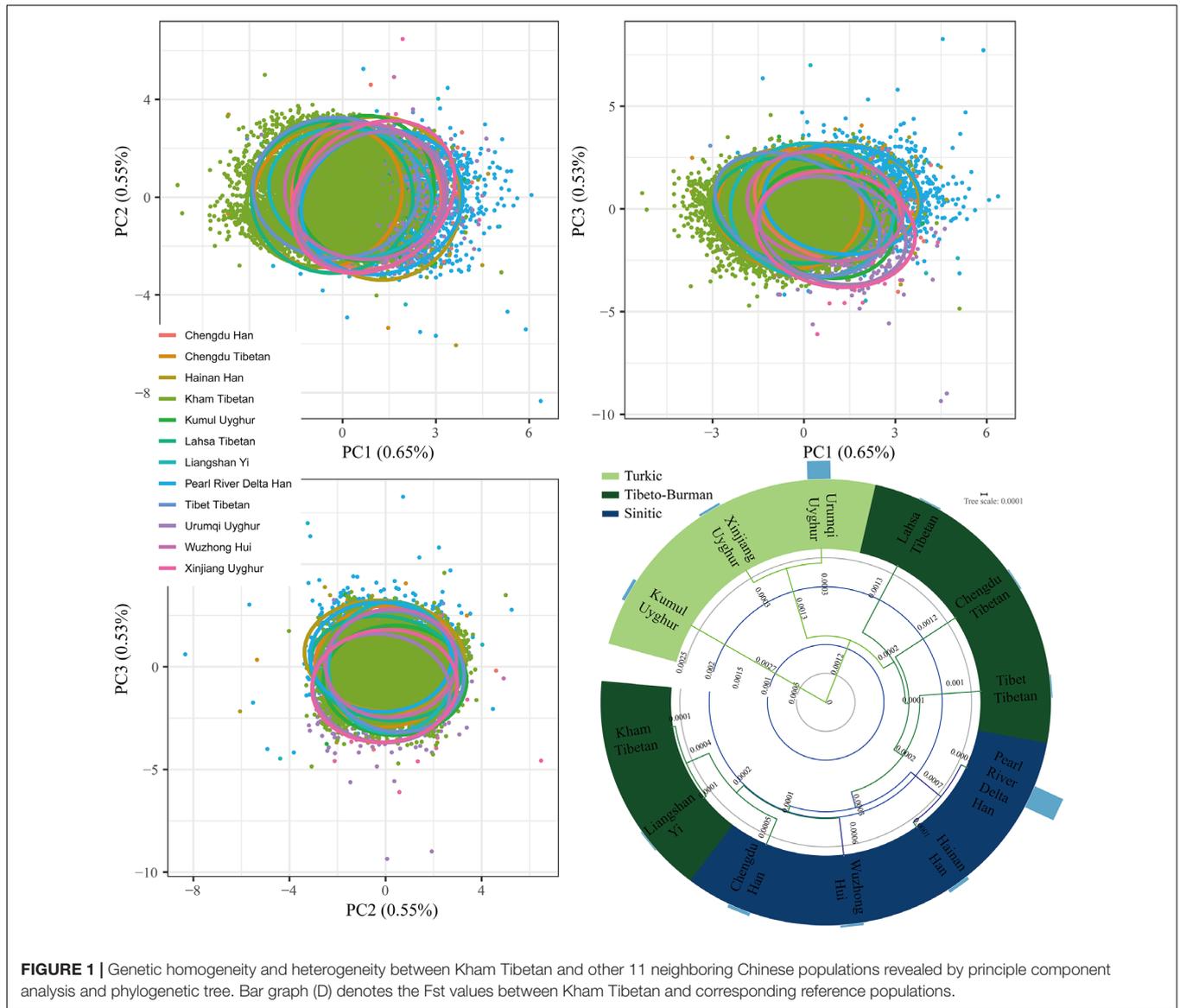


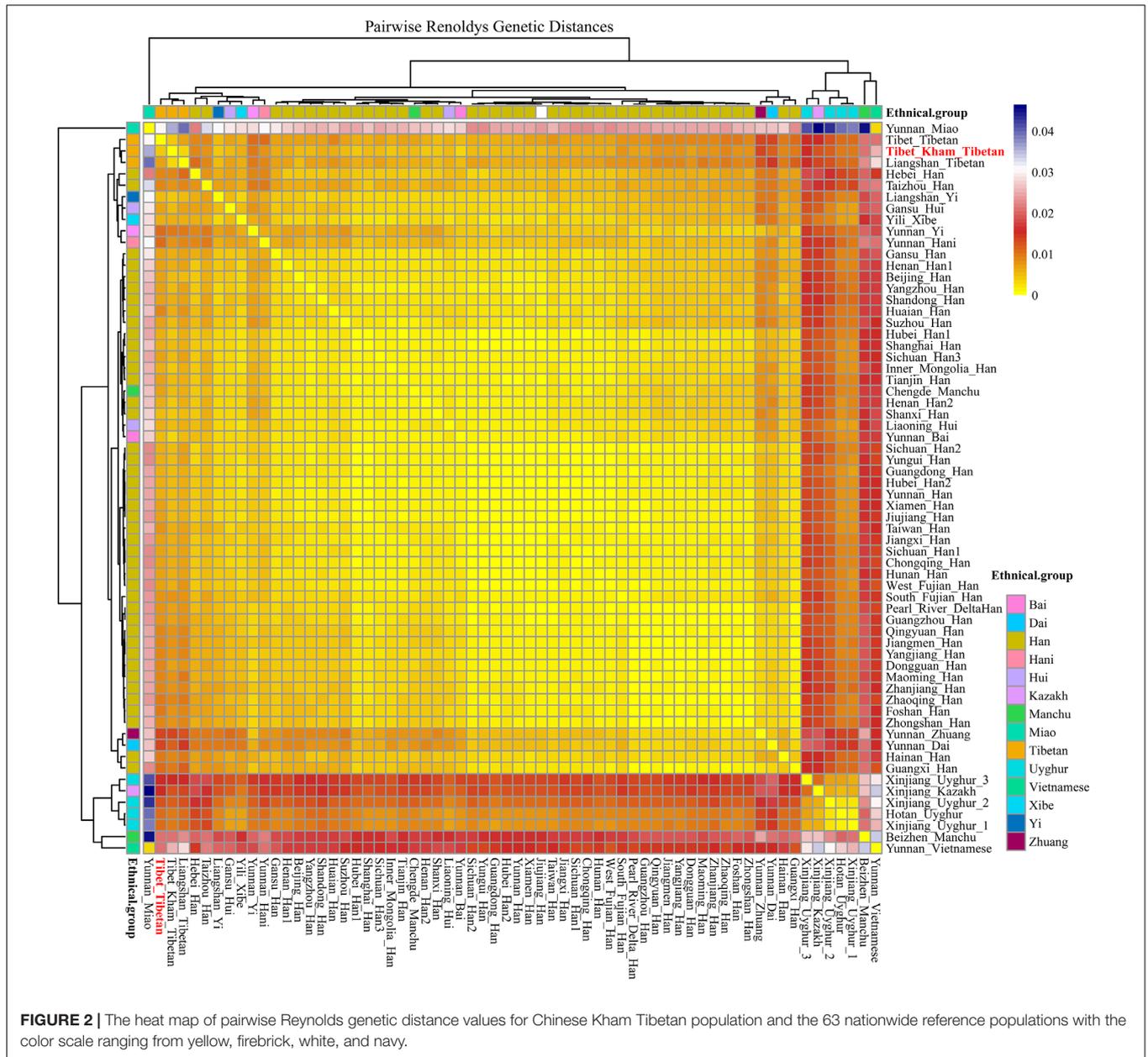
FIGURE 1 | Genetic homogeneity and heterogeneity between Kham Tibetan and other 11 neighboring Chinese populations revealed by principle component analysis and phylogenetic tree. Bar graph (D) denotes the F_{st} values between Kham Tibetan and corresponding reference populations.

