

SCIENTIFIC REPORTS



OPEN

Genetic structure and forensic characteristics of Tibeto-Burman-speaking Ü-Tsang and Kham Tibetan Highlanders revealed by 27 Y-chromosomal STRs

Guanglin He¹, Zheng Wang¹, Yongdong Su², Xing Zou¹, Mengge Wang¹, Xu Chen³, Bo Gao⁴, Jing Liu¹, Shouyu Wang¹ & Yiping Hou¹

Culturally diverse Tibetans (Ü-Tsang, Kham and Ando) harboring a unique molecular mechanism that allows them to successfully adapt to hypoxic environments in the Qinghai-Tibet Plateau have been a subject of great interest in medical genetics, linguistics, archeology and forensic science. However, forensic characteristics and genetic variations of the Y-chromosomal 27-marker haplotype included in the Yfiler Plus system in the Ü-Tsang and Kham Tibeto-Burman-speaking Tibetans remain unexplored. Thus, we genotyped 27 Y-STRs in 230 Shigatse Ü-Tsang Tibetans (SUT) and 172 Chamdo Kham Tibetans (CKT) to investigate the forensic characterization and genetic affinity of Chinese Tibetan Highlanders. The haplotype diversities were 0.999962028 in SUT and 0.999796002 in CKT. Forensic diversity measures indicated that this 27-Y-STR amplification system is appropriate for routine forensic applications, such as identifying and separating unrelated males in deficiency paternity cases, male disaster victims and missing person identification and determining male components in sexual assault cases. Moreover, the genetic relationships among 63 worldwide populations (16,282 individuals), 16 Asian populations, and 21 Chinese populations were analyzed and reconstructed using principal component analysis, multidimensional scaling plots and a phylogenetic tree. Considerable genetic differences were observed between Tibetan populations and other geographically/ethnically diverse populations (Han Chinese). Our studied SUT and CKT have a genetically closer relationship with Gansu Ando Tibetans than with other Asians. In total, our analyses indicated that subpopulation structures exist among Asian and Chinese populations, and population-specific reference databases should be established for forensic applications.

The Qinghai-Tibet Plateau, also referred to as the Himalayan Plateau, stretches approximately 1,000 kilometers (km) from the Kunlun Mountains and Qilian Mountains in the north to the inner Himalayan Range in the south and 2,500 km from northwestern Yunnan and western Sichuan in the east to the rugged Karakoram Range of northern Kashmir in the west. This elevated plateau with an average elevation exceeding 4,500 meters harbors the world's highest summits (Mount Everest) and is called the Roof of the World. Tibetans, with a population of over 7.5 million, settled in the Qinghai-Tibet Plateau and other parts of China. There are also significant numbers of Tibetans (approximately 1.5 million) residing in India, Pakistan, Bhutan, Nepal and other countries. Ancestry history and genetic origin as well as the processes of Tibetan adaptation in extreme environments (high altitude,

¹Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu, Sichuan, 610041, China. ²Forensic Identification Center, Public Security Bureau of Tibet Tibetan Autonomous Region, Lhasa, Tibet Tibetan Autonomous Region, 850000, China. ³Department of Clinical Laboratory, the First People's Hospital of Liangshan Yi Autonomous Prefecture, Xichang, Sichuan, 615000, China. ⁴Yili Public Security Bureau, Yili, Xinjiang Uygur Autonomous Region, 418000, China. Guanglin He, Zheng Wang and Yongdong Su contributed equally. Correspondence and requests for materials should be addressed to Y.H. (email: forensic@scu.edu.cn)

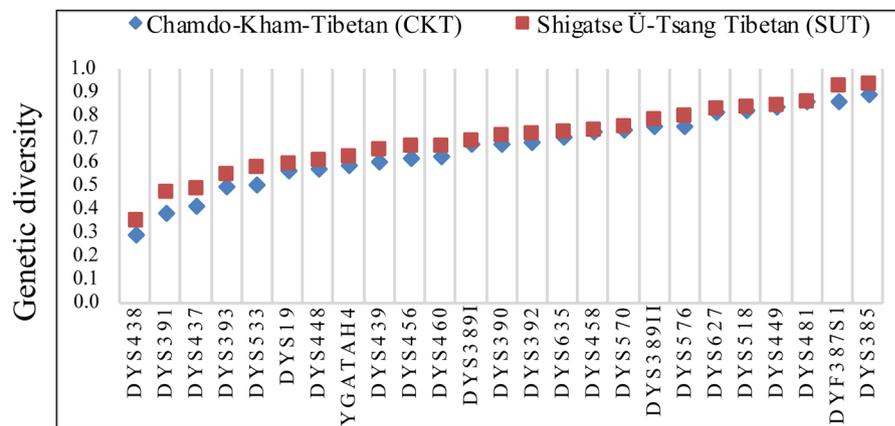


Figure 1. Genetic diversity (GD) of 27 Y-chromosomal STRs included in the Yfiler Plus amplification system in two studied Tibeto-Burman-speaking Ü-Tsang and Kham Tibetan Highlanders.

low temperature, hypobaric hypoxia and high ultraviolet radiation) have attracted the attention of archeologists, anthropologists, and geneticists^{1–3}. Archeological evidence suggested that the initial stage occupation of the Tibetan Plateau occurred 40–25 thousand years ago (kya)⁴. Y-chromosome evidence based on the haplogroup distribution of D-M174 found that the earliest modern human settlement and migration in East Asia occurred approximately 60 kya with multiple origins of Tibetans and complex admixture occurring during the subsequent northward migration⁵. Maternal fine-scale mitochondrial genome variations revealed that Tibetan-specific mtDNA lineages entered the Tibetan Plateau in both the Late Paleolithic and mid-Holocene (M16)⁶. Recently, new geological, paleoclimatological and archeological evidence suggested that the earliest occupations of the Tibetan Plateau could date back as early as 30 kya². Recently, novel genetic findings via Y-chromosomal, mitochondrial and genome-wide autosomal markers demonstrated the Upper Paleolithic colonization and Neolithic expansion of anatomically modern humans in Tibet⁷, as well as the coexistence of Paleolithic and Neolithic ancestries in modern Tibetans⁸. Early whole-genome association studies have successfully identified that the genetic variants in the high-altitude adaptation genes (EPAS1 and EGLN1) are correlated significantly with hemoglobin concentration^{9–14}. Additional genetic research indicated that at least three evolutionary mechanisms, including admixture with surrounding Highlanders, interbreeding with archaic Denisovans, and natural selection, have driven Tibetans to adapt to the extreme environments^{15–18}.

