

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/318961058>

The genetic admixture in Tibetan–Yi Corridor

Article in *American Journal of Physical Anthropology* · August 2017

DOI: 10.1002/ajpa.23291

CITATIONS

5

READS

683

16 authors, including:



Senwei Tang

The Chinese University of Hong Kong

11 PUBLICATIONS 272 CITATIONS

[SEE PROFILE](#)



Xiaotian Yao

11 PUBLICATIONS 601 CITATIONS

[SEE PROFILE](#)



Ivy Hui-Yuan Yeh

Nanyang Technological University

32 PUBLICATIONS 103 CITATIONS

[SEE PROFILE](#)



Qiajun Du

Lanzhou University

2 PUBLICATIONS 10 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



The genetic history of Hmong-Mien-speaking populations [View project](#)



Genetic structure and demographical history of Trans-Eurasian, Tai-Kadai, Hmong-Mien-speaking populations [View project](#)

1
2
3

6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The Genetic Admixture in Tibetan-Yi Corridor

Hong-Bing Yao¹, Senwei Tang², Xiaotian Yao², Hui-Yuan Yeh³, Wanhu Zhang⁴, Zhiyan Xie⁴, Qiajun Du⁵, Liying Ma¹, Shuoyun Wei¹, Xue Gong¹, Zilong Zhang¹, Quanfang Li¹, Bingying Xu⁶, Hu-Qin Zhang⁷, Gang Chen², Chuan-Chao Wang^{8,9,10,t}

1. Key Laboratory of Evidence Science of Gansu Province, Gansu Institute of Political Science and Law, Lanzhou 730070, China
2. WeGene, Shenzhen 518040, China
3. School of Humanities and School of Medicine, Nanyang Technological University, Singapore 639798.
4. People's Hospital of Gaotai, Gaotai 734300, Gansu Province, China
5. Lanzhou University Second Hospital Clinical Laboratory, Lanzhou 730000, Gansu Province, China
6. School of Forensic Medicine, Kunming Medical University, Kunming, 650500, China
7. The Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China
8. Department of Anthropology and Ethnology, Xiamen University, Xiamen 361005, China
9. Department of Archaeogenetics and Eurasia3angle research group, Max Planck Institute for the Science of Human History, D-07745 Jena, Germany
10. Department of Genetics, Harvard Medical School, Boston, MA 02115, United States

tCorrespondence: Dr. Chuan-Chao Wang, Kahlaische Strasse 10, 07745 Jena, Germany, Email: wang@shh.mpg.de, Tel: +49 (0) 3641 686-648, Fax: +49 (0) 3641 686-667.

Objectives: The Tibetan-Yi Corridor located on the eastern edge of Tibetan Plateau is suggested to be the key region for the origin and diversification of Tibeto-Burman speaking populations and the main route of the peopling of the Plateau. However, the genetic history of the populations in the Corridor is far from clear due to limited sampling in the northern part of the Corridor.

Materials and Methods: We collected blood samples from 10 Tibetan and 10 Han Chinese individuals from Gansu province and genotyped about 600,000 genome-wide single nucleotide polymorphisms (SNPs).

Results: Our data revealed that the populations in the Corridor are all admixed on a genetic cline of deriving ancestry from Tibetans on the Plateau and surrounding lowland East Asians. The Tibetan and Han Chinese groups in the north of the Plateau show significant evidence of low-level West Eurasian admixture that could be probably traced back to 600~900 years ago.

Discussion: We conclude that there have been huge population migrations from surrounding

54 lowland onto the Tibetan Plateau via the Tibetan-Yi Corridor since the initial formation of
55 Tibetans probably in Neolithic Time, which leads to the current genetic structure of Tibeto-
56 Burman speaking populations.
57

58 **Keywords:** Han Chinese, Tibetan, gene flow

INTRODUCTION

The Sino-Tibetan languages are a family of more than 400 languages, including two subfamilies, namely Tibeto-Burman and Chinese, which are spoken by over a billion people all over East Asia, Southeast Asia, and parts of South Asia (Martisoff, 1991). The linguistic affinity between Tibeto-Burman and Chinese are well established with many cognates between Proto-Tibeto-Burman and Old Chinese (Martisoff, 1991). The split time for Tibeto-Burman and Chinese was estimated around 6 thousand years ago (kya) based on lexical evidence and cladistic methods (Wang, 1998). Archaeological evidence also indicated that the ancestors of Sino-Tibetan populations that probably could be associated with Neolithic farming populations lived around at least 6 kya in western China (Barton et al., 2009; Shelach et al., 2000; Yang et al., 2012). Despite intense linguistic and archaeological researches, little has been known about how the Tibeto-Burman and Chinese diversified in western China.

The genetic evidence, especially from the paternal Y chromosome and maternal mitochondrial DNA (mtDNA), has shed more light on the history of Sino-Tibetan populations during the past two decades. Y chromosome suggests Tibeto-Burman populations are an admixture of the northward migrations of the initial settlers of East Asia with haplogroup D-M175 in the Late Paleolithic age, and the southward Di-Qiang people with dominant haplogroup O-M134 (xM117) and O-M117 via Tibetan-Yi Corridor through a series of migrations since the Neolithic Age (Su et al., 2000; Kang et al., 2012; Qi et al., 2013; Wang et al., 2014). The Tibetan-Yi Corridor located on the eastern edge of the Tibetan Plateau ranging from the south of Gansu province to the north of Yunnan province is suggested to be the key diversification region for various Tibeto-Burman groups (Shi, 2005). Y-chromosomal Haplogroup O-M134 (xM117) and O-M117 are also characteristic lineages of Han Chinese, comprising 11.4% and 16.3%, respectively (Yan et al., 2011, 2014). However, Haplogroup O-002611, another dominant paternal lineage of Han Chinese, is found at very low frequencies in Tibeto-Burman populations, suggesting this lineage might not have participated in the formation of Tibeto-Burman populations (Wang et al., 2013, 2014; Yan et al., 2011, 2014; Yao et al., 2017). On the maternal mtDNA side, the high frequencies of northern Asian specific haplogroup A, D, G, and M8 suggest a northern Asian origin of Tibeto-Burman speakers (Qi et al., 2013; Qin et al., 2010; Zhao et al., 2009). The genetic relics of the Late Paleolithic ancestors of Tibeto-Burman populations have also been reported, such as haplogroup M62 (Zhao et al., 2009). 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 (Wen et al., 2004). The Tibetans and Lolo-Burmese speaking groups tend to have quite different genetic compositions based on the frequency data of 15 autosomal short tandem repeats (STRs), which is probably due to long-term isolations and genetic drift (Li et al., 2015; Yao et al., 2017).