Xinjiang Uyghur populations, Kazakh, Benzhen Manchu and Kham Tibetan had overall higher genetic differences compared with the others (Figure 2). To get a worldwide view of genetic similarities and differences of the Kham Tibetan, we made the other comprehensive population comparison which focused on the Kham Tibetan and 52 worldwide reference populations on the basis of 16 overlapped STRs (CSF1PO, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D2S1338, D3S1358, D5S818, D7S820, D8S1179, FGA, TH01, TPOX, and vWA). The detailed language family and geographical origins are listed in Supplementary Figure S3 and Supplementary Table S7. Kham Tibetan had a genetic affinity with the Chengdu Tibetan (0.0023) and Liangshan Tibetan (0.0035), followed by Liangshan Yi (0.0039), and owned significant genetic differences with African AmaXhosa (0.0510), AmaZulu (0.0428), and Native American (0.0402). Apparent genetic affinity within-continent populations, such as East Asians, can

be detected in the heat map (Supplementary Table S8 and Figure 3).

Principal Components Analyses Among 64 Nationwide and 53 Worldwide Populations

Principal component analyses based on the genetic data has been widely used in correcting for population stratification to avoid false negative or positive results in the genome wide association studies, making qualified ancestry inferences in the human history reconstruction and detecting population substructure (Patterson et al., 2006; Pickrell and Pritchard, 2012). We first performed PCA among 64 populations on the basis of the allelic frequency distribution (Figures 4A,B). The top 10 components could extract a total of 73.390% genetic variants (PC1: 28.372%; PC2: 14.048%; PC3: 12.879%; PC4: 8.230%;

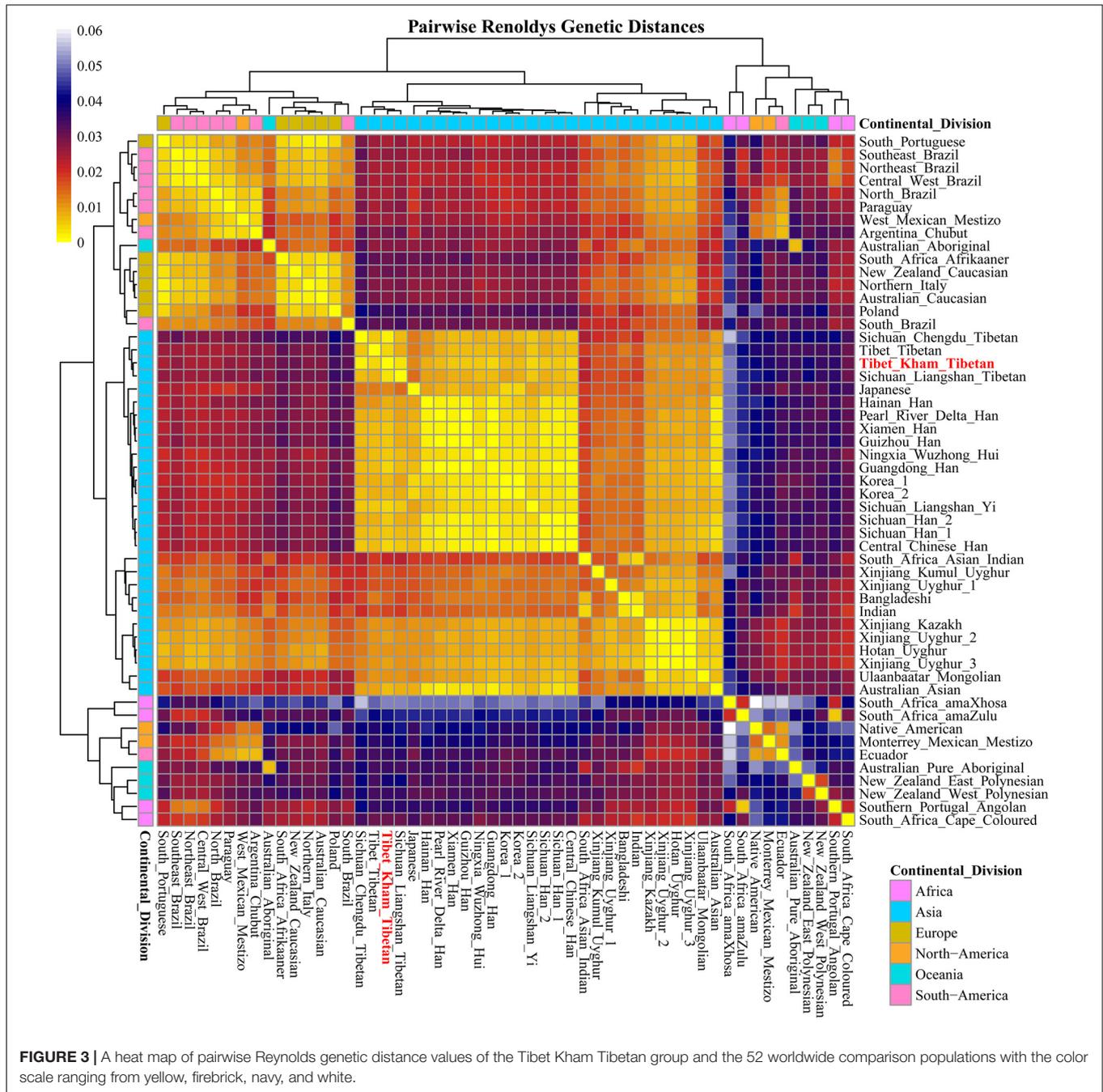


PC5: 3.892%; PC6: 2.906%; PC7: 2.697%; PC8: 2.357%; PC9: 2.123%; PC10: 1.885%). **Figure 4A** was constructed on the basis of the first two components. We observed a clear separation between Han Chinese populations and minority ethnicities with the exception of the Chengde Manchu and Yunnan Bai. PC1 separated Yunnan Miao and Vietnamese from the others, and PC2 separated five Turkic-speaking populations from others. Tibet Kham Tibetan could be distinguished by both PC1 and PC2, and located in the left corner of the first quadrant upper. We subsequently explored the population substructure among 53 worldwide populations via the genetic polymorphisms of 16 polymorphic STRs (**Figures 4C,D**). Around 84.663% genetic variants had been extracted by the first 10 components. The PC1 to PC10 were, respectively, consisted of 33,742, 15,083, 11,883,

6,524, 5,303, 3,119, 2,985, 2,164, 2,023, and 1,837%. Synthetic map based on the combination of PC1 and PC2 was presented in **Figure 4C**. East Asian populations were distinguished in the PC1 and three African origin populations and seven European populations were separated in the PC2.