In addition, there are several subgroups with different cultural backgrounds in the Tibetan population separated by the Alpine and Gorge Regions, including Kham Tibetans, Ü-Tsang Tibetans and Ando Tibetans. Kham Tibetans settled in the rugged terrain located at the boundary of the eastern region in the Tibet Autonomous Region, the western region of Sichuan Province and part of the southern region of Qinghai and Gansu Provinces as well as smaller portions of the northern region of Yunnan Province. Linguists and anthropologists regard this region as the Tibetan-Yi Corridor or Ethnic Corridor of Southwest China due to the complex landscapes of culturally and linguistically distinct ethnic groups. Most of the people residing in the Tibetan-Yi Corridor speak the Qiangic language as their mother language; however, Kham Tibetans speak the Tibetic language. Ü-Tsang Tibetans are also called central Tibetans, Dbus, or Ü Tibetans and mainly reside in the central and western region of the Tibet Autonomous Region. Ü-Tsang Tibetan use the standard conservative Tibetan language. Ando Tibetans mainly live in Qinghai Province and some regions of the Sichuan and Gansu Provinces. Their language is Andolese. Zhang *et al.* recently identified the differentiated population genetic structure among the aforementioned geographically different Tibetans³. Thus, a clear elucidation and understanding of the genetic polymorphisms and forensic characteristics of forensic genetic markers (maternal mitochondrial DNA, paternal Y-chromosomal and autosomal STRs) are necessary and important.

Y-chromosome variations situated within the nonrecombining region, which are paternally inherited and haploid, have received increased attention as classical genetic markers in evolutionary and forensic genetic studies, such as tracing and reconstructing the paternal population history and evaluating population structure and relationships^{19,20}. Approximately 4,500 Y-chromosome short tandem repeats (Y-STRs) have been reported and subsequently used in reconstructing human male population evolutionary history and other multidisciplinary research²¹. In the forensic field, Y-STRs can be used in paternal relationship identification in deficiency paternity cases, male disaster victims and missing person identification and can help determine the male components in sexual assault cases and family research in forensics²². In the past decade, many commercial amplification systems that include 17 to 24 Y-STR loci, for instance, Yfiler (17 Y-STRs), PowerPlex Y23 (23 Y-STRs), AGCU Y24 and GFS Y24 (24 Y-STRs), have been developed and used in forensics and population genetics^{23–26}. Recently, a next-generation Y-STR genotyping system comprising 27 Y-STR loci, including all 17 loci (DYS456, DYS438, DYS458, DYS635, DYS439, DYS448, DYS389I, DYS19, DYS391, DYS392, DYS437, DYS389II, DYS390, DYS393, DYS385a/b and Y-GATA H4) included in the Yfiler kit, seven rapidly mutating Y-STR loci (DYF387S1a/b, DYS449, DYS518, DYS570, DYS576 and DYS627)²⁷, and three highly polymorphic Y-STR loci (DYS460, DYS481 and DYS533)²⁸ was developed for forensic Y-STR database establishment and forensic routine applications. Current population genetic studies have so far focused on a limited number of ethnically, geographically and linguistically restricted populations or targeted autosomal and X-chromosomal markers^{29–34}.

No. of observed haplotypes	CKT					SUT					Qinghai Ando Tibetan				
	MHT	EXT	PPY12	Yfiler	Yfiler Plus	MHT	EXT	PPY12	Yfiler	Yfiler Plus	MHT	EXT	PPY12	Yfiler	Yfiler Plus
1(unique)	76	86	86	117	166	147	160	169	210	228	426	430	432	434	434
2	10	9	11	21	3	17	17	17	10	1	38	36	35	34	34
3	6	8	8	3		3	5	3			3	3	3	3	3
4	4	4	4	1		5		3							
5	1					1	3								
6		1	2			1	1	1							
9						1									
10	1	1													
11	1														
12		1	1												
16	1														
Sample Size	172	172	172	172	172	230	230	230	230	230	511	511	511	511	511
No. of haplotypes	100	110	112	142	169	175	189	193	220	229	467	469	470	471	471
Fraction of unique haplotypes	0.4419	0.5000	0.5000	0.6802	0.9651	0.6391	0.6957	0.7348	0.9130	0.9913	0.8337	0.8415	0.8454	0.8493	0.8493
Match probability	0.0249	0.0182	0.0163	0.0082	0.0060	0.0088	0.0073	0.0066	0.0047	0.0044	0.0023	0.0023	0.0023	0.0023	0.0023
Haplotype diversity	0.9808	0.9876	0.9895	0.9976	0.9998	0.9956	0.9971	0.9978	0.9996	>0.9999	0.9996	0.9997	0.9997	0.9997	0.9997

Table 1. The forensic parameters estimated for different combinations of Y-STRs in three different Tibetan populations MHT: minimal haplotype; EHT: the extended haplotype; PPY12: PowerPlex Y haplotype; Yfiler: Yfiler PCR Amplification Kit; Yfiler Plus: Yfiler® Plus PCR Amplification Kit. SUT: Shigatse Ü-Tsang Tibetans (SUT), CKT: Chamdo Kham Tibetans.

However, existing genetic data from the Y-chromosome of Tibetan Highlanders (only Gansu and Qinghai Ando Tibetans) are insufficient to investigate forensic features and characterize the complex genetic relationships between and within Tibetans and other East Asians³⁵. To gain insights into the genetic diversity and forensic characteristics of forensic Y-chromosomal genetic markers, we genotyped and analyzed 27 Y-STRs in 230 Shigatse Ü-Tsang Tibetans (SUT) and 172 Chamdo Kham Tibetans (CKT). Subsequently, we performed comprehensive population comparisons corresponding to cultural, geographical or linguistic groups to dissect the genetic similarities and differences between our studied Tibetan populations and Chinese reference populations or other adjacent Tibetan-Burman speaking populations^{20,35–42}.

