

# SCIENTIFIC REPORTS



OPEN

## Genetic structure of Tibetan populations in Gansu revealed by forensic STR loci

Received: 03 November 2016  
Accepted: 15 December 2016  
Published: 23 January 2017

Hong-Bing Yao<sup>1,\*</sup>, Chuan-Chao Wang<sup>2,3,\*</sup>, Jiang Wang<sup>4</sup>, Xiaolan Tao<sup>1</sup>, Lei Shang<sup>5</sup>, Shao-Qing Wen<sup>2</sup>, Qiajun Du<sup>6</sup>, Qiongying Deng<sup>7</sup>, Bingying Xu<sup>8</sup>, Ying Huang<sup>8</sup>, Hong-Dan Wang<sup>9</sup>, Shujin Li<sup>10</sup>, Bin Cong<sup>10</sup>, Liying Ma<sup>1</sup>, Li Jin<sup>2,11</sup>, Johannes Krause<sup>3</sup> & Hui Li<sup>2</sup>

The origin and diversification of Sino-Tibetan speaking populations have been long-standing hot debates. However, the limited genetic information of Tibetan populations keeps this topic far from clear. In the present study, we genotyped 15 forensic autosomal short tandem repeats (STRs) from 803 unrelated Tibetan individuals from Gansu Province (635 from Gannan and 168 from Tianzhu) in northwest China. We combined these data with published dataset to infer a detailed population affinities and genetic substructure of Sino-Tibetan populations. Our results revealed Tibetan populations in Gannan and Tianzhu are genetically very similar with Tibetans from other regions. The Tibetans in Tianzhu have received more genetic influence from surrounding lowland populations. The genetic structure of Sino-Tibetan populations was strongly correlated with linguistic affiliations. Although the among-population variances are relatively small, the genetic components for Tibetan, Lolo-Burmese, and Han Chinese were quite distinctive, especially for the Deng, Nu, and Derung of Lolo-Burmese. Han Chinese but not Tibetans are suggested to share substantial genetic component with southern natives, such as Tai-Kadai and Hmong-Mien speaking populations, and with other lowland East Asian populations, which implies there might be extensive gene flow between those lowland groups and Han Chinese after Han Chinese were separated from Tibetans. The dataset generated in present study is also valuable for forensic identification and paternity tests in China.

The Sino-Tibetan languages, spoken by over a billion people all over East Asia and Southeast Asia, have been classified into two subfamilies, namely Chinese and Tibeto-Burman<sup>1</sup>. The linguistic connection between Chinese and Tibeto-Burman are well established. Chinese was suggested to split away from Tibeto-Burman around 6 thousand years ago (kya) based on lexical evidence<sup>2</sup>.

During the past two decades, genetic evidence, especially from the maternal mitochondrial DNA (mtDNA) and the paternal Y chromosome, has shed more light on the history of Sino-Tibetan populations. MtDNA evidence reveals a northern Asian origin of Tibetans, due to the high frequencies of northern Asian specific

<sup>1</sup>Key Laboratory of Evidence Science of Gansu Province, Gansu Institute of Political Science and Law, Lanzhou 730070, China. <sup>2</sup>State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai, 200433, China. <sup>3</sup>Department of Archaeogenetics and Eurasia3angle research group, Max Planck Institute for the Science of Human History, Kahlaische Straße 10, 07745 Jena, Germany. <sup>4</sup>College of Animal Sciences and Veterinary Medicine, Henan Agricultural University, Zhengzhou 450002, Henan Province, China. <sup>5</sup>Key Laboratory of Forensic Genetics, Institute of Forensic Science, Ministry of Public Security, Beijing, 100038, China. <sup>6</sup>Lanzhou University Second Hospital Clinical Laboratory, Lanzhou 730000, Gansu Province, China. <sup>7</sup>Department of Anatomy, Guangxi Medical University, Nanning 530021, China. <sup>8</sup>School of Forensic Medicine, Kunming Medical University, Kunming, 650500, China. <sup>9</sup>Medical Genetic Institute of Henan Province, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University, Zhengzhou, China. <sup>10</sup>Hebei Key Laboratory of Forensic Medicine, Department of Forensic Medicine, Hebei Medical University, Shijiazhuang, 050017, China. <sup>11</sup>CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 200031 Shanghai, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to H.-B.Y. (email: yaohongb@126.com) or C.-C.W. (email: wang@shh.mpg.de) or H.L. (email: LHCA@Fudan.edu.cn)