Multidimensional Scaling Analyses

To further illustrate and dissect the genetic relationships between the Tibet Kham Tibetan and 63 nationwide groups, as well as 52 worldwide populations, we performed multidimensional scaling analyses using the national-scale and world-scale pairwise genetic distance matrixes. As shown in **Figure 5**, Kham Tibetan was localized close with Liangshan Tibetan and Tibet Tibetan and located alone in the fourth quadrant of the coordinate axis. Han



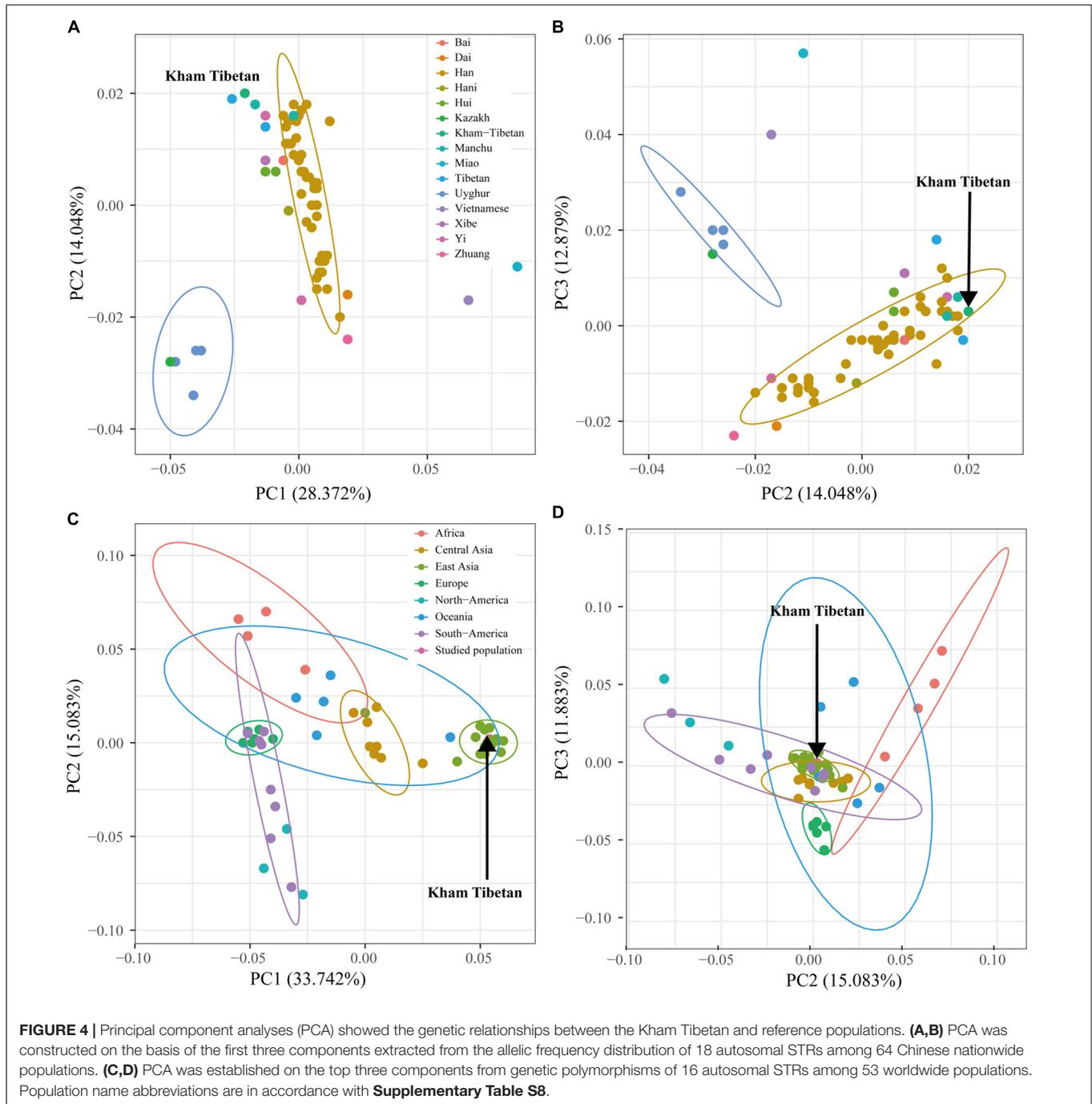
Chinese populations, except for Guangxi Han, Hebei Han and Taizhou Han, fell close to each other and were generally close to four Chinese minority ethnicities (Gansu Hui, Liaoning Hui, Yunnan Bai, and Yili Xibe). Other Chinese minorities, including five Turkic-speaking populations (Uyghur and Kazakh), five Tibetan-Burman-speaking populations (Tibetan and Yi), Miao, Zhuang, Dai and Vietnamese, formed a loose cluster and distinguished from Han Chinese populations.

We also carried out a new MDS which projected worldwide populations. It is evident that the worldwide population substructures were concordant with continental boundaries

(Africa, Europe, South Asia, Central Asia, East Asia, America, and Oceania), which is in accordance with the observed patterns of population genetic relationship in the PCA (Figure 6).

Phylogenetic Reconstruction Among Two Datasets

We finally carried out phylogenetic relationship reconstructions on the basis of the neighbor-joining (N-J) algorithm. A N-J tree based on the pairwise Reynolds' genetic distance among 64 Chinese populations (Supplementary Table S6 and Figure 7)



suggested that the Tibet Kham Tibetan was genetically closer to the surrounding Tibeto-Burman-speaking populations. Kham Tibetan was first grouped with Liangshan Tibetan, and then subsequently grouped with Tibet Tibetan and Liangshan Yi. Tibeto-Burman-speaking genetic affinity cluster was first pooled with the Chinese minority cluster, which consisted of six Altaic-speaking populations, Benzhen Manchu, and finally pooled the Han Chinese populations' cluster which was mixed with several ethnic minorities (Chengde Manchu, Yunnan Bai, Hani, Yi, Bai, and Zhuang). Six populations were served as the outliers

in this N-J tree (two Sichuan Han populations, Hebei Han, Yunnan Vietnamese and Miao). A continuity phylogenetic relationship reconstruction was performed between the Kham Tibetan and a large set of contemporary worldwide populations. **Figure 8** showed that a genealogical link was located mainly in close linguistic, ethnical and geographical proximity. Linguistic proximity could be evidently observed in Asian populations, which included Sinitic-, Tibeto-Burman-population cluster in the East Asia, Altaic-speaking populations in the Central Asia, Indo-European-speaking groups in the Europe and so on. Populations

