Received: 8 May 2017 | Revised: 14 July 2017 | Accepted: 23 July 2017

DOI: 10.1002/ajpa.23291

RESEARCH ARTICLE

WILEY American Journal of PHYSICAL ANTHROPOLOGY

The genetic admixture in Tibetan-Yi Corridor

Hong-Bing Yao[1] | Senwei Tang[2] | Xiaotian Yao[2] | Hui–Yuan Yeh[3] | Wanhu Zhang[4] | Zhiyan Xie[4] | Qiajun Du[5] | Liying Ma[1] | Shuoyun Wei[1] | Xue Gong[1] | Zilong Zhang[1] | Quanfang Li[1] | Bingying Xu[6] | Hu-Qin Zhang[7] | Gang Chen[2] | Chuan-Chao Wang[8,9,10]

[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, 639798, Singapore

[4]People's Hospital of Gaotai, Gaotai, Gansu Province 734300, China

[5]Lanzhou University Second Hospital Clinical Laboratory, Lanzhou, Gansu Province 730000, 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, Jena, D-07745, Germany

[10]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115

Correspondence
Chuan-Chao Wang, Kahlaische Strasse 10, 07745 Jena, Germany.
Email: wang@shh.mpg.de

Funding information
Natural Science Foundation of Gansu Province, Grant/Award Number: 1308RJZA190; Scientific Research Project for Colleges of Gansu province, Grant/Award Number: 2014A-085; Scientific Research Project for Colleges of Gansu province, Grant/Award Number: 2015A-105; Natural Science Foundation for Young Scientists of China, Grant/Award Number: 51501042; Lanzhou Research Program of Science and Technology, Grant/Award Number: 2016-3-122; Natural Science Foundation, Grant/Award Number: H2302, C.C.W is supported by Nanqiang Outstanding Young Talents Program of Xiamen University, Max Planck Society and Harvard Medical School, European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, Grant/Award Number: 646612 granted to Martine Robbeets.

Abstract

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

KEYWORDS
gene flow, Han Chinese, Tibetan

1 | 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 (Kang et al., 2012; Qi et al., 2013; Su et al., 2000; 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, Wang, Li, Li, & Jin, 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 avaiable in recent years, but mainly focus on their genetic basis in adapting high-altitude environments (Beall et al., 2010; Jeong et al., 2014; Petousi et al., 2014; Simonson et al., 2010; Wang et al., 2011; Wuren et al., 2014; Xu et al., 2011; Yi et al., 2010). 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) suggeted 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 perspective that most of the Tibetan gene pool diverged from that of Han Chinese about 15 kya to 9 kya and the shared ancestry of Tibetan-enriched sequences dates back to 62–38 kya, which is consistent with Y chromosome evidence of two-phase for the origin of Tibeto-Burman populations. Jeong et al (Jeong et al., 2016) reported ancient genomes from the Chokhopani, Mebrak, and Samdzong sites spanning 3 to 1 kya in Nepal suggesting the Tibetan Plateau experienced millennia of genetic continuity which continues to the present day.

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

2 | MATERIALS AND METHODS

2.1 | Sampling and genotyping

We collected blood samples from 10 unrelated individuals from Tibetan and the other 10 unrelated individuals from Han Chinese in south Gansu province (Figure 1). 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. Informed consent was obtained from all individual participants included in the study. DNA isolation and purification were following the standard lysis protocol, and DNA was purified using the QIAamp DNA mini kit (QIAGEN, Hilden, Germany). Genotyping was performed on the Affymetrix WeGene V1 Arrays covering 596,744 SNPs at the Shanghai Jiaotong University, Shanghai. The WeGene V1 arrays were designed to identify all known paternal Y-chromosome and maternal mtDNA lineages with 18963 Y-chromosome and 4448 mtDNA phylogenetic relevant SNPs. We genotyped our samples using WeGene arrays because we want to generate informative Y chromosome and mtDNA results. The dataset generated during the current study is available upon request to the corresponding author when the article is published.

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

2.2 | Data merging

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

2.3 | Principal component analysis

We used smartpca (version: 13050), part of the EIGENSOFT package (Patterson, Price, & Reich, 2006), to carry out Principal Component Analysis (PCA). We performed PCA on present-day populations and then projected the ancient samples using the lsqproject: YES option, which accounts for samples with substantial missing data. We did not perform any outlier removal iterations (numoutlieriter: 0). We set all other options to the default. We assessed statistical significance with a Tracy-Widom test using the twstats program of EIGENSOFT. All the first six principal components that we discuss and plot in what follows were highly statistically significant ($p < 10^{-12}$).