Results and Discussion

The Y-chromosomal haplotype reference database (YHRD) was developed 16 years ago and is a robust and reliable resource to revisit the worldwide or regional male genetic landscape, dissect Y-specific population differentiation and evaluate the forensic characteristics. Currently, this database includes 77,805 Yfiler haplotypes (17-Y-STRs), 12,430 PowerPlex Y23 haplotypes (23-Y-STRs) and 26,933 Yfiler Plus haplotypes (27-Y-STRs) from China. China is the world's most populous country, consisting of 56 ethnicities from more than 30 administrative divisions. However, no Yfiler Plus haplotype data were available for CKT and SUT residing in the Tibet Tibetan autonomous region. In this study, we first examined the patterns of Y-chromosome variations in these two culturally different Tibetan populations and investigated the genetic heterogeneity and homogeneity between Tibetan populations and other Asian, nationwide or Himalayan-related reference populations^{20,35–41}.

Genetic polymorphisms and forensic characteristics of 27 Y-STRs in Tibetans. In total, 230 SUT and 172 CKT subjects were successfully genotyped, and the corresponding genotype data and haplotype data are presented in Supplementary Table S1. In addition, all our raw genotypes were submitted to the YHRD with accession numbers of YA004561 (CKT) and YA004562 (SUT). Three duplicate alleles at three single-copy loci (15, 17 at DYS19 locus; 18, 19 at DYS576 locus; 18, 19 at DYS570 locus) were observed. Intermediate alleles of 19.2 at the DYS627 locus in one SUT and 37.2 at the DYS518 locus in four CKT people were detected. The allele frequencies and genetic diversity of 23 single-copy Y-STR loci and two multicopy loci are presented in Fig. 1 and Supplementary Tables S2 and 3. A total of 144 alleles with corresponding frequencies ranging from 0.0058 to 0.8256 in CKT and 160 alleles with corresponding allele frequencies ranging from 0.0043 to 0.7826 in SUT were observed in the 23 single-copy loci. GD values spanned from 0.2936 (DYS438) to 0.8630 (DYS481) in CKT and from 0.3512 (DYS391) to 0.8625 (DYS518) in SUT. Twenty-eight allelic combinations with 9 different alleles at the DYS387S1a/b locus and 47 allelic combinations with 13 different alleles at the DYS385a/b locus in SUT were observed. In addition, we detected 18 allelic combinations with 7 different alleles at the DYS387S1a/b locus and 31 allelic combinations with 12 distinct alleles at the DYS385a/b locus in CKT.

Among 172 CKT genotypes, 169 different haplotypes consisting of 166 unique haplotypes and three duplicated haplotypes were detected. The match probability (MP) and HD were 0.0060 and 0.999796, respectively. A total of 229 different haplotypes were observed in 230 SUT, of which 228 were unique and one was a duplicate. The MP and HD were 0.0044 and 0.999962, respectively. To comprehensively evaluate the forensic discrimination