The genome-wide data for Tibetans become available in recent years, but mainly focus on their genetic basis in adapting high-altitude environments (Beall et al., 2010; Simonson et al., 2010; Yi et al., 2010; Xu et al., 2011; Wang et al., 2011; Jeong et al., 2014; Wuren et al., 2014; Petousi et al., 2014). Wang et al (Wang et al., 2011) reported that Tibetans are genetically similar with other East Asian populations compared with West Eurasian and South Asians. Jeong et al (Jeong et al., 2014) suggested Tibetans are a mixture of ancestral populations related to the Sherpa and Han Chinese. Lu et al (Lu et al., 2016) reported from whole-genome

1
2
3 perspective that most of the Tibetan gene pool diverged from that of Han Chinese about 15
4 kya to 9 kya and the shared ancestry of Tibetan-enriched sequences dates back to 62–38 kya,
5 which is consistent with Y chromosome evidence of two-phase for the origin of Tibeto-
6 Burman populations. Jeong et al (Jeong et al., 2016) reported ancient genomes from the
7 Chokhopani, Mebrak, and Samdzong sites spanning 3 to 1 kya in Nepal suggesting the
8 Tibetan Plateau experienced millennia of genetic continuity which continues to the present
9 day.

10
11
12 The origin and diversification of Tibeto-Burman populations seem to involve substantial
13 genetic admixture with surrounding lowland populations viewed from the above previous
14 studies. However, the limited markers of mtDNA and Y chromosome and insufficient
15 sampling of genome-wide study are not enough to give a comprehensive understanding of the
16 genetic history and admixture process of Tibeto-Burman populations. In addition, Tibetan and
17 Han Chinese populations of Gansu province, the northern edge of Tibetan-Yi Corridor, have
18 seldom been studied genetically. Therefore, we here analyze about 600,000 genome-wide
19 SNPs from 20 samples collected from Tibetan and Han Chinese populations in south Gansu
20 province to explore the genetic structure and admixture of Tibeto-Burman populations.

21 22 23 **MATERIALS AND METHODS**

24 25 **Sampling and genotyping**

26
27 We collected blood samples from 10 unrelated individuals from Tibetan and the other 10
28 unrelated individuals from Han Chinese in south Gansu province (Figure 1). Our study was
29 approved by the Ethnic Committee of Gansu Institute of Political Science and Law. The study
30 was conducted in accordance with the human and ethical research principles of Gansu
31 Institute of Political Science and Law. Informed consent was obtained from all individual
32 participants included in the study. DNA isolation and purification were following the standard
33 lysis protocol, and DNA was purified using the QIAamp DNA mini kit (QIAGEN, Hilden,
34 Germany). Genotyping was performed on the Affymetrix WeGene V1 Arrays covering
35 596,744 SNPs at the Shanghai Jiaotong University, Shanghai. The WeGene V1 arrays were
36 designed to identify all known paternal Y-chromosome and maternal mtDNA lineages with
37 18963 Y-chromosome and 4448 mtDNA phylogenetic relevant SNPs. We genotyped our
38 samples using WeGene arrays because we want to generate informative Y chromosome and
39 mtDNA results. The dataset generated during the current study is available upon request to
40 the corresponding author when the paper is published.

41 42 43 44 **Data merging**

45
46 We merged our 20 samples with previously published populations from International
47 HapMap Project Phase 3 (International HapMap Consortium, 2003), Human Genome
48 Diversity Project (HGDP) (Li et al., 2008), Simons Genome Diversity Project (SGDP)
49 (Mallick et al., 2016), and ancient Nepalese (Jeong et al., 2016) and present-day Tibetans of
50 Lhasa and Yunnan province (Beall et al., 2010; Wang et al., 2011). We finally generated a
51 combined dataset covering 304180 SNPs that were used in subsequent analysis.

52 53 54 **Principal Component Analysis**

55
56 We used smartpca (version: 13050), part of the EIGENSOFT package (Patterson et al., 2006),
57 to carry out Principal Component Analysis (PCA). We performed PCA on present-day

1
2
3 populations and then projected the ancient samples using the `lsqproject: YES` option, which
4 accounts for samples with substantial missing data. We did not perform any outlier removal
5 iterations (`numoutlieriter: 0`). We set all other options to the default. We assessed statistical
6 significance with a Tracy-Widom test using the `twstats` program of EIGENSOFT. All the first
7 six principal components that we discuss and plot in what follows were highly statistically
8 significant ($P < 10^{-12}$).
9

10 *f₃-statistics*

11 We computed statistics of the form f_3 (Mbuti; X, Y) using the `qp3Pop` program of
12 ADMIXTOOLS (Patterson et al., 2012; Reich et al., 2009), which measure the shared genetic
13 drift between populations X and Y since their separation from an African outgroup Mbuti.
14

15 For each Tibeto-Burman and Han Chinese population Z in turn, we computed statistics of the
16 form $f_3(Z; X, Y)$ where X and Y are worldwide populations. A significantly negative f_3 -value
17 provides unambiguous evidence of admixture in Z from populations that are related, perhaps
18 distantly, to population X and Y.
19

20 *f₄-statistics*

21 We computed f_4 -statistics of the form $f_4(X, Y; Test, Outgroup)$ using the `qpDstat` program of
22 ADMIXTOOLS (Patterson et al., 2012; Reich et al., 2009) with default parameters to show if
23 population *Test* is symmetrically related to X and Y or shares an excess of alleles with either
24 of the two, with standard errors computed with a block jackknife.
25

26 *f₄-ratio estimations*

27
28 We used the `qpF4ratio` program of ADMIXTOOLS (Patterson et al., 2012; Reich et al., 2009)
29 with default parameters to estimate the admixture proportions of tested populations with the
30 proposed sources.
31

32 **ADMIXTURE analysis**

33
34 We carried out model-based clustering analysis using ADMIXTURE 1.23 (Alexander et al.,
35 2009), combining the present-day worldwide populations and ancient Nepalese samples with
36 our 20 individuals. We used PLINK v1.90 (Chang et al., 2015) to thin the dataset of 304,180
37 autosomal SNPs to remove SNPs in strong linkage disequilibrium, employing a window of
38 200 SNPs advanced by 25 SNPs and an r^2 threshold of 0.4 (with the flag: `--indep-pairwise`
39 `200 25 0.4`). A total of 175,483 SNPs remained for analysis after this procedure. We ran
40 ADMIXTURE with default 5-fold cross-validation (`--cv=5`), varying the number of ancestral
41 populations between $K=2$ and $K=16$ in 100 bootstraps with different random seeds. We used
42 the unsupervised ADMIXTURE approach, in which allele frequencies for non-admixed
43 ancestral populations are unknown and are computed during the analysis. We used point
44 estimation and terminated the block relaxation algorithm when the objective function $\Delta <$
45 0.0001 . We chose the best run according to the highest log likelihood. We used cross-
46 validation to identify an “optimal” number of clusters. We observed the lowest CV errors for
47 $K=12$.
48
49

50 **Weighted linkage disequilibrium (LD) Analysis**

51 LD decay was calculated using ALDER (Loh et al., 2013) to infer admixture parameters
52 including dates and mixture proportions.
53

54 **Y chromosomal and mtDNA haplogroup assignment**

1
2
3 The WeGene V1 Arrays were designed to identify all known Y chromosomal and mtDNA
4 lineages with 18963 Y-chromosome and 4448 mtDNA phylogenetic relevant SNPs. We
5 assign the Y chromosomal and mtDNA haplogroups using in-house tools following the
6 International Society of Genetic Genealogy (2016). Y-DNA Haplogroup Tree 2016, Version:
7 1.87, Date: 29 March 2016, <http://www.isogg.org/tree/> 30 March 2016; and mtDNA tree Build
8 16 (van Oven et al., 2009), <http://www.phylotree.org/>.
9
10

