



Genetic variation and forensic characterization of highland Tibetan ethnicity revealed by autosomal STR markers

Guanglin He¹ · Zheng Wang¹ · Yongdong Su² · Xing Zou³ · Mengge Wang¹ · Jing Liu¹ · Yiping Hou¹

Received: 28 October 2017 / Accepted: 12 December 2017 / Published online: 8 January 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Understanding the origin and genetic background of Chinese high-altitude Tibetans play a pivotal role in medical genetics, archeology, anthropology, and forensics. In this study, to investigate the forensic characterization and genetic diversity of Chinese Tibetan, allele frequencies and corresponding forensic statistical parameters of 15 autosomal STRs included in the AmpF[®]STR[®] Sinofiler[™] kit were obtained from 1220 Tibetan individuals residing in Lhasa country, Tibet Autonomous Region. We identified 191 alleles with corresponding allele frequencies varied from 0.0004 to 0.3984. The combined probability of discrimination and the combined probability of exclusion are 0.99999999999999997 and 0.9999996, respectively. Our study provided the valuable dataset for forensic individual identification and parentage testing in the high-altitude Tibetan population. In addition, comprehensive population comparisons among 30 Chinese populations via PCA, AMOVA, MDS, and N-J tree demonstrated that the genetic components of Tibet Tibetan have received gene introgression from surrounding lowland populations (Such as Gansu Hui and Yunnan Bai) and Tibetan keeps the close genetic relationship with geographic neighboring populations.

Keywords Highland Tibetan · Short tandem repeats · Genetic polymorphism · Forensic genetics

The Tibet Autonomous Region, located in the west of China, occupies part of the Chinese western region between the Xinjiang Uyghur autonomous region and Qinghai Province in the north; the Kingdom of Bhutan, Nepal, and India on the south; Kashmir and Pakistan in the west; and Sichuan and Yunnan Province to the east. Most of the Tibetan minority ethnicity group, with a total population of approximately 7.8 million, live inside China (7.5 million) according to the 2010 census. The language widely used in the Chinese Tibetan population is generally considered to be the Tibetan language

in the Tibeto-Burman languages of the Sino-Tibetan language families. These present-day Tibeto-Burman-speakers are thought to be descendants of the ancient Qiang people who have occupied the Tibetan Plateau for thousands of years, and most of them believe and practice Tibetan Buddhism.

The genetic structure of Tibetans who reside in a high-altitude plateau (average and highest elevations are 4900 and 8848 m above sea level, respectively) still remain mysterious [1]. Tibetans had a special adaptation of these extreme environments, such as arterial oxygen content and hemoglobin concentration reduction, as well as the abnormal response to hypoxic pulmonary vasoconstriction and resting ventilation, comparing with Andean high-altitude residents and non-adaptive lowland visitors [2]. Previous genetic evidence provided by Tatum et al. [3] suggested that genetic polymorphisms located in the EGLN1, EPAS1, and PPARA genes were significantly associated with the high-altitude adaptation in Tibetans. Besides, other genetic studies demonstrated that modern Tibetans had the highest rate of allele sharing with archaic hominins (especially for Denisovans) [4–6] and the estimated divergence time between Tibetan and Han was approximately 9000 to 15,000 years ago [1]. Given the demographic history and genetic background of Tibetan, more attention should be devoted to the forensic

Guanglin He and Zheng Wang contributed equally to this work.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00414-017-1765-5>) contains supplementary material, which is available to authorized users.

✉ Yiping Hou
profhou@yahoo.cn

¹ Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, Chengdu, China

² Forensic Identification Center, Public Security Bureau of Tibet Autonomous Region, Lhasa, China

³ Department of Forensic Medicine, College of Basic Medicine, Chongqing Medical University, Chongqing, China

characteristics and efficiency of mitochondrial and nuclear DNA genetic markers.