haplogroup A, D, G, and M8<sup>3–5</sup>. The genetic relics of the Late Paleolithic ancestors of Tibeto-Burman populations have also been reported, such as haplogroup M62<sup>5</sup>. Y chromosome suggested Tibeto-Burman populations are an admixture of the northward migrations of East Asian initial settlers with haplogroup D-M175 in the Late Paleolithic age, and the southward Di-Qiang people with dominant haplogroup O3a2c1\*-M134 and O3a2c1a-M117 in the Neolithic Age<sup>6–8</sup>. Haplogroup O3a2c1\*-M134 and O3a2c1a-M117 are also characteristic lineages of Han Chinese, comprising 11.4% and 16.3%, respectively<sup>9,10</sup>. However, another dominant paternal lineage of Han Chinese, haplogroup O3a1c-002611, is found at very low frequencies in Tibeto-Burman populations, suggesting this lineage might not have participated in the formation of Tibeto-Burman populations<sup>6,9–11</sup>. Sex-biased admixture has also been observed during the formation of Tibeto-Burman populations. Southern Tibeto-Burman populations exhibit a stronger influence of northern immigrants on the paternal lineages and a more extensive contribution of southern natives to the maternal lineages<sup>12</sup>. Likewise, the southern natives have made a greater contribution to the maternal lineages of southern Han Chinese<sup>13</sup>. Tibeto-Burman populations tend to cluster with North Asian and Tai-Kadai populations rather than Han Chinese based on the frequency data of 15 autosomal short tandem repeats (STRs)<sup>14</sup>. A genome-wide study of PanAsia SNP project reveals that Han Chinese populations show varying degrees of admixture between a northern Altaic cluster and a Sino-Tibetan/Tai-Kadai cluster<sup>15</sup>. But Tibetan populations were not included in the PanAsia project. The analyses of more than 30 deeply sequenced genomes of Tibetans in Tibet Autonomous Region give consistent results with Y chromosomes that most of the Tibetan gene pool diverged from that of Han Chinese about 15 kya to 9 kya. The shared ancestry of Tibetan-enriched sequences dates back to 62–38 kya, representing Paleolithic colonization of the plateau<sup>16</sup>. An ancient DNA-based study using ancient Nepalese genomes of the Chokhopani, Mebrak and Samdzong sites spanning 3 to 1 kya demonstrates that the Tibetan Plateau experienced millennia of genetic continuity which continues until the present day<sup>17</sup>.

From previous studies, the origin of Sino-Tibetan populations seems to involve substantial genetic admixture with surrounding populations. However, the limited markers of mtDNA and Y chromosome and small sample sizes and insufficient sampling of genome-wide study are far from enough to give a comprehensive understanding about the genetic history and admixture process of Sino-Tibetan populations. In addition, Tibetan populations of Gansu province, the key area for the diversification of Amdo Tibetans, have seldom been studied genetically. Therefore, we analyze 15 autosomal STRs in 635 and 168 unrelated individuals from two Tibetan populations in Gannan and Tianzhu of Gansu province to explore the genetic structure of Tibetan populations in northwest China and to test population affinities and the level of admixture of Sino-Tibetan populations with surrounding populations.

## Methods

We collected blood samples of 635 and 168 unrelated individuals from two Tibetan populations in Gannan and Tianzhu, Gansu province. Our study was approved by the Ethnic Committee of Gansu Institute of Political Science and Law. The study was conducted in accordance with the human and ethical research principles of Gansu Institute of Political Science and Law. All individuals were adequately informed and signed their informed consent before their participation. For each sample, genomic DNA was extracted according to the Chelex-100 method and proteinase K protocol<sup>18</sup>. The 15 most widely used forensic loci were amplified simultaneously using AmpFISTR Sinofiler PCR Amplification Kit (Applied Biosystems, Foster City, CA, USA) at the D8S1179, D21S11, D7S820, CSF1PO, D3S1358, D13S317, D16S539, D2S1338, D19S433, vWA, D18S51, D5S818, FGA, D6S1043 and D12S391 STR loci. The PCR products were analyzed with the 3500XL DNA Genetic Analyzer and Genemapper ID-X software (Applied Biosystems, Foster City, CA, USA).