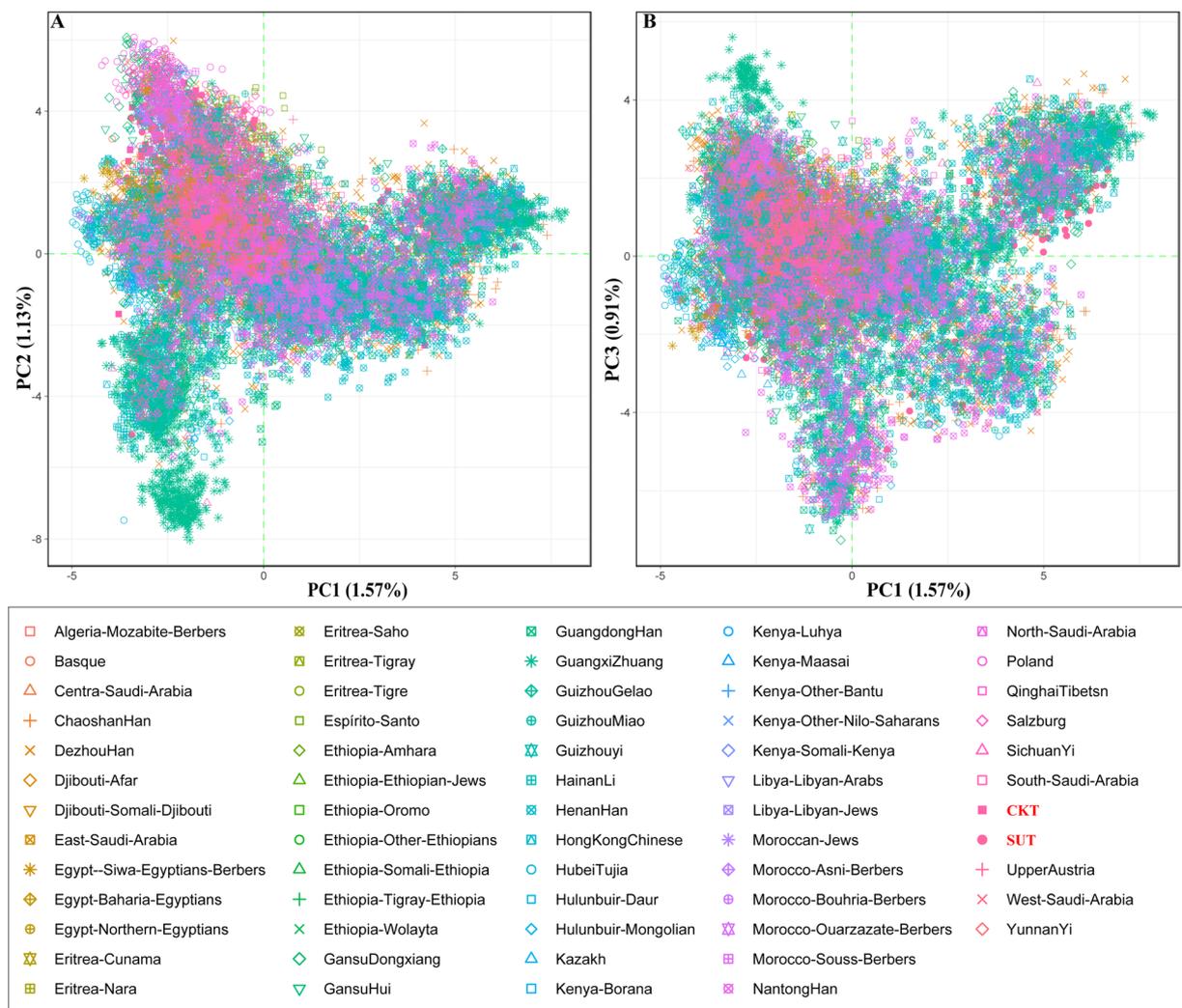


Figure 2. Principal component analysis among 63 worldwide populations.

power of different marker compositions in three culturally different Tibetan populations, we combined our newly genotyped data with previously published Yfiler Plus haplotype data that included 27 Y-STRs from the Qinghai Ando Tibetan population³⁵. The forensic parameters of different marker sets, including the minimal haplotype, the extended haplotype, the PowerPlex Y haplotype, and the Yfiler and Yfiler® Plus PCR Amplification systems in CKT, SUT and previously reported Qinghai Ando Tibetans³⁵, are presented in Table 1. The results indicated that the forensic discrimination power increases when the number of analyzed markers increases, especially the six included rapidly mutating Y-STRs⁴³. This Y-STR genotyping system is discriminating and polymorphic in the different Tibetan populations with different cultural backgrounds and is suitable as a powerful tool for identifying paternal lineages, linking suspects and crime samples and conducting forensic family research in Tibetan populations.

Worldwide population genetic relationship inferred from the raw genotype data. To explore the population genetic relationship capitalizing the worldwide Y-chromosomal STR variations, we merged our data with the publicly available data from 27 Y-STRs from 15,880 individuals (Table S4). The top three principal components (PC) could explain 3.61% of the variation (Fig. 2). No clear population stratification was observed. This may be caused by the high proportion of shared Y-chromosome haplotypes in modern male populations. The pairwise F_{st} genetic distances among 63 populations were calculated and visualized as the heat map in Fig. S1. The Dezhou Han population harbors the smallest genetic distance between both CKT (0.0112) and SUT (0.0054). We can also identify the relatively small genetic distances between Tibetan and Northeast African populations (Fig. 3), which may be caused by the small sample number of these African populations. A phylogenetic relationship reconstruction based on the pairwise F_{st} genetic distances generally shows the distinct genetic affinity between African and Asian populations, but we also found confusing cluster patterns between Northeast African populations and others. Two newly studied Tibetan populations maintain a close phylogenetic relationship with Qinghai Tibetan populations (Fig. S2). To obtain a clearer genetic landscape of Tibetan populations and more adjacent populations with larger sample sizes, we focused on the genetic diversity in Asia and China.

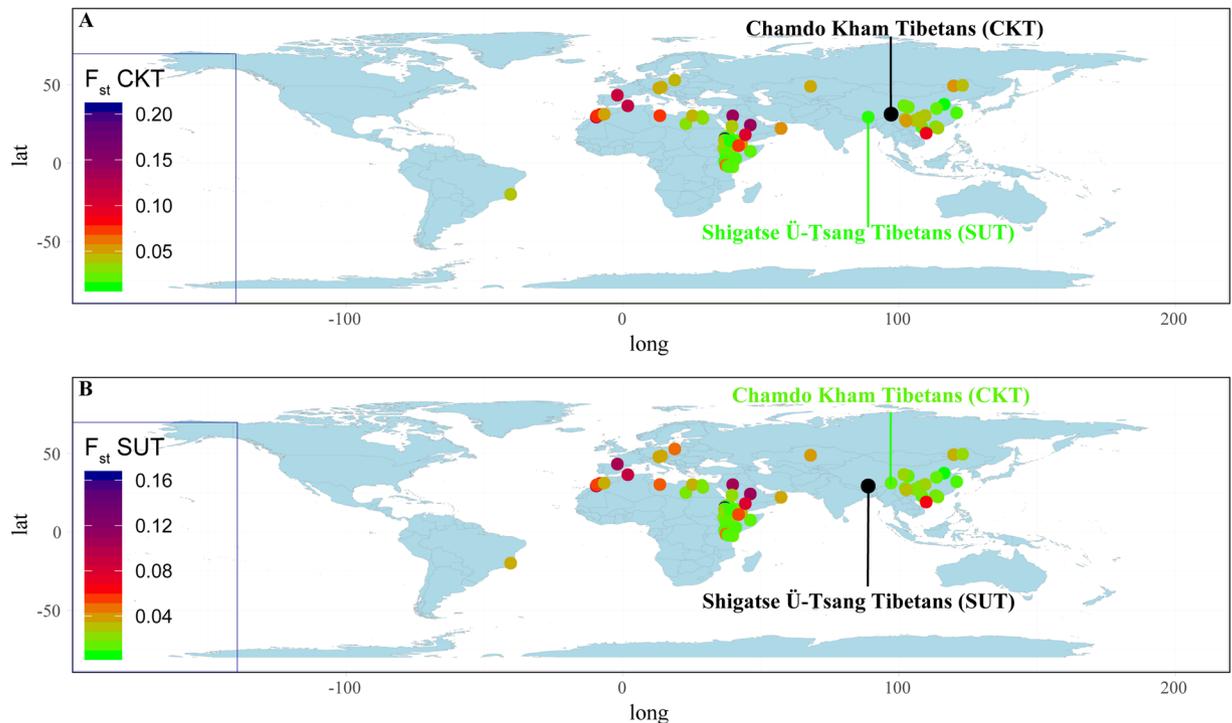


Figure 3. Heatmaps show the pairwise genetic distances between Tibetans and the other 61 reference populations.