11 RESULTS

12
13 We firstly performed PCA to provide a broad overview of population structure across East
14 Asia (Figure 2). We show five broad clusters correlating well with geographic and linguistic
15 categories within East Asia: a southern cluster with Austronesian, Tai-Kadai, and
16 Austroasiatic speaking groups; a Han Chinese cluster; a Tibetan cluster; a Japanese cluster;
17 and an Altaic cluster consisting of Turkic, Tungusic and Mongolic-speaking groups in north
18 China. Our Tibetan and Han Chinese samples from Gansu province together with other
19 Tibeto-Burman speaking populations in the Corridor (Tibetan_Yunnan, Naxi, and Yi) appear
20 as potentially admixed populations on the PCA occupying an intermediate position between
21 Tibetan and Han Chinese.
22

23
24 In the model-based ADMIXTURE clustering analysis, we used cross-validation to identify an
25 “optimal” number of clusters. We observed the lowest CV errors for $K=12$. At $K=12$, we
26 observed three ancestral components specific to individuals in East Asia. One of these
27 components is enriched in the ancient Nepalese and found to be at highest proportions in
28 Tibetans. The second is enriched in Taiwan Austronesians but is also prevalent in Han
29 Chinese. The third component is enriched in Yakut, a Turkic-speaking population in Siberia.
30 We found our Tibetan and Han Chinese samples of Gansu province are genetically similar
31 with Han Chinese in northern China and other Corridor populations (Supplementary Figure 1).
32
33
34

35 We next calculated an outgroup f_3 -statistics of the form $f_3(\text{Mbuti}; X, Y)$ to quantify
36 population differentiation across East Asia observed by PCA (Supplementary Table 1 and
37 Figure 3). The Han_Gansu cluster with other lowland populations, but Tibetan_Gansu group
38 tightly together with other Tibeto-Burman speaking populations in the Corridor. The f_3 -
39 statistics correlate well with the patterns observed via PCA that the populations in the
40 Tibetan-Yi Corridor share an affinity with both Tibetans on the Plateau and Han Chinese and
41 other East Asians in the lowland. We also plotted the outgroup f_3 -statistics in the form of $f_3(X,$
42 $\text{Chokhopani}; \text{Mbuti})/f_3(X, \text{Ami}; \text{Mbuti})$ to visualize the allele sharing of various East Asian
43 populations with Ami and Chokhopani (an ancient sample tracing back to 3kya in Nepal
44 showing genetic continuity with present-day Tibetans) in Figure 4a. We observed a clear
45 cline of differences in Tibetan related ancestry with Sherpa and Tibetan_Lhasa sharing the
46 most genetic drift with ancient Chokhopani and Cambodian sharing the least. Consistent with
47 the PCA plot, the populations in the Tibetan-Yi Corridor are in the middle of this Tibetan
48 ancestry cline.
49
50
51

52 We applied a formal admixture test using f_3 -statistics in the form of $f_3(Z; X, Y)$ where Z is our
53 tested group and X and Y are worldwide populations that might be the genetic sources for
54 modeling the admixture in population Z . We observed significant signals of admixture ($Z < -5$)
55 in Han_Gansu, Yi, and Tibetan_Yunnan for Tibetan related and lowland East Asian related
56 ancestry (Table 1).
57
58
59
60

1
2
3 We proceed to use f_4 -ratio based estimates in the form of f_4 (Chokhopani, Mbuti; Test,
4 Ami)/ f_4 (Chokhopani, Mbuti; Tibetan_Lhasa, Ami) to quantify the proportions of Tibetan
5 related ancestry East Asian groups. We assigned the Tibetan_Lhasa and Ami as ancestral
6 source populations based on the f_3 -statistics. We observed a consistent genetic cline with Fig.
7
8 3a in term of Tibetan related ancestry (Table 2). The Tibetan_Yunnan, Naxi, Yi, and
9 Tibetan_Gansu derive 90.1%, 80.1%, 69.6%, and 48.0% Tibetan related ancestry,
10 respectively.

11
12 We also detected the evidence of West Eurasian admixture into our Han Chinese samples in
13 Gansu as shown in the significant negative Z-score of f_3 -statistics when assuming West
14 Eurasian populations as sources (Table 1 and Supplementary Table 2). The outgroup f_3 -
15 statistics we calculated in Figure.4b show Han Chinese and Tibetan populations in Gansu
16 share more genetic drift with French compared with other Han Chinese and Tibetans. We
17 confirmed the West Eurasian admixture using f_4 -statistics in the form of (West Eurasians,
18 Mbuti; Han_Gansu/Tibetan_Gansu, Han) in which “Han” is the Han Chinese of HGDP in
19 Table 3, where the significant positive statistics suggest West Eurasians share more allele
20 with Han_Gansu and Tibetan_Gansu compared with Han Chinese of HGDP. We can get
21 significant positive values when putting other Corridor populations (Tibetan_Yunnan, Naxi,
22 and Yi) in place of “Han” in the above f_4 -statistics, which suggest there is substructure in
23 those Corridor populations with the northern ones specially sharing an affinity with West
24 Eurasians.
25
26
27

28 We estimated the admixture time and lower bounds on the admixture proportion using the
29 linkage disequilibrium (LD)-based admixture inference implemented in ALDER (Loh et al.,
30 2013) and showed the results in Table 4. For Tibetan highlander and lowland East Asian
31 admixture in Corridor populations, we weighted LD curves with Corridor groups as test
32 populations and Tibetan_Lhasa and CHB or CHD as possible source populations. We also
33 took advantage of the one-reference inference capabilities of ALDER to only use CHB or
34 CHD as the source to estimate the lower bounds of the admixture proportions. The average
35 admixture times for Corridor populations range from 5 to 70 generations (about 150 to 2100
36 years ago assuming 30 years a generation). The Tibetans in Yunnan are suggested to derive
37 11%-18% ancestry from Han Chinese. For the West Eurasian admixture, we computed
38 weighted LD curves with Han_Gansu and Tibetan_Gansu as the test populations and CEU
39 and CHB as the sources. The average admixture times for Han Chinese and Tibetan in Gansu
40 range from about 20 to 30 generations ago (about 600 to 900 years assuming 30 years a
41 generation), suggesting relatively recent gene flow from West Eurasia to northwest China.
42 We also estimated mixture fractions of at least 2.4% to 3.3% CEU-related ancestry for those
43 populations. Changing the starting point of the LD fit does not qualitatively affect the results
44 (Supplementary Document 1). We caution that the date estimates might not reflect the initial
45 admixture in present-day populations; instead, it is an average date of population mixture. If
46 the admixture did not happen immediately when two populations met or occurred many times
47 over an extended period, the true start of mixture would be more ancient.
48
49
50
51
52

53 The paternal Y chromosome also gave evidence for the West Eurasian admixture, as we
54 identified West Eurasian characteristic lineages E and J in our Han Chinese and Tibetan
55 samples (Table 5). The dominant lineage in Han Chinese of Gansu is O2, which is in
56 consistent with the general paternal profile of other Han Chinese groups (Yan et al., 2011,
57 2014). However, the paternal history of Tibetans_Gansu is more complicated with haplogroup
58 C, N, and Q. The maternal mtDNA lineages of our samples are consistent with the general
59
60

1
2
3 profile of this region with high frequency of D4 (Qi et al., 2013; Wang et al., 2014; Qin et al.,
4 2010).