Allele frequency, heterozygosity, polymorphism information content (PIC), discrimination power (DP), probability of paternity exclusion (PPE) were calculated using PowerStatsV12 (<http://www.promega.com/>). Tests for Hardy–Weinberg equilibrium were performed in Arlequin v3.5.1.3<sup>19</sup>. Since the statistical analyses in this study were on the basis of Bayesian-clustering algorithm, raw genotypic data of 13 STRs (excluding D6S1043 and D12S391) from 59 populations all around the world were extracted to determine population affinity<sup>14,20–50</sup>. Analysis of molecular variance (AMOVA), average number of pairwise differences, pairwise *F*<sub>st</sub>, Slatkins linearized *F*<sub>st</sub>, and coancestry coefficients were all calculated in Arlequin v3.5.1.3<sup>19</sup> using genotype data. The detailed population genetic structure was performed using model-based clustering method implemented in Structure 2.3.4<sup>51,52</sup> under assumptions of admixture, LOCPRIOR model, and correlated allele frequencies. Each run used 100,000 estimation iterations for *K* = 2 to 12 after a 20,000 burn-in length with several replicates. Posterior probabilities for each *K* were computed for each set of runs. Graphical display for Matrix plot of genetic distance and population structure were carried out in R statistical software v3.0.2<sup>53</sup> and Distruct v1.1<sup>54</sup>.