Overall genetic affinity between Tibetans and other Asian populations. To obtain a generally comprehensive picture of genetic architecture between two newly studied Tibetan populations and Asian populations, we first merged our haplotype data from 402 samples with previously published haplotype data from 9,577 samples in fourteen populations (Fig. 4A). Pairwise R_{st} standard genetic distances among sixteen populations were calculated and visualized in Fig. 4B. CKT was genetically closest to SUT ($R_{st} = 0.029$), followed by Central Asian populations of Pakistan ($R_{st} = 0.1002$), Kazakhstan ($R_{st} = 0.1292$) and Afghanistan ($R_{st} = 0.1458$). Shigatse Tibetans had the smallest genetic distance from CKT ($R_{st} = 0.029$), followed by the East Asian populations of Nantong Han ($R_{st} = 0.0652$), Henan Han ($R_{st} = 0.0723$) and Shanghai Han ($R_{st} = 0.738$). Genetic homogeneity and heterogeneity on the basis of the R_{st} distance matrix were further explored via multidimensional scaling plots (MDS). As shown in Fig. 4C, CKT and SUT maintain a relatively close relationship and are located in the first quadrant. East Asian populations and one group from the Philippines are clustered together and positioned in the second quadrant. The Southeast Asian populations from Thailand and Laos are localized in the third quadrant. Central Asian populations (Pakistan, Kazakhstan and Afghanistan) and one South Asian population from India are grouped in the fourth quadrant. Consistent patterns of genetic affinity among sixteen populations were reconstructed in the neighbor-joining tree (Fig. 4D). East Asian and Southeast Asian populations form one branch, and our newly studied populations combined with the Central Asian population and South Asian population form the other branch. In total, the genetic similarities observed in the Asian populations correspond to geographical divisions.

Population structure and phylogenetic relationships among Chinese populations. Genetic affinity and population structure between CKT, SUT and Chinese reference populations^{20,35–41} were then evaluated using the pairwise genetic distances, MDS and phylogenetic relationship reconstruction (Fig. 5A). The pairwise R_{st} distances with corresponding p values are presented in Supplementary Table S5, and a heatmap of the pairwise genetic distance matrix is presented in Fig. 5B. Among the 21 Chinese populations, CKT has the closest genetic relationships with Gansu Ando Tibetans ($R_{st} = 0.0281$) and SUT ($R_{st} = 0.0291$) and has significantly distinct genetic relationships with the Hainan Li ($R_{st} = 0.2174$) and Yanbian Korean ($R_{st} = 0.2174$) populations, which has overall temperate genetic variances (average \pm standard variance: 0.1478 ± 0.0511). Similar patterns of genetic variation and population relationships of SUT were observed with an average standard variance of 0.0726 ± 0.0318 . Population genetic relationships are subsequently reconstructed using the MDS and neighbor-joining tree. As shown in Fig. 5C, MDS shows a genetic homogeneity among three Tibetan populations (Gansu Ando, Chamdo Kham and Shigatse Ü-Tsang), as well as genetic similarity within eight Han Chinese populations (Shanghai, Chaoshan, Hainan, Shenzhen, Beijing, Chongqing, Jinling and China), Guizhou Gelao and Hainan Lingao. Significant genetic differences were observed among the Yanbian Korean, Gansu Hui, Hanna Li, Gansu Dongxiang, Qinghai Tibetan, Inner Mongolia Daur and Hubei Tujia populations. Two clusters of population relationship patterns were observed in the phylogenetic tree: four Tibetan populations and one Dongxiang population form one genetic affinity cluster, and other Chinese minorities and Han Chinese populations form the other cluster (Fig. 5D).