2.4 | f_3-statistics

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

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

2.5 | f_4-statistics

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

2.6 | f_4-*ratio* estimations

We used the qpF4ratio program of ADMIXTOOLS (Patterson et al., 2012; Reich et al., 2009) with default parameters to estimate the

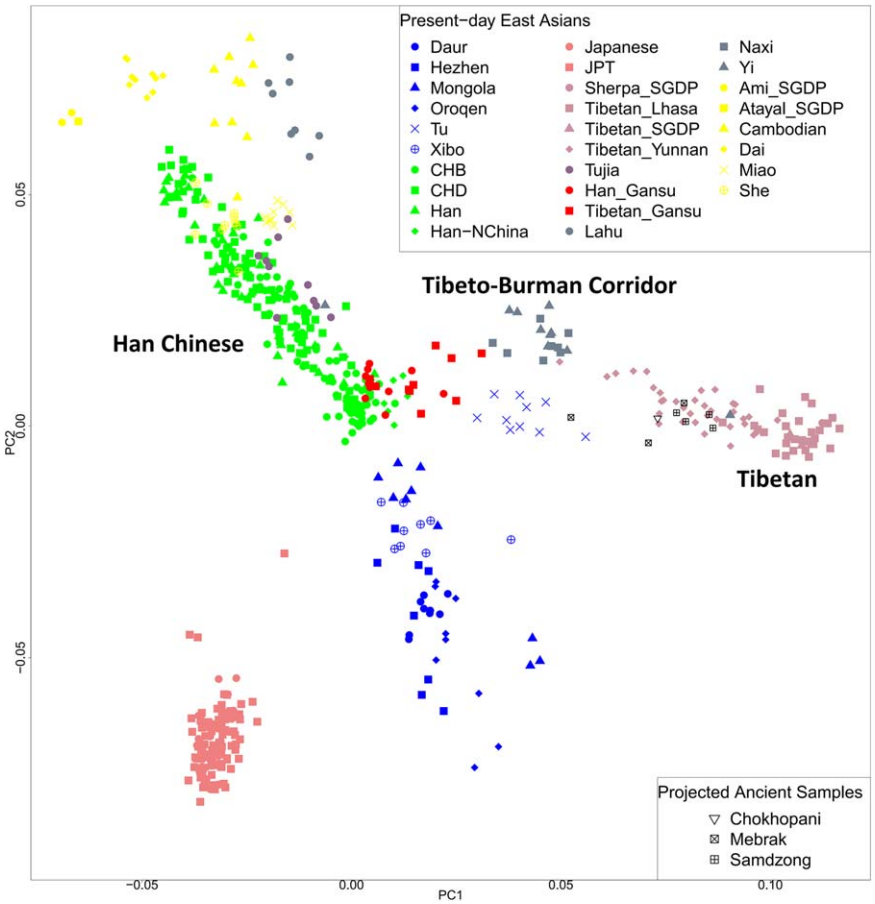

FIGURE 2 Principal Component Analysis (PCA) of Tibetan and Han Chinese samples in Gansu Province with other East Asian populations. CHB: Han Chinese in Beijing, China; CHD: Chinese in metropolitan Denver, CO, United States; JPT: Japanese in Tokyo, Japan; Han—NChina: Han Chinese in northern China

admixture proportions of tested populations with the proposed sources.

2.7 | ADMIXTURE analysis

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

2.8 | Weighted linkage disequilibrium (LD) analysis

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

2.9 | Y chromosomal and mtDNA haplogroup assignment

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

3 | RESULTS

We firstly performed PCA to provide a broad overview of population structure across East Asia (Figure 2). We show five broad clusters correlating well with geographic and linguistic categories within East Asia: a southern cluster with Austronesian, Tai-Kadai, and Austroasiatic speaking groups; a Han Chinese cluster; a Tibetan cluster; a Japanese