Short tandem repeats (STRs) widely distributed in the entire genome with high discriminatory power are used routinely in forensic individual identification and parentage testing due to the rapid technological revolution of capillary electrophoresis (CE)-based and massively parallel sequencing (MPS)-based STR genotyping [7]. Unfortunately, genetic polymorphisms and forensic characteristics of forensic commonly used STR loci in the high-altitude Tibetan population residing in Tibet Autonomous Region as well as the genetic relationships between the Tibetan population and Chinese neighboring populations remain uncharacterized.

In continuation of our previous work focused on East Asian populations [8–12], a total of 1220 bloodstain samples (917 males and 303 females) were collected from Lhasa country in the Tibet Autonomous Region (Fig. S1). Our study had been approved by the institutional review boards at the Sichuan University. All participants had ancestors residing in this region at least three generations and signed informed consents before sample collection. The PureLink Genomic DNA Mini Kit (Life Technologies, Invitrogen, USA) was used to isolate genomic DNA based on the manufacturer's instructions. DNA concentration and purification were measured using a Nanodrop-2000C (Thermo Fisher Scientific). We employed AmpF Λ STR $\text{\textcircled{R}}$ Sinofiler TM PCR Amplification Kit (Applied Biosystems, Foster City, CA, USA) comprising of 15 autosomal STRs (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, D5S818, D13S317, D16S539, D2S1338, D19S433, vWA, D12S391, D18S51, D6S1043, and FGA) plus a sex-determining locus Amelogenin for STRs genotyping on a ProFlex TM PCR System (Applied Biosystems) following the manufacturer's specification. The amplified products were isolated and detected using capillary electrophoresis in the ABI 3130 genetic analyzer (Applied Biosystems) based on corresponding recommendations. Allele identification was conducted on the GeneMapperID software v3.2. Our study with the positive control (9947A) and reagent blank in each batch of genotyping was conducted in the Forensic Genetic Department, Sichuan University, which was a laboratory that passed the international accreditation of Y-chromosome STR haplotype reference database (YHRD, <https://yhrd.org/>) and ISO 17025, as well as the China National Accreditation Service for Conformity Assessment (CNAS).

The allele frequencies and corresponding statistic parameters of forensic interest (probability of discrimination (PD); probability of exclusion (PE); typical paternity index (TPI); random match probability (PM)) were estimated using the modified PowerStates (version 1.2) spreadsheet (Promega, Madison, WI, USA). In addition, the observed heterozygosity (He), expected heterozygosity (Ho), Hardy-Weinberg equilibrium, and Linkage Disequilibrium testing were assessed employing the Arlequin version 3.5.2 [13]. Subsequently,

population relationships between the Tibet Tibetan population and 29 reference populations were investigated. Pairwise F_{st} and corresponding p values were calculated by Arlequin version 3.5.2 [13] based on Locus-by-Locus allele frequency distributions. We also performed a principal component analysis (PCA) using the Multivariate Statistical Package (MVSP) version 3.22 software [14] and calculated pairwise Nei's genetic distances via PHYLIP ver. 3.5 packages [15] based on allele frequency distributions. Multidimensional Scaling Plots (MDS) were performed with IBM SPSS Statistics version 21 (SPSS, Chicago, IL, USA). A Neighbor-Joining tree based on the genetic distance matrix was constructed with Mega 7.0 [16].

Table S1 presents the results of Linkage Disequilibrium analysis. No linkage inheritance of 105 pairwise groups was identified in the Tibetan population after Bonferroni correction ($p < 0.05/105 = 0.0005$). As shown in Table S2, all STRs are in accordance with the Hardy-Weinberg equilibrium with the exception of D6S1043 ($p = 0.0003$) after Bonferroni adjustment. The allele frequencies and forensic parameters of 15 autosomal STRs in the Tibet Tibetan population are presented in Table S2 and Fig. S2. A total of 191 alleles were identified with corresponding allele frequencies varied from 0.0004 to 0.3984. The PD, PE, and TPI spanned from 0.8620 (D3S1358) to 0.9726 (D6S1043), 0.4219 (CSF1PO) to 0.7291 (FGA), and 1.6442 (CSF1PO) to 3.7654 (FGA), respectively. The H_o varied from 0.6959 at locus of CSF1PO to 0.8672 at locus of FGA with a mean of 0.8073 ± 0.0490 . The H_e spanned from 0.7142 (D3S1358) to 0.8781 (D6S1043) with a mean of 0.8118 ± 0.0504 . The combined probability of discrimination and the combined probability of exclusion are 0.999999999999999997 and 0.9999996, respectively. Our findings suggested that the investigated 15 autosomal STR loci are informative and valuable in the forensic human identification, disaster victim identification, and kinship analysis.