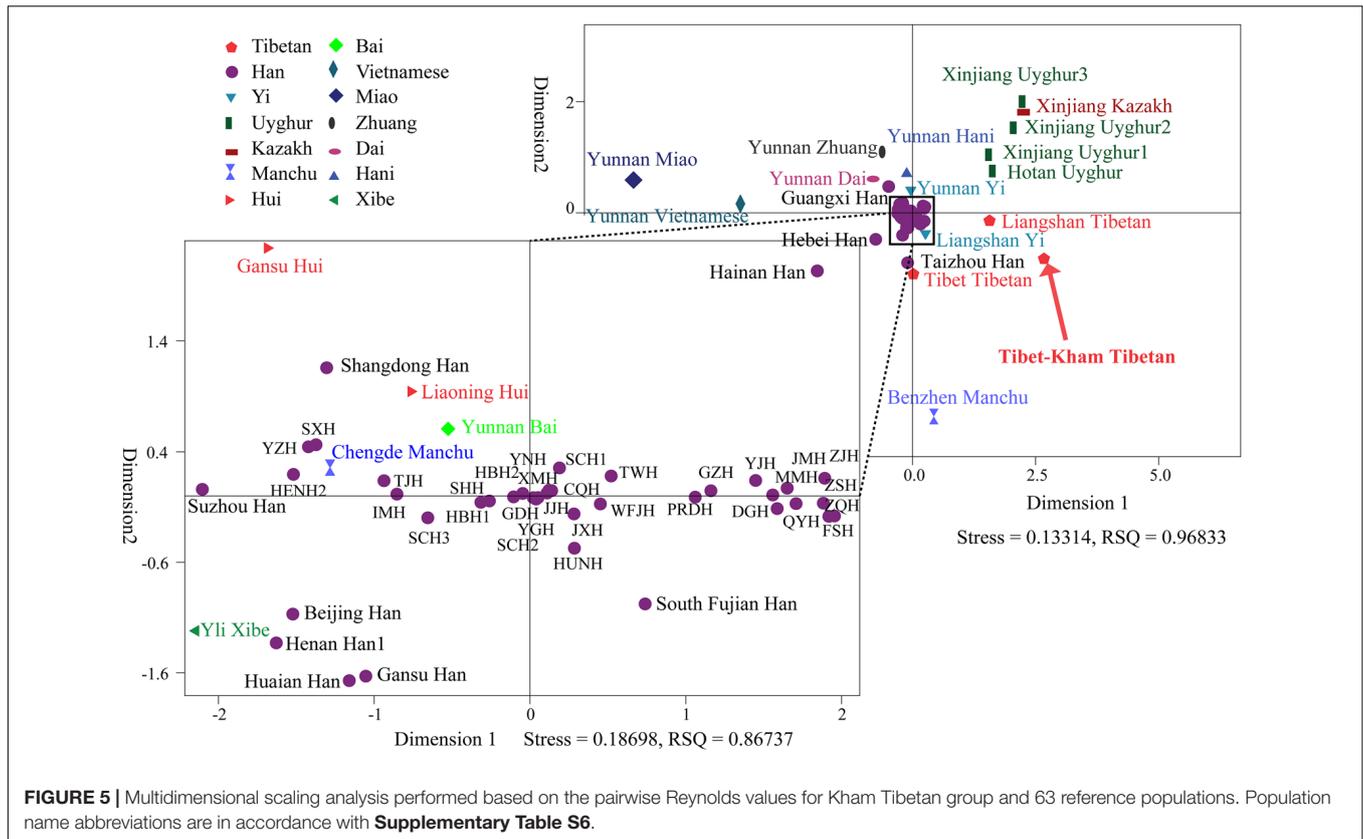


FIGURE 5 | Multidimensional scaling analysis performed based on the pairwise Reynolds values for Kham Tibetan group and 63 reference populations. Population name abbreviations are in accordance with **Supplementary Table S6**.

from one continent or language family are genetically closer to each other than other geographically or linguistically diverse populations. Genetic similarities from continentally different populations were observed between south Asian Indian and South African Indian, South Portugal Angolan and African Cape-Colored, African Afrikaner and Polish, New Zealand and Australian Caucasian and European Caucasian. All of these populations with ethnic proximity had recent large-scale population colonization, migration and genetic admixture.

DISCUSSION

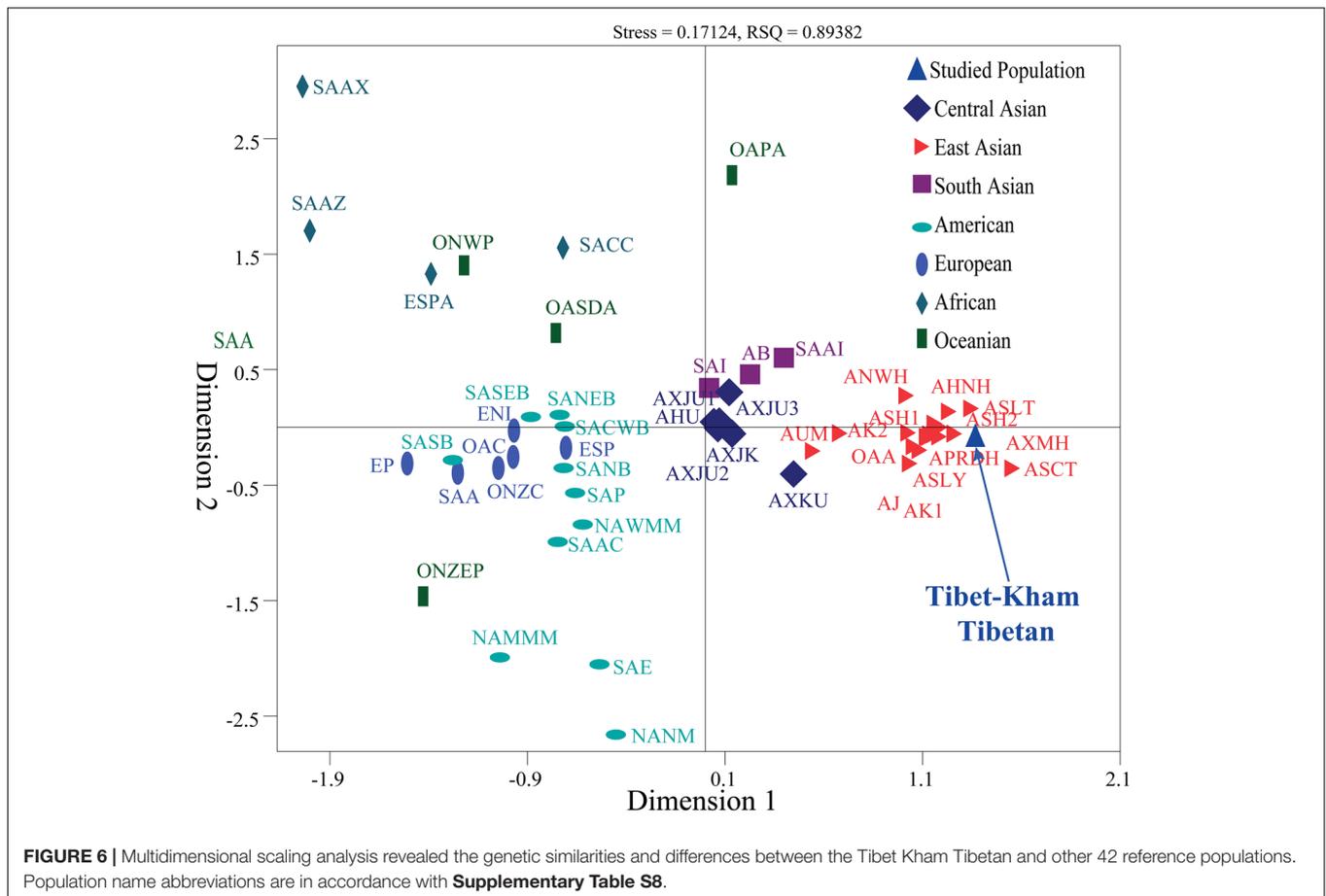
Genetic Polymorphisms and Forensic Characteristics of Kham Tibetan