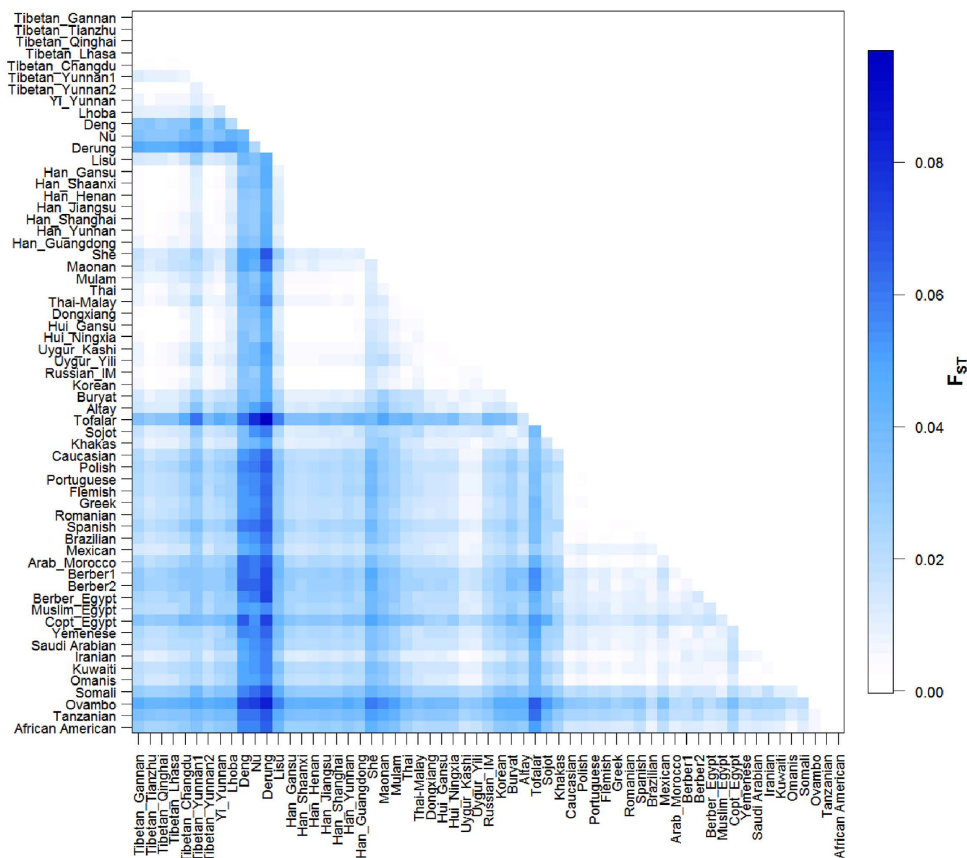
## Results

**Forensic parameter analysis.** Fifteen STR loci were genotyped in two populations sampled from Gannan and Tianzhu of Gansu province and their allele frequencies along with a number of genetic and forensic parameters of interest are provided in Supplementary Table 1 and 2. No significant deviation was observed for Hardy–Weinberg equilibrium tests, indicating that our samples well represent the populations. The loci in both populations were highly discriminating with DP ranging from 0.852 to 0.974, demonstrating that those loci are useful for forensic identification.

**Interpopulation genetic distances.** We performed various parameters of genetic diversity and distances to infer population structure between Tibetans in Gannan and Tianzhu, as well as compared them with previously studied populations. The Tibetans in Gannan and Tianzhu fall into the general profile of Tibetan groups, showing extremely small genetic distances with other Tibetan populations. The within-population component of genetic variation, estimated here as 99.14% (Table 1), accounts for most of genetic diversity of the 20 Sino-Tibetan populations. The small among-population and among-group variance components support the genetic affinity among

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among group	2	175.190	0.01925	0.37
Among populations within groups	17	306.945	0.02549	0.49
Within populations	11204	58131.8	5.18848	99.14
Total	11223	58613.9	5.23323	

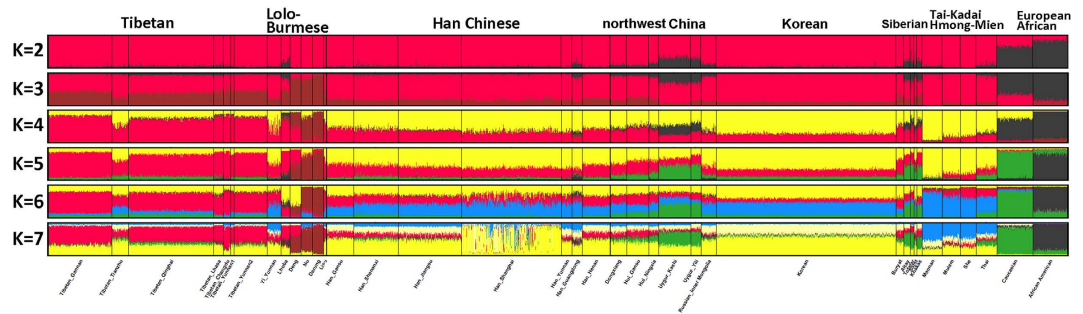
**Table 1.** AMOVA results for 13 autosomal STRs at population and group scales. 20 Sino-Tibetan populations have been classified into 3 groups: Han Chinese, Tibetan, and Lolo-Burmese.



**Figure 1.** Plots of pairwise  $F_{ST}$  of Tibetan in Gannan, Tianzhu and other 59 worldwide populations.

the Sino-Tibetan populations. The pairwise  $F_{ST}$  comparisons in Fig. 1 also confirm the genetic similarity between Tibetan and Han Chinese populations, with almost all the values below 0.01. However, Deng, Nu, and Derung seem like to be outliers of the Sino-Tibetan profile due to almost all the  $F_{ST}$  values between each of them with other Sino-Tibetan populations are above 0.03. The  $F_{ST}$  between other Lolo-Burmese populations with Tibetan or Han Chinese were also slightly higher than the values between Tibetan and Han Chinese. Tibetan and Han Chinese also show close genetic relationship with Russian samples collected from Inner Mongolia in north China and Korean in East Asia, but not with the population in South Siberia, such as Buryat, Altay, Tofalar, Sojot, and Khakas. The genetic distances between Tibetan and Han Chinese with Tai-Kadai speaking (Maonan, Mulam, and Thai) or Hmong-Mien speaking (She) populations were also larger than those between Tibetan and Han Chinese. The Muslim populations in northwest China exhibited relatively small genetic distances with Tibetan and Han Chinese, revealing the substantial gene flow from Sino-Tibetan populations into Muslim people during their Islamization. Average number of pairwise differences, Slatkins linearized  $F_{ST}$ , and coancestry coefficients also reveal a very similar pattern with the pairwise  $F_{ST}$  (Supplementary Table 3).