Analysis of molecular of variance (AMOVA) between the Tibet Tibetan and 29 Chinese neighboring populations was carried out. The F_{st} and corresponding p -values are presented in Table S3. There are significant differences between the Tibet Tibetan and Xinjiang Uyghur1 at nine Loci; Xiamen Han at seven loci; Hubei Han at six loci; Sichuan Han, Chongqing Han, Yunnan Han, and Yungui Han at five loci; Yunnan Yi and Yunnan Zhuang at three loci; Shanxi Han, Xinjiang Uyghur2, and Xinjiang Uyghur3, and Yili Xibe at one locus. To further characterize the population structure across the Tibetan and 29 neighboring mainland populations, we subsequently performed the principal component analysis based on allele frequency distributions. As presented in Fig. 1 and Fig. S3 and S4, the first three components accounted for 53.163% of total variances: the first (PCA1), second (PCA2), and third (PCA3)

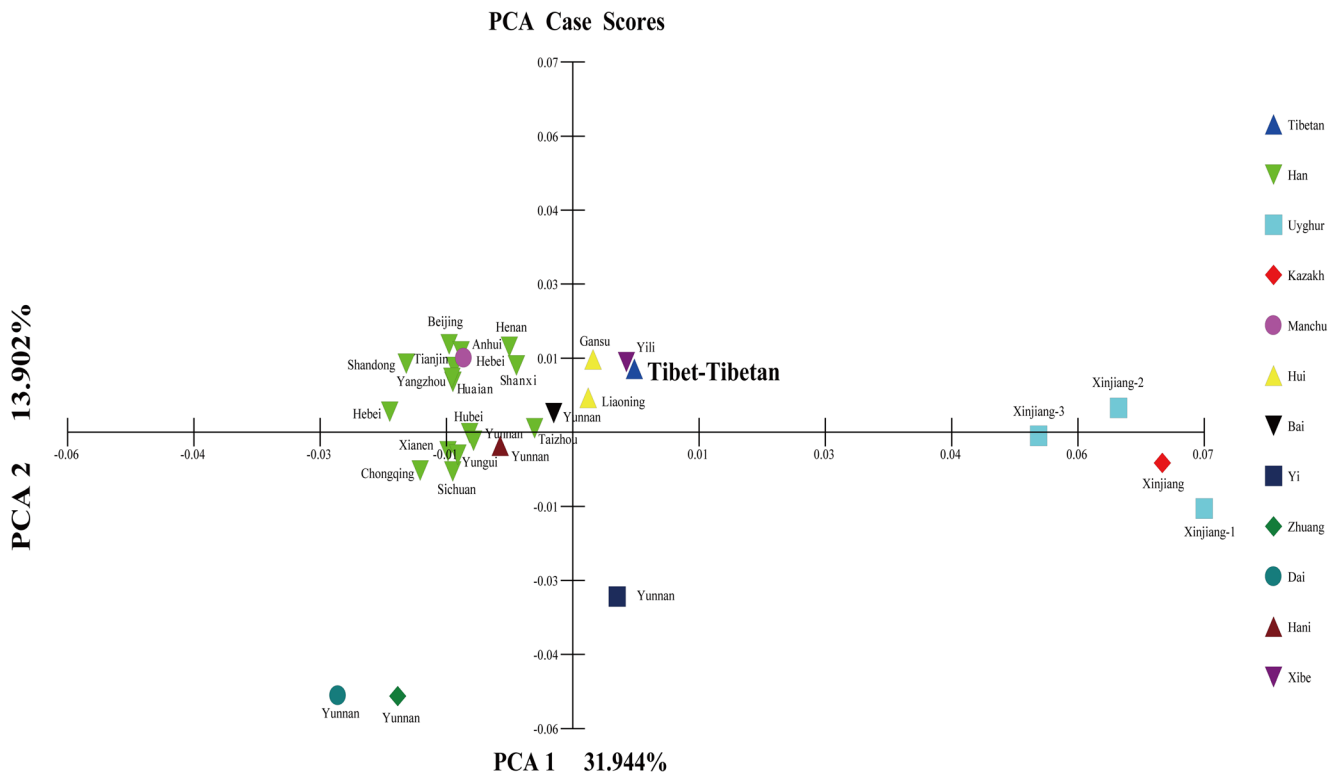


Fig. 1 The principal component analysis (PCA) of highland Tibetan and 29 reference populations based on PCA1 and PCA2