The characterization and identification of genetic diversity of forensic genetic markers across ethnically diverse populations are important before employing one kind of markers or one amplification system in the forensic cases. Knowledge of the frequency and distribution of forensic markers (SNPs, STRs, insertion/deletion, multi-InDel, microhaplotype and so on) should be accurately obtained and understood to evaluate the forensic efficiency and paternity probability. Ho, He, and PIC values observed in this study indicated that the 18 autosomal STRs are high diversity and polymorphic in the Tibet Kham Tibetan. The overall forensic efficiency values of the combined power of discrimination (CPD) and the combined probability

of exclusion (CPE) are 0.999999999999999999974 and 0.999999931, respectively. This new PCR amplification system is more polymorphic and informative compared with the forensic effectiveness of 21 non-Combined DNA Index System (CODIS) autosomal STRs included in the AGCU 21+1 system, which CPD and CPE are, respectively, 0.999999999999999999999 and 0.999997 in the Liangshan Tibetan and 0.99999999999999999993 and 0.999999 in the Liangshan Yi (He et al., 2018f). Simultaneously, the discrimination and exclusion powers of this new-developed system in the Kham Tibetan are better than the previously wide-used AmpFISTR® Sinofiler™ kit, in which CPD and CPE values in 1,220 Tibetans are 0.99999999999999999997 and 0.9999996, respectively (He et al., 2018c). Moreover, the forensic efficiency is also better than 19 X-chromosomal STRs included in the AGCU X19 kit in the Tibetan population (He et al., 2018a). Thus, next-generation autosomal STRs amplification system of DNATyper™ 19 is suitable for the routine forensic applications: individual identification, parentage testing, the national database establishment, missing person identification and so on.

Genetic Relationships Between Tibetan and Nationwide or Worldwide Reference Populations

Microsatellites with the features of easy typing and availability of large numbers have been widely used to study the genetic diversity, relationship among different human populations. A previous simulation genetic study conducted



by Nei and Takezaki (1996) suggested that a more reliable phylogenetic relationship within closely related populations than between distantly related groups could be revealed by microsatellite loci. Thus, we carried out the PCA, MDS and N-J phylogenetic relationship reconstruction on the basis of genetic variations of two datasets (one dataset comprises of 18 autosomal STRs in 64 nationwide populations, and the other one consists of 16 autosomal STRs in 53 worldwide populations) to obtain an overview of genetic relationships, population substructure of Tibetans and adjacent populations. Pairwise Reynolds genetic distances indicated an affinity between the Kham Tibetan and other Tibeto-Burman-speaking populations, including Liangshan Tibetan and Yi, Chengdu Tibetan and Tibet Tibetan, suggesting their similar origin and the natural selection process. Comparisons of nationwide to worldwide genetic variation distribution also showed the significant genetic distinctions between Han Chinese populations and other East Asians and other continental residing groups. Our findings further confirmed the patterns of diversity and substructures revealed by ancestry-informative markers (Wang Z. et al., 2018), and previous population genetic findings (Qi et al., 2013; Huerta-Sanchez et al., 2014; Lu et al., 2016; Wang L.X. et al., 2018). Zhang et al. (2017) suggested Tibetan and Han Chinese populations are diverged at 6.2–16 kya and subsequently diverged with adjacent Sherpa at 3.2–11.3 kya. Recent genetic studies indicated that at

least four modern ancestry sources (East Asian, South Asian, Central Asian and Siberian, and western Eurasian and Oceanian) and four archaic ancestry sources [Neanderthal-like, Denisovan-like, ancient-Siberian-like, and even unknown ancestries which is a part of Non-modern human sequences or archaic-like signals in Tibetan gene pool identified by the S^* method (Browning et al., 2018) with the exception of aforementioned three components] exist in the modern Tibetan, as well as revealed at least two Neolithic expansions and one Paleolithic colonization (Qi et al., 2013; Huerta-Sanchez et al., 2014; Lu et al., 2016; Wang L.X. et al., 2018). These complex processes of demographic population history and genetic adaptation shaped the unique population relationship observed in the present study of this high-altitude adaptive Tibetan population.

Population Substructure in China

Our results showed that Han Chinese populations – long believed to the decedents of Yanhuang Emperors who shared similar cultural artifacts and underwent several southward migrations as well as an admixture with southern indigenous minorities – presented a population stratification (Wen et al., 2004). Significant genetic difference between North-China Han and South-China Han was identified, which is consistent with the earlier research findings via maternal mitochondrial, paternal Y-chromosomal and autosomal genetic materials