**Clustering by structure analysis.** We then applied a model based clustering algorithm in Structure to infer the detailed genetic ancestry at individual level. This approach will place individuals into  $K$  clusters, where  $K$  is set in advance but can be varied. The results for  $K = 2$  to 7 are shown in Fig. 2 and Supplementary Table 4. At  $K = 2$ , a clear distinction is observed between the present-day Europeans and Africans with populations from Asia. At  $K = 3$ , a component maximized in present-day Lolo-Burmese populations separates especially Deng, Nu, and



**Figure 2. Estimated population genetic structure of Tibetan in Gannan and Tianzhu and other 36 worldwide populations.** Note that the bold names “Tibetan”, “Lolo-Burmese”, “Han Chinese”, “northwest China”, “Korean”, “Siberian”, “Tai-Kadai Hmong-Mien”, “European”, and “African” refer to the group classifications of present-day populations based on language and geographic affinity. Those names are not the labels for the inferred ancestral population in Structure analysis.

Derung from other Asian populations. At  $K = 4$ , a component maximized in present-day Tibetan populations appears, which also comprises about 20–50% of Han Chinese and southern Siberian populations, but is greatly reduced in present-day Maonan samples. At  $K = 5$ , Europeans are separated from Africans. The present-day Uyghur and southern Siberian populations seem to have half of this European maximized component. The next cluster at  $K = 6$  corresponds to present-day Tai-Kadai and Hmong-Mien speaking populations in south China with a component maximized in those southern native groups. This southern native component comprises almost half of the Han Chinese and Siberian gene pool. At  $K = 7$ , the supposed southern native genetic component in present-day Han Chinese and Siberian populations forms a new cluster. It's quite clearly from the Structure analysis that Han Chinese, Korean and populations in northwest China share the similar membership, but the Siberian groups share more genetic component with Europeans. The origin of Tibetan populations seemed to involve extensive gene flow with Han Chinese. Southern natives, such as Tai-Kadai and Hmong-Mien speaking groups, share substantial genetic component with Han Chinese, Muslim populations in northwest China, and Yi in Yunnan, but not with Tibetan populations. The Tibetans in Gannan of Gansu province are genetically very similar with Tibetans in Tibet, Qinghai and Yunnan, but Tibetans in Tianzhu of Gansu province seem to have more lowland East Asian genetic components than other Tibetan populations have.

## Discussion

The origin and diversification of Sino-Tibetan populations have become long-standing hot topics among linguists, population geneticists, anthropologists, and archaeologists. However, the limited genetic information of Tibetan populations in northwest China has made this topic far from clear. Here, we genotyped 15 forensic autosomal STRs from 635 and 168 unrelated Tibetan individuals from Gannan and Tianzhu of Gansu province, together with published forensic dataset to infer a detailed genetic structure of Sino-Tibetan populations. The Tibetans in Gannan share a very similar genetic makeup with other Tibetan populations from Tibet, Qinghai, and Yunnan. While Tibetans in Tianzhu County seem like to share more genetic component with lowland East Asians, such as Han Chinese, Muslim and Korean populations, which is understandable as those Tibetans are surrounded by Han Chinese and Chinese Muslims. The genetic structure of studied Sino-Tibetan populations is strongly correlated with linguistic affiliations, as we can detect three distinctive genetic components for Tibetan, Lolo-Burmese, and Han Chinese although the among-population variances are relatively small. The Yi of Yunnan province, one of Lolo-Burmese speaking populations, is found out to be an admixture between Tibetan, Han Chinese, and southern natives (Tai-Kadai and Hmong-Mien speaking groups). However, other Lolo-Burmese populations, such as Deng, Nu, and Derung, form a distinctive cluster, which is probably due to long-term isolations and genetic drift as those populations are all small and living on hunting and gathering<sup>12</sup>.