8 DISCUSSION

9
10 The Sino-Tibetan language family comprises more than 400 languages that are spoken by
11 over a billion people distributed in East Asia and Southeast Asia, including the Tibeto-
12 Burman and Chinese subfamilies (Martisoff, 1991). Despite intense linguistic, archaeological,
13 and genetic researches, how the Tibeto-Burman groups originated and diversified and how
14 they dispersed remain major open questions. The Tibetan-Yi Corridor is located on the
15 eastern edge of Tibetan Plateau, and is very diverse in both geography and culture and
16 suggested to be the main region for the diversification of Tibeto-Burman groups (Shi, 2005).
17 Taking advantage of the high-density genotyping data in Tibetan and Han Chinese
18 populations collected from the northern part of the Corridor, we conducted the comprehensive
19 genome-wide study and provided a genomic landscape and admixture history of populations
20 in the Corridor.
21
22
23

24 Our findings clearly show that the populations in the Tibetan-Yi Corridor are admixed
25 deriving ancestry from Tibetan highlanders and surrounding lowland East Asians, such as
26 Han Chinese and various southern groups speaking Austronesian, Tai-Kadai, and
27 Austroasiatic languages. The Tibetan and Han Chinese in the northern terminal of the
28 Corridor also have significant evidence of West Eurasian admixture. Our results confirm that
29 the Tibetan-Yi Corridor is an active contacting region for Tibetan and Han Chinese and also
30 the key region for the formation and diversification of Tibeto-Burman groups (Shi, 2005).
31
32

33 The archaeological and genetic evidence show that the Tibeto-Burman populations are an
34 admixture of the initial settlers of East Asia probably in the Late Paleolithic Age and the
35 Neolithic farming populations from the Upper and Middle Yellow River Basin (Barton et al.,
36 2009; Su et al., 2000; Kang et al., 2012; Qi et al., 2013; Wang et al., 2014). Our results here
37 give evidence that there are huge population migrations from surrounding lowland onto the
38 Tibetan Plateau via the Tibetan-Yi Corridor since the initial formation of Tibetans in
39 Neolithic Time, which suggest the large mountainous regions are not barriers for human
40 diffusion.
41
42

43 Han Chinese as a whole has long been suggested to be a homogeneous group due to recent
44 population expansions (Chen et al., 2009; Xu et al., 2009; Nothnagel et al., 2017). However,
45 we detected genetic substructure in Han Chinese populations that the Han Chinese in
46 northwest China show a low level of West Eurasian influence. The Tibetan groups in
47 northwest China also have this West Eurasian attraction compared with Tibetans in other
48 regions. The time estimation suggests the admixture happened probably in recent times within
49 the last 1000 years. This raises the possibility that the admixture probably resulted from
50 commercial, religious, and cultural network interlinking the historical trade routes between
51 the West Eurasia and East Asia, such as the well-known Silk Road.
52

54 ACKNOWLEDGMENTS

55
56 This work was supported by the Natural Science Foundation of Gansu Province
57 (1308RJZA190), Scientific Research Project for Colleges of Gansu province (2014A-085),
58

Scientific Research Project for Colleges of Gansu province (2015A-105), Natural Science Foundation for Young Scientists of China (51501042), Lanzhou Research Program of Science and Technology (2016-3-122), and Natural Science Foundation (H2302). C.C.W is supported by Nanqiang Outstanding Young Talents Program of Xiamen Univeristy, Max Planck Society and Harvard Medical School. 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.

REFERENCES

- Alexander, D.H., Novembre, J., Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655-1664.
- Barton, L., Newsome, S.D., Chen, F.H., Wang, H., Guilderson, T.P., Bettinger, R.L. (2009). Agricultural origins and the isotopic identity of domestication in northern China. *Proc Natl Acad Sci U S A.* 106, 5523-5528.
- Beall, C.M., Cavalleri, G.L., Deng, L., Elston, R.C., Gao, Y., Knight, J., Li, C., Li, J.C., Liang, Y., McCormack, M., Montgomery, H.E., Pan, H., Robbins, P.A., Shianna, K.V., Tam, S.C., Tsering, N., Veeramah, K.R., Wang, W., Wangdui, P., Weale, M.E., Xu, Y., Xu, Z., Yang, L., Zaman, M.J., Zeng, C., Zhang, L., Zhang, X., Zhaxi, P., Zheng, Y.T. (2010). Natural selection on EPAS1 (HIF2a) associated with low hemoglobin concentration in Tibetan highlanders. *Proc. Natl Acad. Sci. USA.* 107, 11459–11464.
- Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J. (2015). Second-generation PLINK, rising to the challenge of larger and richer datasets. *Gigascience.* 4,7. <https://www.cog-genomics.org/plink2>.
- Chen, J., Zheng, H., Bei, J.X., Sun, L., Jia, W.H., Li, T., Zhang, F., Seielstad, M., Zeng, Y.X., Zhang, X., Liu, J. (2009). Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *Am J Hum Genet.* 85, 775-785.
- Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D.B., Pritchard, J.K., Beall, C.M., Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat Commun.* 5,3281.
- Jeong, C., Ozga, A.T., Witonsky, D.B., Malmström, H., Edlund, H., Hofman, C.A., Hagan, R.W., Jakobsson, M., Lewis, C.M., Aldenderfer, M.S., Di Rienzo, A., Warinner, C. (2016). Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc Natl Acad Sci U S A.* 113, 7485-7490.
- Kang, L., Lu, Y., Wang, C., Hu, K., Chen, F., Liu, K., Li, S., Jin, L., Li, H., Genographic Consortium. (2012). Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the eastern Himalayas. *Ann Hum Genet.* 76, 92-99.
- Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science.* 319, 1100-1104.
- Li, Y., Hong, Y., Li, X., Yang, J., Li, L., Huang, Y., Wang, C., Li, H., Xu, B. (2015). Allele frequency of 19 autosomal STR loci in the Bai population from the southwestern region of mainland China. *Electrophoresis.* 36, 2498-503.