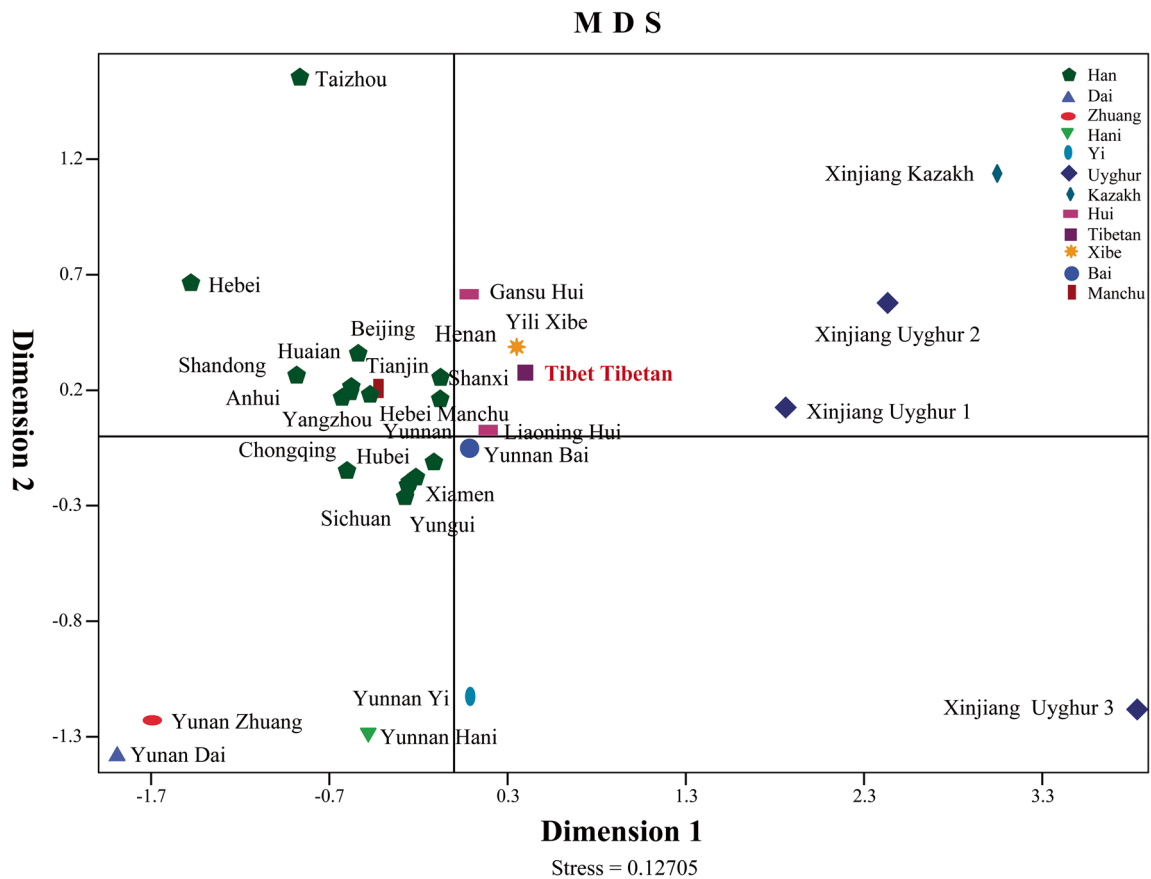
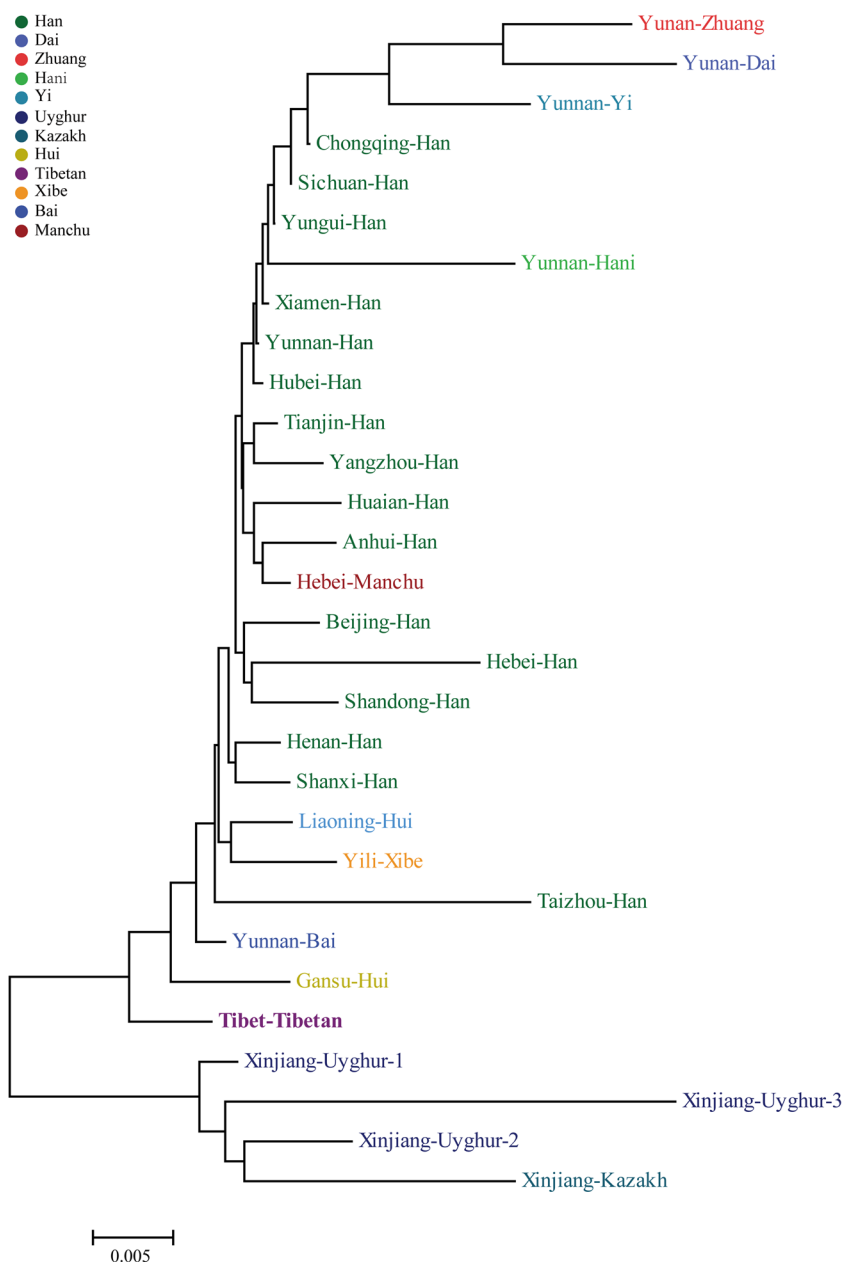


Fig. 2 Multidimensional scaling analysis (MDS) based on autosomal STR pairwise Nei's genetic distances estimates among Tibet Tibetan and 29 Chinese reference populations