Previous studies, especially using mtDNA and Y chromosome, had suggested the North Asian origin of Tibetan populations<sup>55,56</sup>. Our results show that the Tibetans are quite distinctive from Siberian populations. The Siberian populations, such as Buryat, Altay, Tofalar, Sojot, and Khakas, share substantial genetic components with European groups which are rarely seen in Tibetan populations. The results are consistent with genome-wide evidence that there is no significant gene flow from West Eurasians into Tibetans<sup>14,15</sup>. We suspect that the proposed northern ancestral group that led to present-day Tibetan populations was probably separated with the lineage that later became the East Asian part of the Siberian groups earlier before the Siberian groups were extensively admixed with West Eurasian lineages. We caution that the geographical distribution of past populations is probably not accurately reflected in present-day distributions. An important direction for future work is to work out the exact phylogenetic relationship of the proposed ancient population branch leading to present-day Tibetan populations to other extant Eurasian groups by sequencing ancient samples from Tibetan Plateau and the Upper-Middle Yellow River Basin.

The genetic makeups of the Tai-Kadai and Hmong-Mien speaking populations in south China are similar with Han Chinese rather than with Tibetan groups. The Muslim populations in northwest China and Korean people also exhibit the similar component cluster pattern. The possible scenario for this observation is that those lowland populations might have extensive gene exchange with Han Chinese after Han Chinese were separated from Tibetans. The autosomal STR results are consistent with uniparental Y chromosomal and mtDNA evidence that southern natives made a greater contribution to the maternal lineages of southern Han Chinese<sup>11</sup>.