- 1
2
3 Loh, P.R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J.K., Reich, D. and Berger, B.
4 (2013). Inferring admixture histories of human populations using linkage
5 disequilibrium. *Genetics*. 193, 1233-1254.
- 6 Lu, D., Lou, H., Yuan, K., Wang, X., Wang, Y., Zhang, C., Lu, Y., Yang, X., Deng, L., Zhou,
7 Y., Feng, Q., Hu, Y., Ding, Q., Yang, Y., Li, S., Jin, L., Guan, Y., Su, B., Kang, L.,
8 Xu, S. (2016). Ancestral Origins and Genetic History of Tibetan Highlanders. *Am J*
9 *Hum Genet*. 99, 580-594.
- 10 Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri,
11 N., Nordenfelt, S., Tandon, A., Skoglund, P., Lazaridis, I., Sankararaman, S., Fu, Q.,
12 Rohland, N., Renaud, G., Erlich, Y., Willems, T., Gallo, C., Spence, J.P., Song, Y.S.,
13 Poletti, G., Balloux, F., van Driem, G., de Knijff, P., Romero, I.G., Jha, A.R., Behar,
14 D.M., Bravi, C.M., Capelli, C., Hervig, T., Moreno-Estrada, A., Posukh, O.L.,
15 Balanovska, E., Balanovsky, O., Karachanak-Yankova, S., Sahakyan, H., Toncheva,
16 D., Yepiskoposyan, L., Tyler-Smith, C., Xue, Y., Abdullah, M.S., Ruiz-Linares, A.,
17 Beall, C.M., Di Rienzo, A., Jeong, C., Starikovskaya, E.B., Metspalu, E., Parik, J.,
18 Villems, R., Henn, B.M., Hodoglugil, U., Mahley, R., Sajantila, A.,
19 Stamatoyannopoulos, G., Wee, J.T., Khusainova, R., Khusnutdinova, E., Litvinov, S.,
20 Ayodo, G., Comas, D., Hammer, M.F., Kivisild, T., Klitz, W., Winkler, C.A., Labuda,
21 D., Bamshad, M., Jorde, L.B., Tishkoff, S.A., Watkins, W.S., Metspalu, M.,
22 Dryomov, S., Sukernik, R., Singh, L., Thangaraj, K., Pääbo, S., Kelso, J., Patterson,
23 N., Reich, D. (2016). The Simons Genome Diversity Project, 300 genomes from 142
24 diverse populations. *Nature*. 538, 201-206.
- 25 Martisoff, J.A. (1991). Sino-Tibetan linguistics, present state and future prospects. *Annu Rev*
26 *Anthropol*. 20, 469–504.
- 27 Nothnagel, M., Fan, G., Guo, F., He, Y., Hou, Y., Hu, S., Huang, J., Jiang, X., Kim, W., Kim,
28 K., Li, C., Li, H., Li, L., Li, S., Li, Z., Liang, W., Liu, C., Lu, D., Luo, H., Nie, S., Shi,
29 M., Sun, H., Tang, J., Wang, L., Wang, C.C., Wang, D., Wen, S.Q., Wu, H., Wu, W.,
30 Xing, J., Yan, J., Yan, S., Yao, H., Ye, Y., Yun, L., Zeng, Z., Zha, L., Zhang, S.,
31 Zheng, X., Willuweit, S., Roewer, L. (2017). Revisiting the male genetic landscape of
32 China, a multi-center study of almost 38,000 Y-STR haplotypes. *Hum Genet*. doi,
33 10.1007/s00439-017-1759-x.
- 34 Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T.,
35 Webster, T., Reich, D. (2012). Ancient admixture in human history. *Genetics*. 192,
36 1065-1093.
- 37 Patterson, N., Price, A.L., Reich, D. (2006). Population structure and eigenanalysis. *PLoS*
38 *Genet*. 2, e190.
- 39 Petousi, N., Croft, Q.P., Cavalleri, G.L., Cheng, H.Y., Formenti, F., Ishida, K., Lunn, D.,
40 McCormack, M., Shianna, K.V., Talbot, N.P., Ratcliffe, P.J., Robbins, P.A. (2014).
41 Tibetans living at sea level have a hyporesponsive hypoxia-inducible factor system
42 and blunted physiological responses to hypoxia. *J Appl Physiol*. 116,893–904.
- 43 Qi, X., Cui, C., Peng, Y., Zhang, X., Yang, Z., Zhong, H., Zhang, H., Xiang, K., Cao, X.,
44 Wang, Y., Ouzhuluobu, Basang, Ciwangsangbu, Bianba, Gonggalanzi, Wu, T., Chen,
45 H., Shi, H., Su, B. (2013). Genetic evidence of paleolithic colonization and neolithic
46 expansion of modern humans on the Tibetan plateau. *Mol Biol Evol*. 30, 1761–1778.
- 47 Qin, Z., Yang, Y., Kang, L., Yan, S., Cho, K., Cai, X., Lu, Y., Zheng, H., Zhu, D., Fei, D., Li,
48 S., Jin, L., Li, H., Genographic Consortium. (2010). A mitochondrial revelation of
49 early human migrations to the Tibetan Plateau before and after the last glacial
50 maximum. *Am J Phys Anthropol*. 143, 555–569.
- 51
52
53
54
55
56
57
58

- 1
2
3 International HapMap Consortium. (2003). The International HapMap Project. *Nature*. 426,
4 789-796.
- 5 Reich, D., Thangaraj, K., Patterson, N., Price, A.L., Singh, L. (2009). Reconstructing Indian
6 population history. *Nature*. 461, 489-494.
- 7
8 Shelach, G. (2000). The earliest Neolithic cultures of northeast China, recent discoveries and
9 new perspectives on the beginning of agriculture. *J World Prehist.* 14, 363–413.
- 10 Shi, S. (2005). *The Tibetan-Yi Corridor, History and Culture*. Sichuan People's Publishing
11 House.
- 12 Simonson, T.S., Yang, Y., Huff, C.D., Yun, H., Qin, G., Witherspoon, D.J., Bai, Z., Lorenzo,
13 F.R., Xing, J., Jorde, L.B., Prchal, J.T., Ge, R. (2010). Genetic evidence for high-
14 altitude adaptation in Tibet. *Science*. 329,72-75.
- 15 Su, B., Xiao, C., Deka, R., Seielstad, M.T., Kangwanpong, D., Xiao, J., Lu, D., Underhill, P.,
16 Cavalli-Sforza, L., Chakraborty, R., Jin, L. (2000). Y chromosome haplotypes reveal
17 prehistorical migrations to the Himalayas. *Hum Genet.* 107, 582-590.
- 18 van Oven, M., Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human
19 mitochondrial DNA variation. *Hum Mutat.* 30, E386-E394.
- 20
21 Wang, C.C., Wang, L.X., Shrestha, R., Zhang, M., Huang, X.Y., Hu, K., Jin, L., Li, H. (2014).
22 Genetic structure of Qiangic populations residing in the western Sichuan corridor.
23 *PLoS One*. 9,e103772.
- 24
25 Wang, B., Zhang, Y.B., Zhang, F., Lin, H., Wang, X., Wan, N., Ye, Z., Weng, H., Zhang, L.,
26 Li, X., Yan, J., Wang, P., Wu, T., Cheng, L., Wang, J., Wang, D.M., Ma, X., Yu, J.
27 (2011). On the origin of Tibetans and their genetic basis in adapting high-altitude
28 environments. *PLoS One*. 6, e17002.
- 29
30 Wang, C.C., Yan, S., Qin, Z.D., Lu, Y., Ding, Q.L., Wei, L.H., Li, S.L., Yang, Y.J., Jin, L., Li,
31 H. (2013). Late Neolithic expansion of ancient Chinese revealed by Y chromosome
32 haplogroup O3a1c-002611. *J Syst Evol.* 51, 280–286.
- 33
34 Wang, W.S.Y. (1998). *In the Bronze Age and Early Iron Age peoples of Eastern Central Asia*.
University of Pennsylvania Museum Publications, 508–534.
- 35
36 Wen, B., Xie, X., Gao, S., Li, H., Shi, H., Song, X., Qian, T., Xiao, C., Jin, J., Su, B., Lu, D.,
37 Chakraborty, R., Jin, L. (2004). Analyses of genetic structure of Tibeto-Burman
38 populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum*
39 *Genet.* 74, 856–865.
- 40
41 Wuren, T., Simonson, T.S., Qin, G., Xing, J., Huff, C.D., Witherspoon, D.J., Jorde, L.B., Ge,
42 R.L. (2014). Shared and Unique Signals of High-Altitude Adaptation in
43 Geographically Distinct Tibetan Populations. *PLoS One*. 9, e88252.
- 44
45 Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., Yang, L., Pan, X., Wang, J., Shen, Y., Wu,
46 B., Wang, H., Jin, L. (2011). A genome-wide search for signals of high-altitude
47 adaptation in Tibetans. *Mol Biol Evol.* 28, 1003-11.
- 48
49 Xu, S., Yin, X., Li, S., Jin, W., Lou, H., Yang, L., Gong, X., Wang, H., Shen, Y., Pan, X., He,
50 Y., Yang, Y., Wang, Y., Fu, W., An, Y., Wang, J., Tan, J., Qian, J., Chen, X., Zhang,
51 X., Sun, Y., Zhang, X., Wu, B., Jin, L. (2009). Genomic dissection of population
52 substructure of Han Chinese and its implication in association studies. *Am J Hum*
53 *Genet.* 85, 762-774.
- 54
55 Yan, S., Wang, C.C., Li, H., Li, S.L., Jin, L. (2011). An updated tree of Y chromosome
56 Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *Eur J*
Hum Genet. 19, 1013–1015.
- 57
58 Yan, S., Wang, C.C., Zheng, H.X., Wang, W., Qin, Z.D., Wei, L.H., Wang, Y., Pan, X.D., Fu,
W.Q., He, Y.G., Xiong, L.J., Jin, W.F., Li, S.L., An, Y., Li, H., Jin, L. (2014). Y