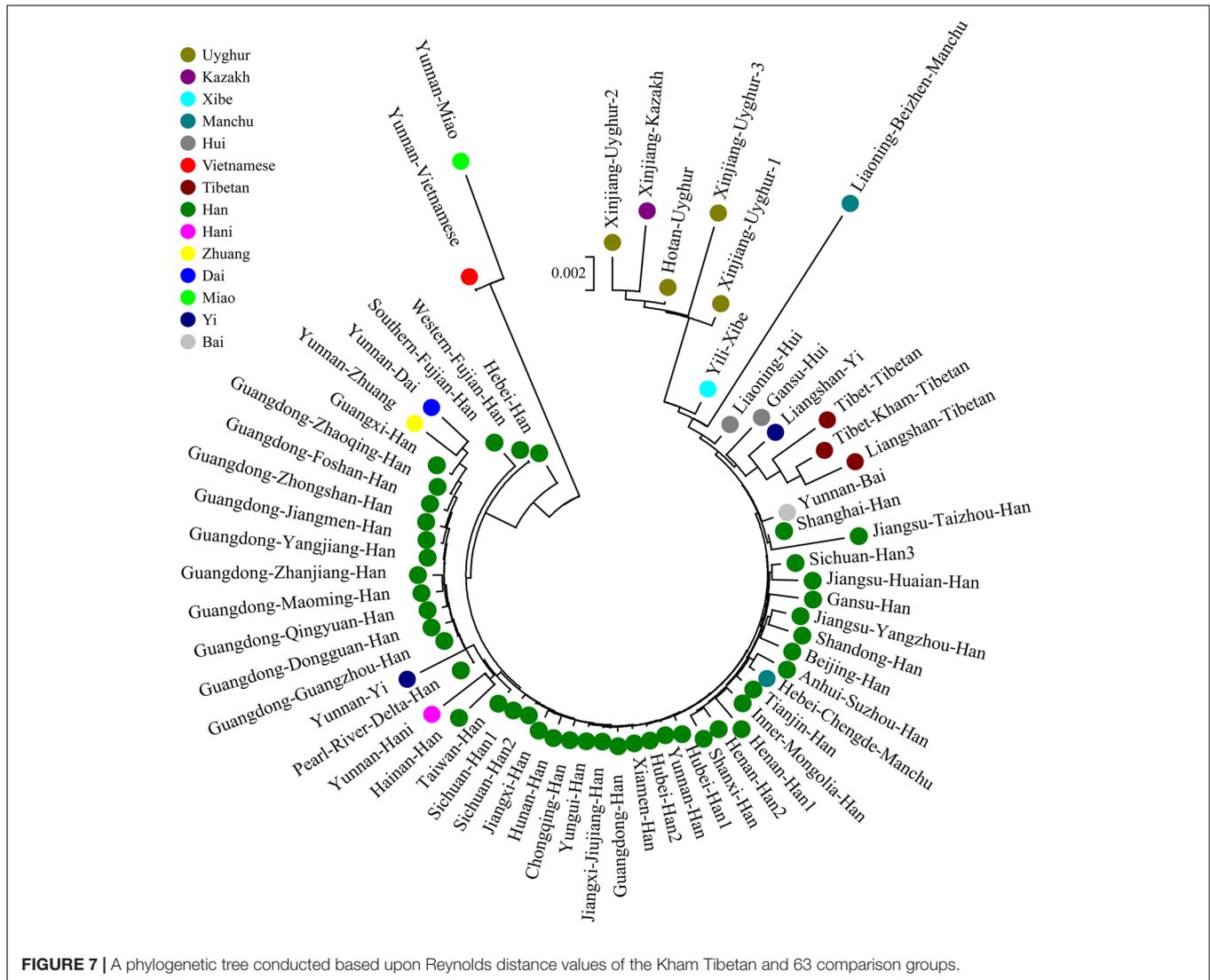


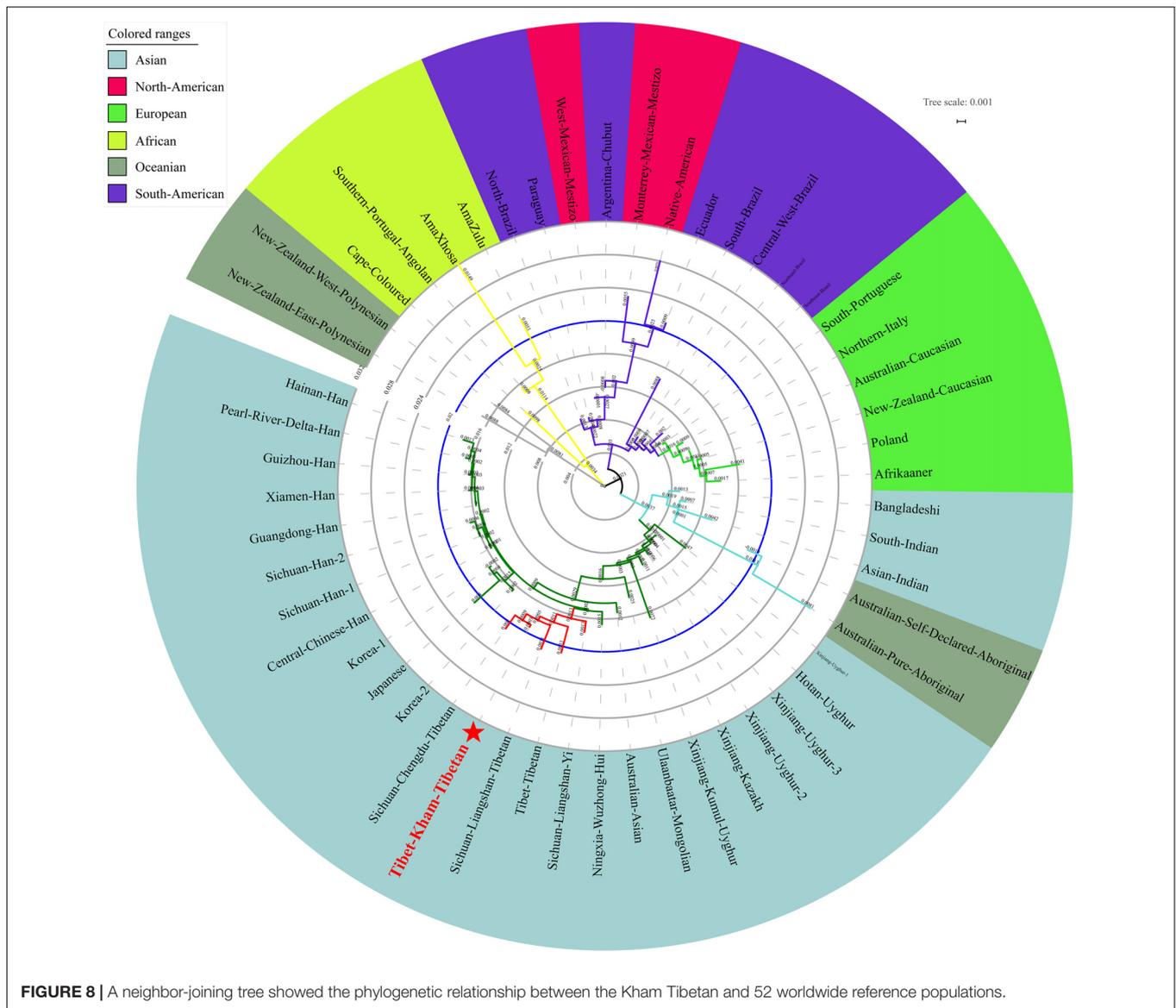
FIGURE 7 | A phylogenetic tree conducted based upon Reynolds distance values of the Kham Tibetan and 63 comparison groups.

(Chen et al., 2009; Xu et al., 2009; Nothnagel et al., 2017; Chiang et al., 2018). This North-to-South cline is dependably supported by our heat map, MDS, N-J phylogenetic relationship reconstruction and PCA analyses, as well as illustrated by the low pairwise Reynolds’ genetic distance within South Han Chinese populations and Northern Han Chinese populations and larger genetic distance between them. China is a country which is rich in the genetic, linguistic, geographical, ethnical, and cultural diversity. There are 55 officially recognized minority ethnicities and Han Chinese, which belong to seven language families [Tai-Kadai, Hmong-Mien, Sino-Tibetan (Sinitic branch and Tibeto-Burman branch), Altaic (Tungusic, Turkic and Mongolic), Austroasiatic; Indo-European and Austronesian] consisting of over 290 different recognized languages. Our population genetic comparison analyses simultaneously revealed that most minorities, especially for Altaic-speaking and Tibeto-Burman-speaking populations, possess different genetic ancestry components at varying degrees compared with other references. In our PCA and MDS analysis, we found most of the

minorities isolated and scattered compared with the tight close Han Chinese cluster. These findings are congruent with the appearance of unique local climate (High-altitude in Tibet) and intermarriage within the same cultural background and clan beliefs (Turkic-speaking populations in northwestern China). In general, separated ethnic-specific origins (56 ethnicities), enormous geographic separation (the Yangzi and Yellow Rivers as well as the Himalayas), potentially existing ongoing and substantial gene flow among ethnically, geographically and linguistically different populations may serve as the Chinese plausible demographic mechanisms to explain the patterns of genetic variations.