## References

- Matisoff, J. A. Sino-Tibetan linguistics: present state and future prospects. *Annu Rev Anthropol* **20**, 469–504 (1991).
- Wang, W. S. Y. In the Bronze Age and Early Iron Age peoples of Eastern Central Asia. *University of Pennsylvania Museum Publications*. 508–534 (1998).
- 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 (2013).
- Qin, Z. *et al.* A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *Am J Phys Anthropol* **143**, 555–569 (2010).
- Zhao, M. *et al.* Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proc Natl Acad Sci USA* **106**, 21230–21235 (2009).
- Wang, C. C. *et al.* Genetic structure of Qiangic populations residing in the western Sichuan corridor. *PLoS One* **9**, e103772 (2014).
- Kang, L. *et al.* Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the eastern Himalayas. *Ann Hum Genet* **76**, 92–99 (2012).
- Kang, L. *et al.* Northward genetic penetration across the Himalayas viewed from Sherpa people. *Mitochondr DNA* **27**, 342–349 (2016).
- Yan, S., Wang, C. C., Li, H., Li, S. L. & Jin, L. An updated tree of Y chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *Eur J Hum Genet* **19**, 1013–1015 (2011).
- Yan, S. *et al.* Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLoS One* **9**, e105691 (2014).
- Wang, C. C. *et al.* Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611. *J Syst Evol* **51**, 280–286 (2013).
- Wen, B. *et al.* Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet* **74**, 856–865 (2004).
- Wen, B. *et al.* Genetic evidence supports demic diffusion of Han culture. *Nature* **431**, 302–305 (2004).
- Kang, L. *et al.* Genetic structures of the Tibetans and the Deng people in the Himalayas viewed from autosomal STRs. *J Hum Genet* **55**, 270–277 (2010).
- HUGO Pan-Asian, S. N. P. Consortium. Mapping human genetic diversity in Asia. *Science* **326**, 1541–1545 (2009).
- Lu, D. *et al.* Ancestral Origins and Genetic History of Tibetan Highlanders. *Am J Hum Genet* **99**, 580–594 (2016).
- Jeong, C. *et al.* Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc Natl Acad Sci USA* **113**, 7485–90 (2016).
- Zhou, Y. Q., Zhu, W., Liu, Z. P. & Wu, W. Q. A quick method of extraction of DNA by Chelex-100 from trace bloodstains. *Fudan Univ J Med Sci* **30**, 379–380 (2003).
- Excoffier, L., Laval, G. & Schneider, S. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**, 47–50 (2005).
- Yan, J. *et al.* Genetic analysis of 15 STR loci on Chinese Tibetan in Qinghai Province. *Forensic Sci Int* **169**, e3–6 (2007).
- Wang, H. D. *et al.* Genetic polymorphisms of 20 short tandem repeat loci from the Han population in Henan, China. *Electrophoresis* **35**, 1509–1514 (2014).
- Zhai, D. *et al.* The allele frequency of 15 STRs among three Tibeto-Burman-speaking populations from the southwest region of mainland China. *Forensic Sci Int Genet* **13**, e22–4 (2014).
- Deng, Q. *et al.* Genetic relationships among four minorities in Guangxi revealed by analysis of 15 STRs. *J Genet Genomics* **34**(12), 1072–9 (2007).
- Zhu, B. F. *et al.* Population genetic analysis of 15 autosomal STR, 3889–3895 (2010).
- Wu, Y. M., Zhang, X. N., Zhou, Y., Chen, Z. Y. & Wang, X. B. Genetic polymorphisms of 15 STR loci in Chinese Han population living in Xi'an city of Shaanxi Province. *Forensic Sci Int Genet* **2**, e15–18 (2008).
- Chen, J. G. *et al.* Population genetic data of 15 autosomal STR loci in Uygur ethnic group of China. *Forensic Sci Int Genet* **6**, e178–9 (2012).
- Zhu, B. F., Shen, C. M., Wu, Q. J. & Deng, Y. J. Population data of 15 STR loci of Chinese Yi ethnic minority group. *Leg Med (Tokyo)* **10**, 220–224 (2008).
- Zhu, J. *et al.* Population data of 15 STR in Chinese Han population from north of Guangdong. *J Forensic Sci* **50**, 1510–1511 (2005).
- Wang, Z. *et al.* Genetic polymorphisms of 15 STR loci in Chinese Hui population. *J Forensic Sci* **50**, 1508–1509 (2005).
- Zhu, B. *et al.* Genetic analysis of 15 STR loci of Chinese Uigur ethnic population. *J Forensic Sci* **50**, 1235–1236 (2005).
- Li, C. *et al.* Genetic polymorphism of 17 STR loci for forensic use in Chinese population from Shanghai in East China. *Forensic Sci Int Genet* **3**, e117–118 (2009).
- Rerkamnuaychoke, B. *et al.* Thai population data on 15 tetrameric STR loci-D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818 and FGA. *Forensic Sci Int* **158**, 234–237 (2006).
- Alenizi, M., Goodwin, W., Ismael, S. & Hadi, S. STR data for the AmpFI STR Identifier loci in Kuwaiti population. *Leg Med (Tokyo)* **10**, 321–325 (2008).
- Coudray, C. *et al.* Population genetic data of 15 tetrameric short tandem repeats (STRs) in Berbers from Morocco. *Forensic Sci Int* **167**, 81–86 (2007).
- Forward, B. W., Eastman, M. W., Nyambo, T. B. & Ballard, R. E. AMPFI STR Identifier STR allele frequencies in Tanzania, Africa. *J Forensic Sci* **53**, 245 (2008).
- Muro, T. *et al.* Allele frequencies for 15 STR loci in Ovambo population using AmpFI STR Identifier Kit. *Leg Med (Tokyo)* **10**, 157–159 (2008).
- Immel, U. D., Erhuma, M., Mustafa, T., Kleiber, M. & Klintschar, M. Population genetic analysis in a Libyan population using the PowerPlex 16 system. *Int Congr Ser* **1288**, 421–423 (2006).
- Coudray, C., Guitard, E., EL-Chennawi, F., Larrouy, G. & Dugoujon, J. M. Allele frequencies of 15 short tandem repeats (STRs) in three Egyptian populations of different ethnic groups. *Forensic Sci Int* **169**, 260–265 (2007).
- Coudray, C. *et al.* Allele frequencies of 15 tetrameric short tandem repeats (STRs) in Andalusians from Huelva (Spain). *Forensic Sci Int* **68**, e21–24 (2007).
- Marian, C. *et al.* STR data for the 15 AmpFI STR identifier loci in the Western Romanian population. *Forensic Sci Int* **170**, 73–75 (2007).
- Sánchez-Diz, P., Menounos, P. G., Carracedo, A. & Skitsa, I. 16 STR data of a Greek population. *Forensic Sci Int Genet* **2**, e71–72 (2008).
- Mertens, G. *et al.* Flemish population genetic analysis using 15 STRs of the Identifier<sup>®</sup> kit. *Int Congr Ser* **1288**, 328–330 (2006).
- Hernández-Gutiérrez, S., Hernández-Franco, P., Martínez-Tripp, S., Ramos-Kuri, M. & Rangel-Villalobos, H. STR data for 15 loci in a population sample from the central region of Mexico. *Forensic Sci Int* **151**, 97–100 (2005).
- Ossmani, H. E., Talbi, J., Bouchrif, B. & Chafik, A. Allele frequencies of 15 autosomal STR loci in the southern Morocco population with phylogenetic structure among worldwide populations. *Leg Med (Tokyo)* **11**, 155–158 (2009).
- Abu Halima, M. S., Bernal, L. P. & Sharif, F. A. Genetic variation of 15 autosomal short tandem repeat (STR) loci in the Palestinian population of Gaza Strip. *Leg Med (Tokyo)* **11**, 203–204 (2009).
- Kutanan, W., Kitpipit, T., Phetpeng, S. & Thanakiatkrai, P. Forensic STR loci reveal common genetic ancestry of the Thai-Malay Muslims and Thai Buddhists in the deep Southern region of Thailand. *J Hum Genet* **59**, 675–681 (2014).