chromosomes of 40% Chinese descend from three Neolithic super-grandfathers.
PLoS One. 9, e105691.

Yang, X., Wan, Z., Perry, L., Lu, H., Wang, Q., Zhao, C., Li, J., Xie, F., Yu, J., Cui, T., Wang, T., Li, M., Ge, Q. (2012). Early millet use in northern China. Proc Natl Acad Sci U S A. 109, 3726–3730.

Yao, H.B., Wang, C.C., Wang, J., Tao, X., Shang, L., Wen, S.Q., Du, Q., Deng, Q., Xu, B., Huang, Y., Wang, H.D., Li, S., Cong, B., Ma, L., Jin, L., Krause, J., Li, H. (2017). Genetic structure of Tibetan populations in Gansu revealed by forensic STR loci. Sci Rep. 7, 41195.

Yao, X., Tang, S., Bian, B., Wu, X., Chen, G. and Wang, C.C. (2017). Improved phylogenetic resolution for Y-chromosome Haplogroup O2a1c-002611. Sci Rep, 7: 1146.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., Zheng, H., Liu, T., He, W., Li, K., Luo, R., Nie, X., Wu, H., Zhao, M., Cao, H., Zou, J., Shan, Y., Li, S., Yang, Q., Asan, Ni, P., Tian, G., Xu, J., Liu, X., Jiang, T., Wu, R., Zhou, G., Tang, M., Qin, J., Wang, T., Feng, S., Li, G., Huasang, Luosang, J., Wang, W., Chen, F., Wang, Y., Zheng, X., Li, Z., Bianba, Z., Yang, G., Wang, X., Tang, S., Gao, G., Chen, Y., Luo, Z., Gusang, L., Cao, Z., Zhang, Q., Ouyang, W., Ren, X., Liang, H., Zheng, H., Huang, Y., Li, J., Bolund, L., Kristiansen, K., Li, Y., Zhang, Y., Zhang, X., Li, R., Li, S., Yang, H., Nielsen, R., Wang, J., Wang, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 329, 75-78.

Zhao, M., Kong, Q.P., Wang, H.W., Peng, M.S., Xie, X.D., Wang, W.Z., Jiayang, Duan, J.G., Cai, M.C., Zhao, S.N., Cidanpingcuo, Tu, Y.Q., Wu, S.F., Yao, Y.G., Bandelt, H.J., Zhang, Y.P. (2009). Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. Proc Natl Acad Sci U S A. 106, 21230–21235.

Table 1. The f_3 -statistics (Z; X, Y) to detect if there is evidence that the population Z is derived from admixture of populations related to population X and population Y. A significantly negative Z-score provides unambiguous evidence of mixture in the population X. The population “Han” we used here are the Han Chinese samples in HGDP.

X	Y	Z	f_3	std.err	Z	SNPs
Ami_SGDP	Tibetan_Lhasa	Han_Gansu	-0.00286	0.000417	-6.867	263098
Han	Sardinian	Han_Gansu	-0.00218	0.000371	-5.871	285195
Han	Tuscan	Han_Gansu	-0.00214	0.000368	-5.813	282956
Han	TSI	Han_Gansu	-0.00203	0.000351	-5.788	285994
Ami_SGDP	Tibetan_Yunnan	Han_Gansu	-0.00237	0.00041	-5.78	263109
Tibetan_Lhasa	Dai	Yi	-0.00213	0.000269	-7.898	276206
Tibetan_Yunnan	Dai	Yi	-0.00152	0.000258	-5.892	276298
Ami_SGDP	Tibetan_Lhasa	Yi	-0.00251	0.000431	-5.813	261503
Ami_SGDP	Tibetan_Yunnan	Yi	-0.0021	0.000414	-5.076	261537
Tibetan_Yunnan	Cambodian	Yi	-0.00129	0.00026	-4.971	278577
Tibetan_Lhasa	Han	Tibetan_Yunnan	-0.00063	0.000124	-5.075	287091
Tibetan_Lhasa	Dai	Tibetan_Yunnan	-0.0008	0.000158	-5.063	285777
Tibetan_Lhasa	CHD	Tibetan_Yunnan	-0.00052	0.000109	-4.786	288061

Tibetan_Lhasa	CHB	Tibetan_Yunnan	-0.00044	0.000105	-4.198	289340
Tibetan_Lhasa	Miao	Tibetan_Yunnan	-0.00058	0.000154	-3.741	285514

Table 2. The f_4 -ratio based estimates in the form of $f_4(\text{Chokhopani, Mbuti; Test, Ami})/f_4(\text{Chokhopani, Mbuti; Tibetan_Lhasa, Ami})$ to estimate Tibetan related ancestry (α) in East Asian populations. Here we only show the estimates with Z-score >3.

Test	sample size	α	std.err	Z
Sherpa_SGDP	2	1.011	0.147	6.884
Tibetan_Yunnan	35	0.901	0.053	17.017
Naxi	8	0.801	0.083	9.665
Tibetan_SGDP	2	0.745	0.148	5.031
Yi	10	0.696	0.082	8.441
CHB	84	0.652	0.062	10.501
Tujia	10	0.615	0.087	7.080
Miao	10	0.611	0.089	6.829
Han-NChina	10	0.580	0.087	6.629
Han_Gansu	10	0.570	0.086	6.633
Han	34	0.560	0.075	7.521
CHD	85	0.559	0.070	8.022
She	10	0.550	0.092	5.962
Tibetan_Gansu	10	0.480	0.102	4.692
JPT	86	0.412	0.089	4.642
Japanese	28	0.400	0.095	4.228

Table 3. The f_4 -statistics (Test, Outgroup; X, Y) are to detect if the Test population share more allele with population X or population Y. The population “Han” we used here are the Han Chinese samples in HGDP.