Worldwide Population Genetic Similarities and Differences via Autosomal STRs

The migration routes and time of the human out of Africa have been subsequently discovered and validated using patterns



of genetic variation in the maternally inherited mitochondrial DNA, paternally inherited Y chromosome and autosomal chromosome. Dramatic events accompanied by the changes in cultural interactions and social structure in prehistoric and historic times, such as worldwide Hunter-Gatherer transition, Bantu expansion in Africa, Agriculture spread from Anatolia to Europe and complex Neolithic/Bronze Age migrations from the Pontic-Caspian Steppe in Europe, Mongol Empire expansion in Eurasia and complex migrations in Oceania and America, have shaped the worldwide genomic variations of anatomically modern human (Nielsen et al., 2017). Nowadays, the sharing data with larger sample size and global population coverage in forensic science provided an opportunity to investigate the worldwide population relationship and substructure. Our results from comparative studies across 53 worldwide ethnically diverse human populations have revealed numerous genetic affinity clusters, including the Asian cluster, American cluster,

European cluster, African cluster, and Oceanian cluster. Our findings are consistent with the accumulation of population- or region-specific genetic variability under the human adaptation model of “going global by adapting the local” (Fan et al., 2016). We observed obvious genetic affinity among intra-continental populations and genetic differentiation among inter-continental populations. Although geographical structuring of worldwide populations at the continental level can be ideally identified via this simple sequence repeat, no expected genetic relationships between continental populations is observed. In this study, African and Oceanian populations clustered first in the MDS, PCA and N-J tree. Africa has substantial ethnic, cultural and linguistic diversity, which is the origin of anatomically modern humans and the source of the worldwide range modern human expansion (Beltrame et al., 2016). Cape-Colored, AmaXhosa, AmaZulu and Southern Portugal Angolan clustered with New Zealand Polynesians. Polynesians distributing across a

triangle of islands in the South Pacific are descendants of mixed Melanesian and East Asian ancestry. Besides, European and American grouped first and European populations kept genetic affinity with each other, including two immigrant Caucasian groups living in Australia and New Zealand. Anatomically modern humans started residing in Europe from 43 kya and underwent different genetic ancestry component admixtures and even population turnover (Damgaard et al., 2018; Mathieson et al., 2018; Olalde et al., 2018). The peopling of the indigenous American lately started approximately 15 kya via the Eurasia and Bering Strait, and then subsequently expanded and widespread settled in the North and South America (Raghavan et al., 2015; Moreno-Mayar et al., 2018). Generally, Africans and Oceanians are both remotely related to Asian, American, and Europeans in the tree, so they clustered together as kind of outliers. Recent population genomic studies on the basis of genetic variants of modern and ancient peoples has also demonstrated that southern Africans are a deep lineage of modern humans (Skoglund et al., 2017) and interbreeding between anatomically modern humans (Europeans, Asians, and Oceanians) and extinct hominins (Neanderthal or Denisovan) occurred (Nielsen et al., 2017; Browning et al., 2018). Beside of these ethnical-specific genetic components contributed our observed patterns, other limitations in the population comparison analyses should be with cautions in understanding population relationships: (1) Mestizos included in our included populations may be influenced the patterns of genetic relationship; (2) the included populations and marker panel density are small and more genetic information of demographically, culturally and linguistically representative is lack; (3) it is well known that high-mutated genetic marker are better used to investigate genetic history in the genetically close populations (intra-continental populations) and have limitations in precisely dissecting genetic structure in geographically isolated for a long time. In Asia, we evidently observed three Asian sub-clusters, which included the Sinitic-speaking, Turkic-speaking, and Altaic-speaking clusters. The patterns of genetic affinity are in accordance with language family boundaries, and are confirmed our previously observed genetic heterogeneity and homogeneity revealed by ancestry-informative single nucleotide polymorphisms (He et al., 2018d; Wang Z. et al., 2018), Y-chromosomal STRs (He et al., 2017a) and X-chromosomal STRs (He et al., 2017b,c, 2018a).

CONCLUSION

In summary, we presented the first batch population data of large sample size (11,402) to comprehensively evaluate the genetic diversity and forensic efficiency of DNATyper™ 19 PCR amplification system in the Kham Tibetan population. Ideal forensic measures observed in this study indicated that the 18 forensic autosomal genetic markers are polymorphic, informative and useful in forensic personal identification, parentage testing and national database establishment in Chinese Kham Tibetans. Additionally, we employed a total

of 64 Chinese nationwide populations and 53 worldwide populations as two reference panels to explore and clarify the genetic origin, genetic relationships between the Kham Tibetan and reference populations. Our comparative analysis results demonstrated that this high-altitude adaptive Kham Tibetan has genetically closer relationships with low-altitude residing Tibeto-Burman-speaking populations (Chengdu Tibetan, and Liangshan Tibetan and Yi). Finally, genetic substructure analyses in the nationwide and worldwide context suggested that the genetic proximity exists along with linguistic, ethnic, and continental geographical boundary. Additional studies with whole-genome sequencing of modern or archaic Kham Tibetans would help in reconstructing Tibetan population history.

ETHICS STATEMENT

This study was carried out according to the Declaration of Helsinki and the recommendations of “Ethical Committee of Sichuan University, China” with written informed consent from all subjects. Our protocol was approved by the “Ethical Committee of Sichuan University” (Approval Number: K2015008).

AUTHOR CONTRIBUTIONS

XZ and GH wrote the manuscript. MW, JL, PC, BG, SW, and ZL collected the samples and extracted DNA. GH, MW, XZ, and JL helped to conduct the statistical analysis. ZW revised the manuscript. YH designed this study. All authors agreed to the submission of the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (81571854 and 81501635), Open project of Key Laboratory of Forensic Genetics in Ministry of Public Security (2017FGKFKT01) and the Fundamental Research Funds for the Central University (2012017yjsy187).

ACKNOWLEDGMENTS

We would like to thank the volunteers who contributed samples for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00630/full#supplementary-material>