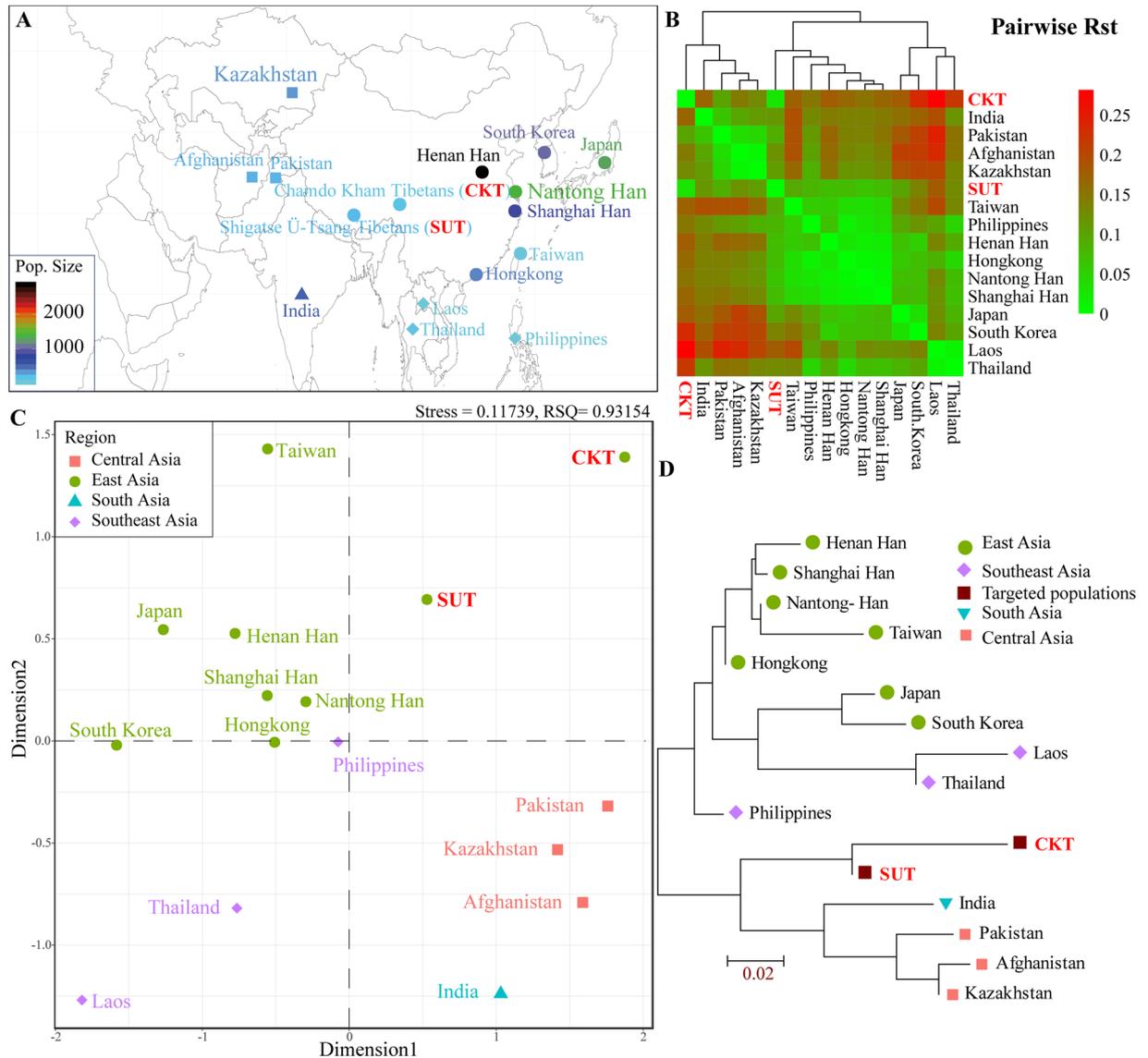


Figure 4. Genetic affinity among 16 Asian populations. (A) The map shows the geographic location and population size of the 16 Asian populations. The map was generated by free software R 3.3.2 (<https://www.R-project.org/>)⁴⁹ using the ggrepel, ggplot2, maps and mapdata packages and then modified using Adobe Illustrator CS6 software. The R script is submitted in Supplementary Note 1. (B) The heatmap shows the pairwise genetic distance among 16 included populations. (C) Multidimensional scaling plots show the genetic similarities among Asian populations. (D) The phylogenetic tree shows the relationship among 16 groups.

Genetic distance comparisons revealed that strong genetic affinity exists among Tibetans along different cultural backgrounds and geographical adjacent minorities (Gansu Dongxiang), as well as significant genetic differentiation between Tibetans and homogeneous Han Chinese populations. Significant genetic differences between Tibetan populations and other reference populations may be explicable by unique demographical history, genetic origin and divergence. Archeological evidence based on Paleolithic sites on the Tibetan Plateau suggested that anatomically modern humans resided in this region with the most extreme environmental conditions over 20 kya^{2,44}. Tibetan matrilineal genetic evidence indicated that Tibetans are descended from Epipaleolithic and Neolithic immigrants from northern China⁶. Recently, a large number of genetic studies found that unique natural selection mechanisms have driven Tibetans to accumulate different genetic legacies, especially genetic mutations in EPAS1 and EGLN1^{9,17}. Lu *et al.* found that Tibetans and the Han Chinese diverged more than 15 kya and found genetic coexistence and continuity between prehistorical highland-foragers (Paleolithic, Neolithic ages) and modern Tibetans¹. Our current study provided a forensic reference database of Chinese Tibetan populations and general population genetic relationships between Chinese Tibetan populations and other global populations. Further genetic studies of the combined testing of modern people and ancient human remains are needed to provide clear and more accurate demographical and evolutionary histories of this Tibetan population adapted to high altitude.

and 230 Shigatse Ü-Tsang male Tibetans residing in West Tibet. All donors were unrelated healthy indigenous people and provided written informed consent. DNA was extracted using the PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific, Waltham, MA, USA) and quantified via the Applied Biosystems 7500 Real-time PCR System (Thermo Fisher Scientific) using the Quantifiler Human DNA Quantification Kit (Thermo Fisher Scientific) following the manufacturer's recommendations.

DNA amplification and genotyping. Co-amplification of 27 Y-STRs (DYS576, DYS19, YGATAH4, DYS448, DYS391, DYS456, DYS390, DYS438, DYS392, DYS627, DYS460, DYS458, DYS449, DYS393, DYS518, DYS570, DYS437, DYS385, DYS389I, DYS635, DYS389II, DYS439, DYS481, DYS387S1 and DYS533) was performed using the Yfiler Plus PCR Amplification Kit (Thermo Fisher Scientific) on the ProFlex 3 × 32 PCR System (Thermo Fisher Scientific) according to the manufacturer's instructions. Amplified DNA products were separated and detected using capillary electrophoresis on a 3500XL Genetic Analyzer (Thermo Fisher Scientific) with the LIZ-600 size standard and ABI Hi-Di™ deionized formamide. Fragment nomenclature was conducted using GeneMapper1 IDX v.1.4 (Thermo Fisher Scientific).

Statistical analysis. Haplotype and allele frequencies of Kham and Ü-Tsang Tibetans were calculated on the basis of the direct counting method. Gene diversity (GD), haplotype diversity (HD) and discrimination capacity (DC) were measured using the following formulas^{26,45}:

$$GD = \frac{N}{N-1} \left(1 - \sum P_{ai}^2\right), \quad (1)$$

$$HD = \frac{N}{N-1} \left(1 - \sum P_{hi}^2\right), \quad (2)$$

$$MP = \sum P_{ai}^2, \quad (3)$$

where N indicates the total number of samples of one population, P_{ai} indicates the allele frequency of i th allele in one locus, and P_{hi} denotes the haplotype frequency of the i th haplotype in one population. Pairwise F_{st} genetic distances between two studied Tibetan populations and 61 other worldwide populations (total of 15,880 raw genotypes) were calculated using Arlequin suite ver 3.5⁴⁶. Principal component analysis based on the raw genotype data was conducted using the convenient online tool for STR Analysis for Forensics (STRAF)⁴⁷. The samples included duplicates or triplicates, and null alleles were removed in the pairwise genetic distance evaluation. Pairwise genetic distances (Rst) and corresponding p values between our Tibetans and other adjacent relative populations with the same or different ethno-geographic origins were calculated using online tools in the YHRD database according to the online instructions in the YHRD databse (<https://yhrd.org/amova>). Finally, multidimensional scaling plot (MDS) and neighbor-joining tree were constructed on the basis of the Rst genetic matrixes using the YHRD online tool and Molecular Evolutionary Genetics Analysis Version 7.0 (Mega 7.0)⁴⁸, respectively. The map used in this study was generated by free software R 3.3.2 (<https://www.R-project.org/>)⁴⁹.