Test	Outgroup	X	Y	f_4	Z	SNPs
French	Mbuti	Tibetan_Gansu	Han	0.000445	3.512	296343
Basque	Mbuti	Tibetan_Gansu	Han	0.000493	3.750	296343
Sardinian	Mbuti	Tibetan_Gansu	Han	0.000403	3.117	296343
Italian	Mbuti	Tibetan_Gansu	Han	0.000410	3.145	296343
Orcadian	Mbuti	Tibetan_Gansu	Han	0.000422	3.219	296343
CEU	Mbuti	Tibetan_Gansu	Han	0.000453	3.589	296343
French	Mbuti	Han_Gansu	Han	0.000498	4.362	296343
Basque	Mbuti	Han_Gansu	Han	0.000527	4.486	296343
Sardinian	Mbuti	Han_Gansu	Han	0.000556	4.862	296343
Italian	Mbuti	Han_Gansu	Han	0.000476	4.034	296343
Orcadian	Mbuti	Han_Gansu	Han	0.000486	4.211	296343
CEU	Mbuti	Han_Gansu	Han	0.000501	4.482	296343

Table 4. Admixture time and lower bound of proportion estimated by ALDER. We only show the results with Z-score >1.

Population	2-ref decay for Tibetan_Lhasa and CHB (generations)	2-ref Z-score	1-ref decay for CHB (generations)	1-ref Z-score	Mixture fraction % lower bound
Tibetan_Gansu	48.03 ± 27.52	1.16	-	-	-
Tibetan_Yunnan	8.49 ± 2.63	3.23	9.55 ± 2.14	4.47	18.1 ± 2.3
Yi	73.42 ± 31.29	1.86	-	-	-

Population	2-ref decay for Tibetan_Lhasa and CHD (generations)	2-ref Z-score	1-ref decay for CHD (generations)	1-ref Z-score	Mixture fraction % lower bound
Tibetan_Gansu	24.19 ± 14.90	1.53	-	-	-
Han_Gansu	28.51 ± 27.53	1.04	-	-	-
Tibetan_Yunnan	5.12 ± 2.38	2.15	6.82 ± 1.59	4.28	11.5 ± 1.8
Naxi	6.59 ± 3.76	1.75	-	-	-
Yi	52.80 ± 20.28	2.60	-	-	-

Population	2-ref decay for CEU and CHB (generations)	2-ref Z-score	1-ref decay for CEU (generations)	1-ref Z-score	Mixture fraction % lower bound
Tibetan_Gansu	23.61 ± 3.31	7.12	20.18 ± 3.79	5.32	3.3 ± 0.3
Han_Gansu	30.87 ± 12.46	2.48	30.34 ± 12.99	2.34	2.4 ± 0.5

Table 5. Y chromosomal and mtDNA haplogroup assignments

Sample	Sex	Population	Y chromosome	mtDNA
Gansu1	male	Tibetan	C2b1b2a-Z32964,B92,Z32965	A11a
Gansu2	male	Tibetan	E1a2b1a2-Z5987	D4
Gansu3	male	Tibetan	C2b1b2a-Z32964,B92,Z32965	D4
Gansu4	male	Tibetan	N1b-L732	F2a
Gansu5	female	Tibetan	-	D4e1a
Gansu6	male	Tibetan	O2a1c2-SK1673,Page74.2	B5a2a1
Gansu7	male	Tibetan	O2a2b1a1a5-CTS10738,M1543,CTS7316,M1726,CTS1017,M1694	M9a1b1
Gansu8	male	Tibetan	O2a1a1c-Page130	M9a1b1
Gansu9	male	Tibetan	Q1a1a1-M120,N14	C1a
Gansu10	male	Tibetan	N1c2b2-L665	D4a
Gansu11	male	Han Chinese	O2b1a-F3338,F2247,F2244,F1770,F837	D4b1b
Gansu12	male	Han Chinese	O2a1c1b1a-F134,F322,F271	D4b2b2b
Gansu13	male	Han Chinese	O2a2a1a2a2a1-F2515,F3469,F2208,F1262	D5c

Gansu14	male	Han Chinese	O2a2a2a-F1226	F1b1
Gansu15	male	Han Chinese	O2a2b1a2b2a-F2326,F2018,F728,F1060	B4c2
Gansu16	male	Han Chinese	O2a1c1a1a1a1a-F856	M8a2
Gansu17	male	Han Chinese	E1a2b1a2-Z5987	M7c1a1
Gansu18	male	Han Chinese	J2a1h-S286,L207.1	F3a1
Gansu19	male	Han Chinese	O2a1c1b1-F238	N9a2
Gansu20	male	Han Chinese	C2e1a1a-M407	A14

Figure Legend

Figure 1: Geographic locations of Han Chinese and Tibetan in Gansu and other referenced East Asian populations in this study.

Figure 2: Principal Component Analysis (PCA) of Tibetan and Han Chinese samples in Gansu Province with other East Asian populations. CHB: Han Chinese in Beijing, China; CHD: Chinese in metropolitan Denver, CO, United States; JPT: Japanese in Tokyo, Japan; Han-NChina: Han Chinese in northern China.

Figure 3: Shared genetic drift among populations, measured by Outgroup f_3 statistics (Mbuti; X, Y). Lighter colours indicate more shared drift.

Figure 4: The Tibeto-Burman Cline (a) and West Eurasian Affinity Cline (b) inferred Outgroup f_3 statistics

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



Figure 1: Geographic locations of Han Chinese and Tibetan in Gansu and other referenced EastAsian populations in this study.
259x197mm (96 x 96 DPI)