Fig. 3 A Neighboring-Joining tree on the basis of Nei's genetic distances of 30 Chinese populations, with colors corresponding to the same ethnic group origin



components respectively accounted for 31.944, 13.902, and 7.317%. The PCA1 could successfully separate three Xinjiang Uyghurs and one Kazakh group from other populations (Fig. 1 and Fig. S3) and PCA2 could differentiate three Yunnan minority groups (Yi, Zhuang, and Dai) from other groups (Fig. 1 and Fig. S4). PCA3 showed a clear differentiation between Taizhou Han, Xinjiang Uyghur2, Uyghur3, and Kazakh from other reference populations (Fig. S3 and Fig. S4). Tibet Tibetan keeps a genetic affinity with Bai (Yunnan), Xibe (Yili), and Hui (Gansu and Liaoning).

A comparison of genetic similarities and differences measured by pairwise Nei's genetic distances between Tibet Tibetan and 29 Chinese mainland populations

(Table S4 and Fig. S5) showed that the largest genetic distance was observed between the Tibet Tibetan and Xinjiang Uyghur3 (0.0544), followed by Xinjiang Kazakh (0.0434), and Tibet Tibetan had a closer relationship with Yunnan Bai (0.0094). Multidimensional scaling analysis based on the aforementioned Nei's genetic matrix was conducted to further explore the genetic background of Tibetans. Figure 2 displayed the MDS results which showed that four Xinjiang minority ethnic groups (one Kazakh and three Uyghurs) and four Yunnan minority ethnic groups (Zhuang, Dai, Yi, and Hani) as well as Taizhou Han keep a relatively isolated with other reference populations. These relationship distributions remained consistent with analytical results

revealed by PCA (Fig. 1) and our previous investigations on the basis of X and Y-chromosome genetic markers [8, 11, 17, 18]. We also are concerned with the genetic history and the phylogenetic relationship between the Tibetan and other Chinese populations and constructed a Neighbor-Joining tree (Fig. 3). Among these highland/lowland or Han Chinese/minority ethnicity populations, two clusters were grouped: one cluster consisting of four Xinjiang minority ethnic groups and the other groups clustered together. Our studied population was clustered with Gansu Hui and then clustered with the branch grouped by Xinjiang ethnic groups. In general, genetic differentiation among different ethnic groups or one ethnicity distributed in distinct administrative divisions is not significantly obvious as our inference according to the history records (geographical and ethnic isolation) due to large-scale human migration under the promotion of starvation and warfare, random marriage, and gene introgression among different populations. At the same time, we will be cautioned that some minority ethnic groups, such as Chinese Muslim populations (Uyghur and Kazakh) and High-latitude adaptive groups (Tibetan) had their distinctive genetic signature under the influence of self-promoting culture and extreme environment.

In summary, population data and forensic characteristics of 15 autosomal STRs in a highland Tibet Tibetan population were obtained and analyzed. Our results demonstrated that the investigated genetic markers are highly polymorphic and informative in the Tibetan population and can be integrated into the new panel of the NGS-based genotyping system in the future forensic high-throughput practice. Genetic background of the Tibetan population revealed by MDS, AMOVA, MDS, and N-J tree illustrated that this high-latitude population has remained with lots of adaptive genetic signatures via positive natural selection under an extreme natural environment and keeps the close genetic relationship with the geographic and ethnic adjacent populations, such as Gansu Hui and Yunnan Bai populations. What is more, significant genetic differences were identified between the highland Tibetan and Xinjiang Uyghur, Kazakh, and other lower-latitude populations.

Acknowledgments This work received the support of grants from the National Key R&D Program of China (2016YFC0800703) and the National Natural Science Foundation of China (81330073 and 81501635) and the Fundamental Research Funds for the Central Universities.

Authors' contribution G.H. and X.Z. wrote the manuscript; Y.S. collected the samples; G.H., X.Z., M.W., and J.L. conducted the experiment and assessed the results; Z.W. modified the manuscript; and Y.H. conceived the experiment. All authors have reviewed and considered the manuscript.

Compliance with ethical standards Our study had been approved by the institutional review boards at the Sichuan University.

Competing interests The authors declare that they have no competing interests.