47. Butler, J. M., Schoske, R., Vallone, P. M., Redman, J. W. & Kline, M. C. Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. *J Forensic Sci* **48**, 908–911 (2003).
48. Pereira, L. *et al.* PopAffiliator: online calculator for individual affiliation to a major population group based on 17 autosomal short tandem repeat genotype profile. *Int J Legal Med* **125**, 629–636 (2011).
49. Yao, H. B. *et al.* Genetic evidence for an East Asian origin of Chinese Muslim populations Dongxiang and Hui. *Sci Rep* **6**, 38656, doi: 10.1038/srep38656 (2016).
50. Li, Y. *et al.* Allele frequency of 19 autosomal STR loci in the Bai population from the southwestern region of mainland China. *Electrophoresis*. **36**, 2498–2503 (2015).
51. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
52. Hubisz, M. J., Falush, D., Stephens, M. & Pritchard, J. K. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* **9**, 1322–1332 (2009).
53. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna (2013).
54. Rosenberg, N. A. Distruct: A program for the graphical display of population structure. *Mol Ecol Notes* **4**, 137–138 (2004).
55. Gayden, T. *et al.* The Himalayas as a directional barrier to gene flow. *Am J Hum Genet* **80**, 884–94 (2007).
56. Torroni, A. *et al.* Mitochondrial DNA analysis in Tibet: implications for the origin of the Tibetan population and its adaptation to high altitude. *Am J Phys Anthropol* **93**, 189–99 (1994).

## Acknowledgements

This work was supported by the Natural Science Foundation of Gansu province (1308RJZA190), Scientific Research Project for Colleges of Gansu province (2014A-085), National Excellent Youth Science Foundation of China (31222030), National Natural Science Foundation of China (31671297, 31071098). C.C.W. is supported by Max Planck Society. The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 646612) granted to Martine Robbeets.

## Author Contributions

H.B.Y., H.L. and C.C.W. supervised the study. H.B.Y. and X.T. collected the samples and did the experiments. C.C.W. and S.L. analyzed the data. J.W., S.Q.W., Q.D., Q.Y.D., B.X., Y.H., H.D.W., S.L., B.C., L.M., contributed to the dataset collection. C.C.W. wrote the manuscript. L.J., J.K., and H.L. were involved in discussions and manuscript revisions. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Yao, H.-B. *et al.* Genetic structure of Tibetan populations in Gansu revealed by forensic STR loci. *Sci. Rep.* **7**, 41195; doi: 10.1038/srep41195 (2017).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017