Quality control. Control DNA 007 was used as the positive control and ddH₂O (Thermo Fisher Scientific) was used as the negative control in each batch of amplification and genotyping. We performed our experiments in the laboratory in the Department of Forensic Genetics, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, which is accredited by the China National Accreditation Service for Conformity Assessment (CNAS) and participated the quality control of the YHARD database. We strictly followed the recommendations published by the DNA Commission of the International Society for Forensic Genetics (ISFG)^{50,51}. All the methods were carried out in accordance with the approved guidelines of Institute of Forensic Medicine, Sichuan University.

References

- Lu, D. *et al.* Ancestral Origins and Genetic History of Tibetan Highlanders. *Am J Hum Genet* **99**, 580–594, <https://doi.org/10.1016/j.ajhg.2016.07.002> (2016).
- Aldenderfer, M. Peopling the Tibetan plateau: insights from archaeology. *High. Alt. Med. Biol.* **12**, 141–147, <https://doi.org/10.1089/ham.2010.1094> (2011).
- Zhang, C. *et al.* Differentiated demographic histories and local adaptations between Sherpas and Tibetans. *Genome Biol* **18**, 115, <https://doi.org/10.1186/s13059-017-1242-y> (2017).
- Madsen, D. B. *et al.* The Late Upper Paleolithic occupation of the northern Tibetan Plateau margin. *J. Archaeol. Sci.* **33**, 1433–1444, <https://doi.org/10.1016/j.jas.2006.01.017> (2006).
- Shi, H. *et al.* Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol* **6**, 45, <https://doi.org/10.1186/1741-7007-6-45> (2008).
- Zhao, M. *et al.* Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci USA* **106**, 21230–21235, <https://doi.org/10.1073/pnas.0907844106> (2009).
- Qi, X. *et al.* Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the tibetan plateau. *Mol Biol Evol* **30**, 1761–1778, <https://doi.org/10.1093/molbev/mst093> (2013).
- Wang, L. X. *et al.* Reconstruction of Y-chromosome phylogeny reveals two neolithic expansions of Tibeto-Burman populations. *Mol. Genet. Genomics*, 1–8, <https://doi.org/10.1007/s00438-018-1461-2> (2018).
- Beall, C. M. *et al.* Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. *Proc Natl Acad Sci USA* **107**, 11459–11464, <https://doi.org/10.1073/pnas.1002443107> (2010).
- Bigham, A. *et al.* Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet* **6**, e1001116, <https://doi.org/10.1371/journal.pgen.1001116> (2010).