REFERENCES

- Beltrame, M. H., Rubel, M. A., and Tishkoff, S. A. (2016). Inferences of African evolutionary history from genomic data. *Curr. Opin. Genet. Dev.* 41, 159–166. doi: 10.1016/j.gde.2016.10.002
- Browning, S. R., Browning, B. L., Zhou, Y., Tucci, S., and Akey, J. M. (2018). Analysis of human sequence data reveals two pulses of archaic denisovan admixture. *Cell* 173, 53.e9–61.e9. doi: 10.1016/j.cell.2018.02.031
- 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
- Chiang, C. W. K., Mangul, S., Robles, C. R., and Sankararam, 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
- Cummings, M. P. (2004). *Phylip Phylogeny Inference Package*. Hoboken, NJ: John Wiley & Sons, Inc. doi: 10.1002/9780471650126.dob0534.pub2
- Damgaard, P. B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliusen, T., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. *Nature* 557, 369–374. doi: 10.1038/s41586-018-0094-2
- 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
- Fan, S., Hansen, M. E., Lo, Y., and Tishkoff, S. A. (2016). Going global by adapting local: a review of recent human adaptation. *Science* 354, 54–59. doi: 10.1126/science.aaf5098
- Gill, P., Fereday, L., Morling, N., and Schneider, P. M. (2006). The evolution of DNA databases—recommendations for new European STR loci. *Forensic Sci. Int.* 156, 242–244. doi: 10.1016/j.forsciint.2005.05.036
- 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
- Gymrek, M. (2017). A genomic view of short tandem repeats. *Curr. Opin. Genet. Dev.* 44, 9–16. doi: 10.1016/j.gde.2017.01.012
- Gymrek, M., Willems, T., Reich, D., and Erlich, Y. (2017). Interpreting short tandem repeat variations in humans using mutational constraint. *Nat. Genet.* 49, 1495–1501. doi: 10.1038/ng.3952
- Hansen, J. (2005). Using SPSS for windows and macintosh: analyzing and understanding data. *Am. Stat.* 59, 113–113. doi: 10.1198/tas.2005.s139
- Hares, D. R. (2012). Expanding the CODIS core loci in the United States. *Forensic Sci. Int. Genet.* 6, e52–e54. doi: 10.1016/j.fsigen.2011.04.012
- He, G., Chen, P., Zou, X., Chen, X., Song, F., Yan, J., et al. (2017a). 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., Zou, X., Li, P., Chen, P., Song, F., et al. (2017b). Forensic characteristics and phylogenetic analyses of the Chinese Yi population via 19 X-chromosomal STR loci. *Int. J. Legal Med.* 131, 1243–1246. doi: 10.1007/s00414-017-1563-0
- He, G., Li, Y., Zou, X., Wang, M., Chen, P., Liao, M., et al. (2017c). Genetic polymorphisms for 19 X-STR loci of Sichuan Han ethnicity and its comparison with Chinese populations. *Legal Med.* 29, 6–12. doi: 10.1016/j.legalmed.2017.09.001
- 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., Su, Y., Zou, X., Wang, M., Liu, J., et al. (2018c). Genetic variation and forensic characterization of highland Tibetan ethnicity revealed by autosomal STR markers. *Int. J. Legal Med.* 132, 1097–1102. doi: 10.1007/s00414-017-1765-5
- He, G., Wang, Z., Wang, M., Luo, T., Liu, J., Zhou, Y., et al. (2018d). 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. (2018e). Genetic variations and forensic characteristics of Han Chinese population residing in the pearl river delta revealed by 23 autosomal STRs. *Mol. Biol. Rep.* 11, 1–9. doi: 10.1007/s11033-018-4264-y
- He, G., Wang, Z., Zou, X., Chen, X., Liu, J., Wang, M., et al. (2018f). Genetic diversity and phylogenetic characteristics of Chinese Tibetan and Yi minority ethnic groups revealed by non-CODIS STR markers. *Sci. Rep.* 8:5895. doi: 10.1038/s41598-018-24291-5
- Hu, H., Petousi, N., Glusman, G., Yu, Y., Bohlender, R., Tashi, T., et al. (2017). Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genet* 13:e1006675. doi: 10.1371/journal.pgen.1006675
- Huerta-Sanchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., et al. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512, 194–197. doi: 10.1038/nature13408
- Kayser, M., and de Knijff, P. (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nat. Rev. Genet.* 12, 179–192. doi: 10.1038/nrg2952
- Kovach, W. L. (2007). *MVSP-A MultiVariate Statistical Package for Windows, ver. 3.1*. Pentraeth: Kovach Computing Services.
- 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
- 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
- Madsen, D. B., Ma, H. Z., Brantingham, P. J., Xing, G., Rhode, D., Zhang, H. Y., et al. (2006). The late upper paleolithic occupation of the northern tibetan plateau margin. *J. Archaeol. Sci.* 33, 1433–1444. doi: 10.1016/j.jas.2006.01.017
- Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szecsenyi-Nagy, A., Rohland, N., Mallick, S., et al. (2018). The genomic history of southeastern Europe. *Nature* 555, 197–203. doi: 10.1038/nature25778
- Moreno-Mayar, J. V., Potter, B. A., Vinner, L., Steinrucken, M., Rasmussen, S., Terhorst, J., et al. (2018). Terminal pleistocene alaskan genome reveals first founding population of native americans. *Nature* 553, 203–207. doi: 10.1038/nature25173
- Nei, M., and Takezaki, N. (1996). The root of the phylogenetic tree of human populations. *Mol. Biol. Evol.* 13, 170–177. doi: 10.1093/oxfordjournals.molbev.a025553
- 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
- Olalde, I., Brace, S., Allentoft, M. E., Armit, I., Kristiansen, K., Booth, T., et al. (2018). The beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555, 190–196. doi: 10.1038/nature25738
- 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
- Qi, X., Cui, C., Peng, Y., Zhang, X., Yang, Z., Zhong, H., et al. (2013). Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. *Mol. Biol. Evol.* 30, 1761–1778. doi: 10.1093/molbev/mst093
- Raghavan, M., Steinrucken, M., Harris, K., Schiffels, S., Rasmussen, S., Degiorgio, M., et al. (2015). Population genetics. genomic evidence for the pleistocene and recent population history of native americans. *Science* 349:aab3884. doi: 10.1126/science.aab3884
- Schneider, P. M. (2007). Scientific standards for studies in forensic genetics. *Forensic Sci. Int.* 165, 238–243. doi: 10.1016/j.forsciint.2006.06.067
- Skoglund, P., Thompson, J. C., Prendergast, M. E., Mittnik, A., Sirak, K., Hajdinjak, M., et al. (2017). Reconstructing prehistoric african population structure. *Cell* 171, 59.e21–71.e21. doi: 10.1016/j.cell.2017.08.049
- Sun, J. X., Helgason, A., Masson, G., Ebenesersdottir, S. S., Li, H., Mallick, S., et al. (2012). A direct characterization of human mutation based on microsatellites. *Nat. Genet.* 44, 1161–1165. doi: 10.1038/ng.2398

- Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., et al. (2009). The genetic structure and history of Africans and African Americans. *Science* 324, 1035–1044. doi: 10.1126/science.1172257
- Wang, D. Y., Gopinath, S., Lagace, R. E., Norona, W., Hennessy, L. K., Short, M. L., et al. (2015). Developmental validation of the globalFiler(R) express PCR amplification kit: A 6-dye multiplex assay for the direct amplification of reference samples. *Forensic Sci. Int. Genet.* 19, 148–155. doi: 10.1016/j.fsigen.2015.07.013
- Wang, L. X., Lu, Y., Zhang, C., Wei, L. H., Yan, S., Huang, Y. Z., et al. (2018). Reconstruction of Y-chromosome phylogeny reveals two neolithic expansions of Tibeto-Burman populations. *Mol. Genet. Genomics* doi: 10.1007/s00438-018-1461-2 [Epub ahead of print].
- Wang, Z., He, G., Luo, T., Zhao, X., Liu, J., Wang, M., et al. (2018). Massively parallel sequencing of 165 ancestry informative SNPs in two Chinese Tibetan-Burmese minority ethnicities. *Forensic Sci. Int. Genet.* 34, 141–147. doi: 10.1016/j.fsigen.2018.02.009
- 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
- Werrett, D. J. (1997). The national dna database. *Forensic Sci. Int.* 88, 33–42. doi: 10.1016/S0379-0738(97)00081-9
- Willems, T., Gymrek, M., Highnam, G., Genomes Project, C., Mittelman, D., and Erlich, Y. (2014). The landscape of human STR variation. *Genome Res.* 24, 1894–1904. doi: 10.1101/gr.177774.114
- Willems, T., Zielinski, D., Yuan, J., Gordon, A., Gymrek, M., and Erlich, Y. (2017). Genome-wide profiling of heritable and de novo STR variations. *Nat. Methods* 14, 590–592. doi: 10.1038/nmeth.4267
- 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
- Zhang, C., Lu, Y., Feng, Q., Wang, X., Lou, H., Liu, J., et al. (2017). Differentiated demographic histories and local adaptations between Sherpas and Tibetans. *Genome Biol.* 18:115. doi: 10.1186/s13059-017-1242-y
- Zhao, M., Kong, Q. P., Wang, H. W., Peng, M. S., Xie, X. D., Wang, W. Z., et al. (2009). Mitochondrial genome evidence reveals successful late paleolithic settlement on the tibetan plateau. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21230–21235. doi: 10.1073/pnas.0907844106

Conflict of Interest Statement: 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.

Copyright © 2018 Zou, Wang, He, Wang, Su, Liu, Chen, Wang, Gao, Li and Hou. 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.