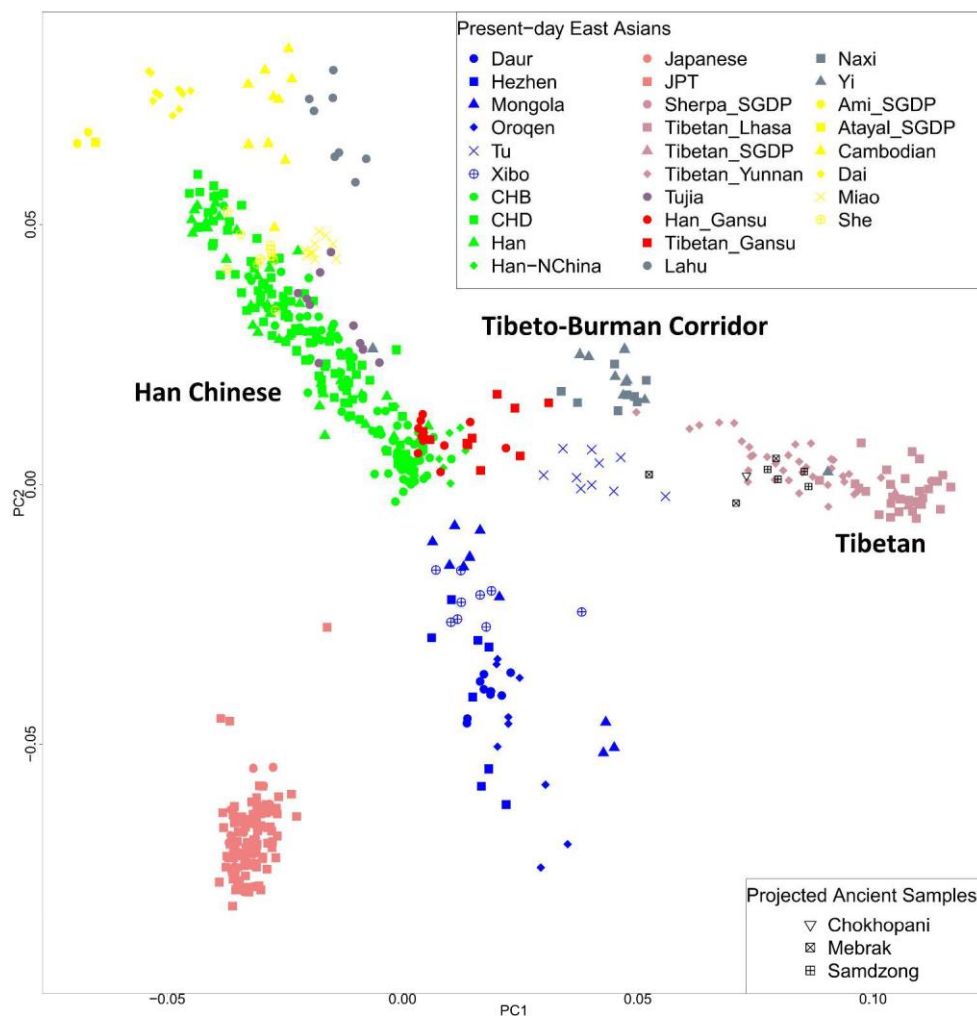


Figure 2: Principal Component Analysis (PCA) of Tibetan and Han Chinese samples in Gansu Province with other East Asian populations. CHB: Han Chinese in Beijing, China; CHD: Chinese in metropolitan Denver, CO, United States; JPT: Japanese in Tokyo, Japan; Han-NChina: Han Chinese in northern China.

423x440mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

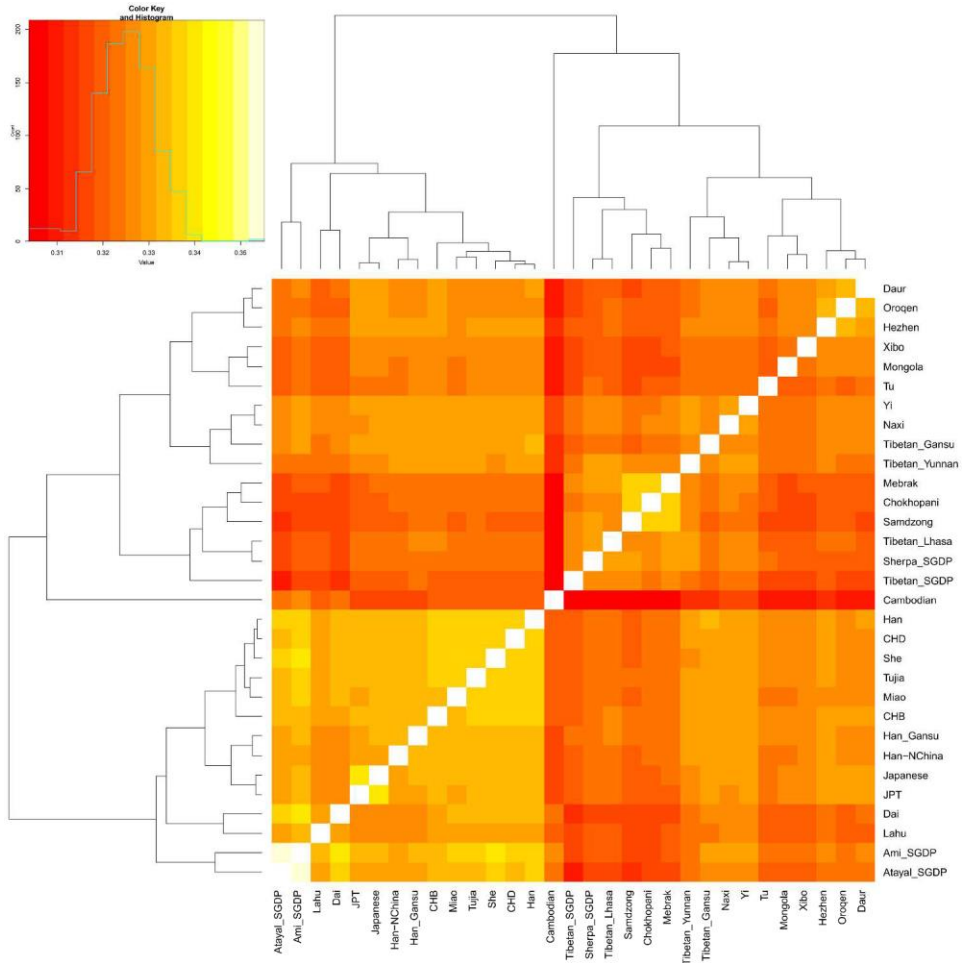


Figure 3: Shared genetic drift among populations, measured by Outgroup f3 statistics (Mbuti; X, Y). Lighter colors indicate more shared drift.

487x487mm (300 x 300 DPI)

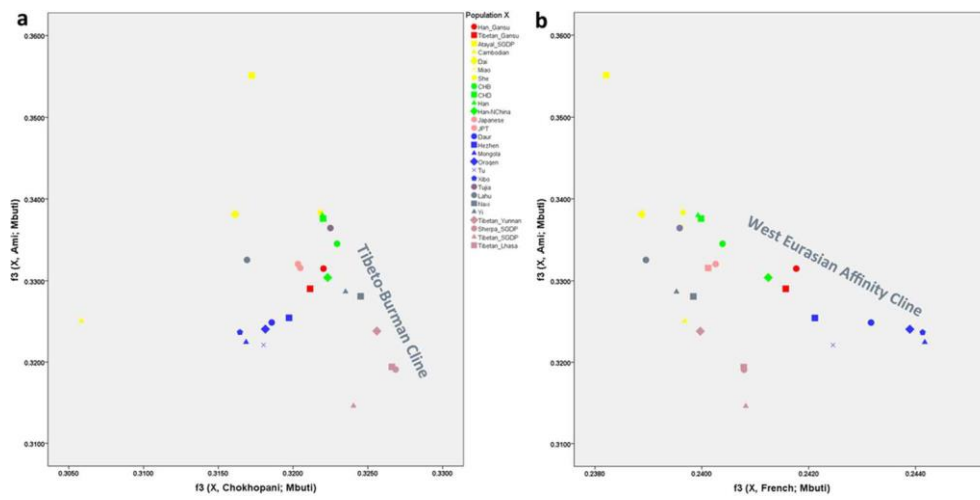


Figure 4: The Tibeto-Burman Cline (a) and West Eurasian Affinity Cline (b) inferred Outgroup f_3 statistics
100x48mm (300 x 300 DPI)