11. Simonson, T. S. *et al.* Genetic evidence for high-altitude adaptation in Tibet. *Sci* **329**, 72–75, <https://doi.org/10.1126/science.1189406> (2010).
12. Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Sci* **329**, 75–78, <https://doi.org/10.1126/science.1190371> (2010).
13. Peng, Y. *et al.* Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Mol Biol Evol* **28**, 1075–1081, <https://doi.org/10.1093/molbev/msq290> (2011).
14. Xu, S. *et al.* A genome-wide search for signals of high-altitude adaptation in Tibetans. *Mol Biol Evol* **28**, 1003–1011, <https://doi.org/10.1093/molbev/msq277> (2011).
15. Huerta-Sanchez, E. *et al.* Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197, <https://doi.org/10.1038/nature13408> (2014).
16. Xiang, K. *et al.* Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. *Mol Biol Evol* **30**, 1889–1898, <https://doi.org/10.1093/molbev/mst090> (2013).
17. Hu, H. *et al.* Evolutionary history of Tibetans inferred from whole-genome sequencing. *PLoS Genet* **13**, e1006675, <https://doi.org/10.1371/journal.pgen.1006675> (2017).
18. Beall, C. M. Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. *Integr Comp Biol* **46**, 18–24, <https://doi.org/10.1093/icb/icj004> (2006).
19. Jobling, M. A. & Tyler-Smith, C. Human Y-chromosome variation in the genome-sequencing era. *Nat Rev Genet* **18**, 485–497, <https://doi.org/10.1038/nrg.2017.36> (2017).
20. Wang, M. *et al.* Forensic characteristics and phylogenetic analysis of two Han populations from the southern coastal regions of China using 27 Y-STR loci. *Forensic Sci Int Genet* **31**, e17–e23, <https://doi.org/10.1016/j.fsigen.2017.10.009> (2017).
21. Poznik, G. D. *et al.* Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* **48**, 593–599, <https://doi.org/10.1038/ng.3559> (2016).
22. Kayser, M. Forensic use of Y-chromosome DNA: a general overview. *Hum Genet* **136**, 621–635, <https://doi.org/10.1007/s00439-017-1776-9> (2017).
23. Zhu, B. F. *et al.* Genetic diversity and haplotype structure of 24 Y-chromosomal STR in Chinese Hui ethnic group and its genetic relationships with other populations. *Electrophoresis* **35**, 1993–2000, <https://doi.org/10.1002/elps.201300574> (2014).
24. Wang, Z., Du, W., He, G., Liu, J. & Hou, Y. Forensic characteristics and phylogenetic analysis of Hubei Han population in central China using 17 Y-STR loci. *Forensic Sci Int Genet* **29**, e4–e8, <https://doi.org/10.1016/j.fsigen.2017.04.013> (2017).
25. Chen, P. *et al.* Genetic diversity and phylogenetic study of the Chinese Gelao ethnic minority via 23 Y-STR loci. *Int J Legal Med* **132**, 1093–1096, <https://doi.org/10.1007/s00414-017-1743-y> (2018).
26. He, G. *et al.* Genetic polymorphism investigation of the Chinese Yi minority using PowerPlex(R) Y23 STR amplification system. *Int J Legal Med* **131**, 663–666, <https://doi.org/10.1007/s00414-017-1537-2> (2017).
27. Ballantyne, K. N. *et al.* Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. *Hum Mutat* **35**, 1021–1032, <https://doi.org/10.1002/humu.22599> (2014).
28. Gopinath, S. *et al.* Developmental validation of the Yfiler(R) Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications. *Forensic Science International. Genetics* **24**, 164–175, <https://doi.org/10.1016/j.fsigen.2016.07.006> (2016).
29. He, G., Wang, Z., Wang, M. & Hou, Y. Genetic Diversity and Phylogenetic Differentiation of Southwestern Chinese Han: a comprehensive and comparative analysis on 21 non-CODIS STRs. *Sci Rep* **7**, 13730, <https://doi.org/10.1038/s41598-017-13190-w> (2017).
30. He, G., Wang, M., Liu, J., Hou, Y. & Wang, Z. 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, <https://doi.org/10.1007/s00414-017-1679-2> (2018).
31. He, G. *et al.* Genetic variation and forensic characterization of highland Tibetan ethnicity revealed by autosomal STR markers. *Int J Legal Med* **132**, 1097–1102, <https://doi.org/10.1007/s00414-017-1765-5> (2018).
32. He, G. *et al.* Forensic ancestry analysis in two Chinese minority populations using massively parallel sequencing of 165 ancestry-informative SNPs. *Electrophoresis* **39**, 2732–2742, <https://doi.org/10.1002/elps.201800019> (2018).
33. He, G. *et al.* 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, <https://doi.org/10.1007/s11033-018-4264-y> (2018).
34. He, G. *et al.* Genetic diversity and phylogenetic characteristics of Chinese Tibetan and Yi minority ethnic groups revealed by non-CODIS STR markers. *Sci Rep* **8**, 5895, <https://doi.org/10.1038/s41598-018-24291-5> (2018).
35. Cao, S. *et al.* Genetic portrait of 27 Y-STR loci in the Tibetan ethnic population of the Qinghai province of China. *Forensic Sci Int Genet* **34**, e18–e19, <https://doi.org/10.1016/j.fsigen.2018.02.005> (2018).
36. Zhang, J. *et al.* Haplotype data for 27 Y-chromosomal STR loci in the Chaoshan Han population, South China. *Forensic Sci Int Genet* **31**, e54–e56, <https://doi.org/10.1016/j.fsigen.2017.08.003> (2017).
37. Fan, H. *et al.* Population analysis of 27 Y-chromosomal STRs in the Li ethnic minority from Hainan province, southernmost China. *Forensic Science International. Genetics* **34**, e20–e22, <https://doi.org/10.1016/j.fsigen.2018.01.007> (2018).
38. Fan, H. *et al.* Genetic analysis of 27 Y-STR loci in Han population from Hainan province, southernmost China. *Forensic Science International. Genetics* **33**, e9–e10, <https://doi.org/10.1016/j.fsigen.2017.12.009> (2018).
39. Liu, Y. *et al.* Haplotype data of 27 Y-STRs analyzed in the Hui and Tujia ethnic minorities from China. *Forensic Sci Int Genet* **35**, e7–e9, <https://doi.org/10.1016/j.fsigen.2018.04.006> (2018).
40. Wang, J. *et al.* Haplotype structure of 27 Yfiler(R)Plus loci in Chinese Dongxiang ethnic group and its genetic relationships with other populations. *Forensic Sci Int Genet* **33**, e13–e16, <https://doi.org/10.1016/j.fsigen.2017.12.014> (2018).
41. Fan, H. *et al.* The evaluation of forensic characteristics and the phylogenetic analysis of the Ong Be language-speaking population based on Y-STR. *Forensic Science International. Genetics*, <https://doi.org/10.1016/j.fsigen.2018.09.008> (2018).
42. Zou, X. *et al.* Population Genetic Diversity and Phylogenetic Characteristics for High-Altitude Adaptive Kham Tibetan Revealed by DNATyper(TM) 19 Amplification System. *Front Genet* **9**, 630, <https://doi.org/10.3389/fgene.2018.00630> (2018).
43. Ballantyne, K. N. *et al.* Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am J Hum Genet* **87**, 341–353, <https://doi.org/10.1016/j.ajhg.2010.08.006> (2010).
44. Yuan, B., Huang, W. & Zhang, D. New evidence for human occupation of the northern Tibetan Plateau, China during the Late Pleistocene. *ChSbu* **52**, 2675–2679, <https://doi.org/10.1007/s11434-007-0357-z> (2007).
45. Nei, M. The theory of genetic distance and evolution of human races. *Jinrui idengaku zasshi. The Japanese journal of human genetics* **23**, 341–369, <https://doi.org/10.1007/BF01908190> (1978).
46. Excoffier, L. & Lischer, H. E. 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, <https://doi.org/10.1111/j.1755-0998.2010.02847.x> (2010).
47. Gouy, A. & Zieger, M. STRAF-A convenient online tool for STR data evaluation in forensic genetics. *Forensic Science International. Genetics* **30**, 148–151, <https://doi.org/10.1016/j.fsigen.2017.07.007> (2017).
48. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870–1874, <https://doi.org/10.1093/molbev/msw054> (2016).
49. Team, R. C. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <https://www.r-project.org/> (2019).

50. Scientific Working Group on DNA Analysis (SWGDM). Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Available at, https://docs.wixstatic.com/ugd/4344b4340_4350e2749756a2242528e2746285a2749755bb2749478f2749754c.pdf (2017).
51. Gusmao, L. *et al.* Revised guidelines for the publication of genetic population data. *Forensic Science International. Genetics* **30**, 160–163, <https://doi.org/10.1016/j.fsigen.2017.06.007> (2017).

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (81571854), the Open Project of Key Laboratory of Forensic Genetics in Ministry of Public Security (2017FGKFKT01) and the Fundamental Research Funds for the Central Universities (20826041A4408, YJ201651 and 2012017jjsy187). We would like to thank the volunteers who contributed samples for this study.

Author Contributions

G.H. and X.Z. wrote the manuscript, Y.S., X.C. and B.G. collected the samples, and G.H., Y.S., M.W., X.Z., J.L. and S.W. conducted the experiment and analyzed the results, Z.W. modified the manuscript, and Y.H. conceived the experiment. All authors have reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-44230-2>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019