References

1. Lu D, Lou H, Yuan K, Wang X, Wang Y, Zhang C, Lu Y, Yang X, Deng L, Zhou Y, Feng Q, Hu Y, Ding Q, Yang Y, Li S, Jin L, Guan Y, Su B, Kang L, Xu S (2016) Ancestral origins and genetic history of Tibetan highlanders. *Am J Hum Genet* 99(3):580–594. <https://doi.org/10.1016/j.ajhg.2016.07.002>
2. Lorenzo FR, Huff C, Myllymaki M et al (2014) A genetic mechanism for Tibetan high-altitude adaptation. *Nat Genet* 46(9):951–956. <https://doi.org/10.1038/ng.3067>
3. Simonson TS, Huff CD, Witherspoon DJ, Prchal JT, Jorde LB (2015) Adaptive genetic changes related to haemoglobin concentration in native high-altitude Tibetans. *Exp Physiol* 100(11):1263–1268. <https://doi.org/10.1113/EP085035>
4. Jeong C, Alkorta-Aranburu G, Basnyat B, Neupane M, Witonsky DB, Pritchard JK, Beall CM, di Rienzo A (2014) Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat Commun* 5: 3281. <https://doi.org/10.1038/ncomms4281>
5. Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E (2015) Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* 16(6):359–371. <https://doi.org/10.1038/nrg3936>
6. Huerta-Sanchez E, Jin X, Asan et al (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512(7513):194–197. <https://doi.org/10.1038/nature13408>
7. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet* 12(3):179–192. <https://doi.org/10.1038/nrg2952>
8. He G, Li Y, Zou X, Zhang Y, Li H, Wang M, Wu J (2017) X-chromosomal STR based genetic structure of Sichuan Tibetan minority ethnicity group and its relationships to various groups. *Int J Legal Med.* <https://doi.org/10.1007/s00414-017-1672-9>
9. He G, Li Y, Wang Z, Liang W, Luo H, Liao M, Zhang J, Yan J, Li Y, Hou Y, Wu J (2017) Genetic diversity of 21 autosomal STR loci in the Han population from Sichuan province, Southwest China. *Forensic Sci Int Genet* 31:e33–e35. <https://doi.org/10.1016/j.fsigen.2017.07.006>
10. He G, Wang Z, Wang M, Hou Y (2017) Genetic diversity and phylogenetic differentiation of southwestern Chinese Han: a comprehensive and comparative analysis on 21 non-CODIS STRs. *Sci Rep* 7(1):13730. <https://doi.org/10.1038/s41598-017-13190-w>
11. He G, Chen P, Zou X et al (2017) Genetic polymorphism investigation of the Chinese Yi minority using PowerPlex(R) Y23 STR amplification system. *Int J Legal Med* 131(3):663–666. <https://doi.org/10.1007/s00414-017-1537-2>
12. He G, Wang M, Liu J, Hou Y, Wang Z (2017) Forensic features and phylogenetic analyses of Sichuan Han population via 23 autosomal STR loci included in the Huaxia Platinum System. *Int J Legal Med.* <https://doi.org/10.1007/s00414-017-1679-2>
13. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
14. Kovach WL (2007) MVSP—a MultiVariate statistical package for Windows, ver. 3.1. Kovach Computing Services, Pentraeth

15. Cummings MP (2004) PHYLIP (phylogeny inference package). Wiley, Hoboken
16. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7):1870–1874. <https://doi.org/10.1093/molbev/msw054>
17. He G, Li Y, Zou X et al (2017) Forensic characteristics and phylogenetic analyses of the Chinese Yi population via 19 X-chromosomal STR loci. *Int J Legal Med* 131(5):1243–1246. <https://doi.org/10.1007/s00414-017-1563-0>
18. He G, Li Y, Zou X, Wang M, Chen P, Liao M, Wu J (2017) Genetic polymorphisms for 19 X-STR loci of Sichuan Han ethnicity and its comparison with Chinese populations. *Leg Med (Tokyo)* 29:6–12. <https://doi.org/10.1016/j.legalmed.2017.09.001>