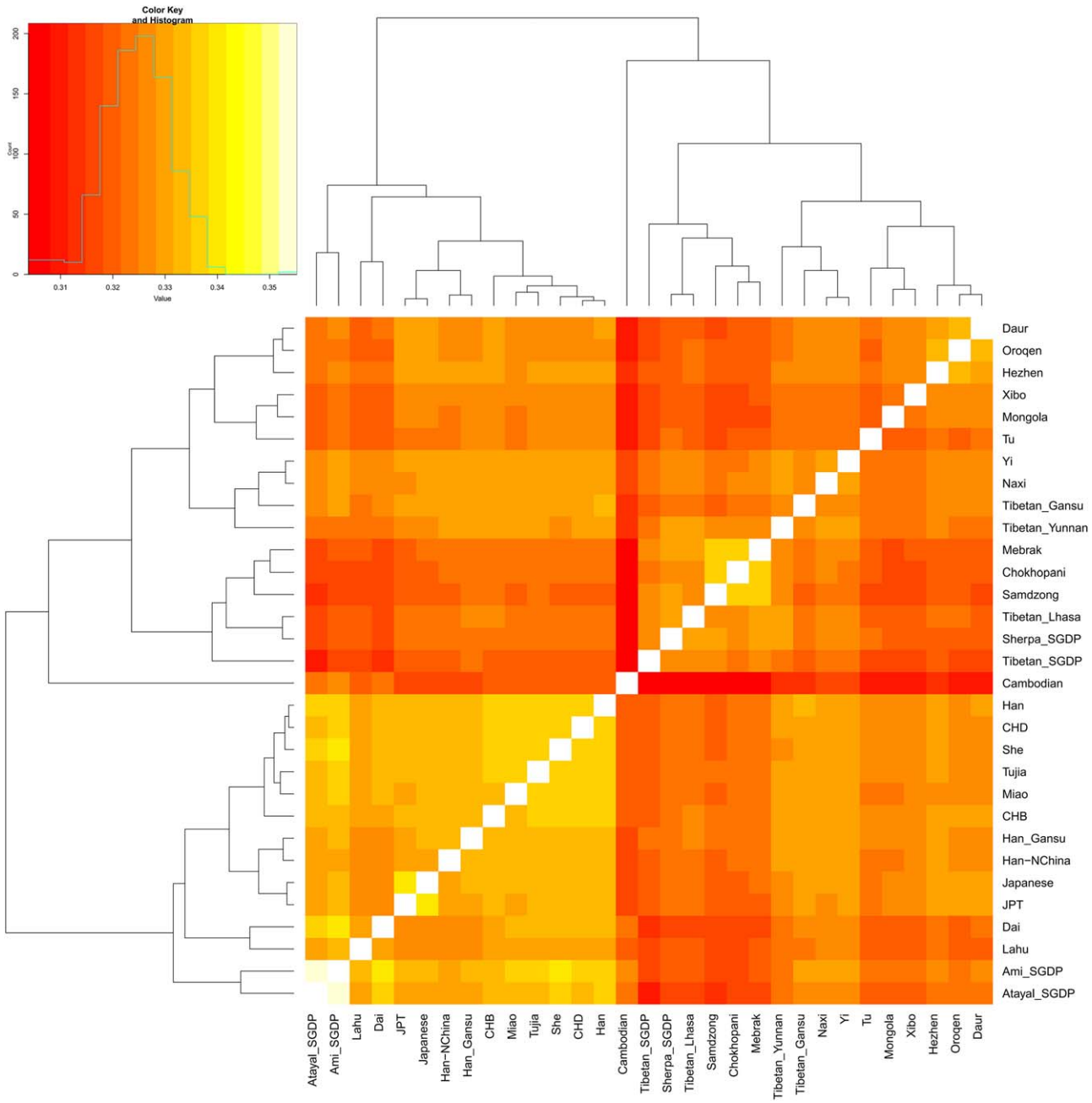

FIGURE 3 Shared genetic drift among populations, measured by Outgroup *f3* statistics (Mbuti; X, Y). Lighter colours indicate more shared drift

cluster; and an Altaic cluster consisting of Turkic, Tungusic and Mongolic-speaking groups in north China. Our Tibetan and Han Chinese samples from Gansu province together with other Tibeto-Burman speaking populations in the Corridor (Tibetan_Yunnan, Naxi, and Yi) appear as potentially admixed populations on the PCA occupying an intermediate position between Tibetan and Han Chinese.

In the model-based ADMIXTURE clustering analysis, we used cross-validation to identify an "optimal" number of clusters. We observed the lowest CV errors for $K = 12$. At $K = 12$, we observed three ancestral components specific to individuals in East Asia. One of these components is enriched in the ancient Nepalese and found to be at highest proportions in Tibetans. The second is enriched in Taiwan Austronesians but is also prevalent in Han Chinese. The third component is enriched in Yakut, a Turkic-speaking population in Siberia. We found our Tibetan and Han Chinese samples of Gansu province are genetically similar with Han Chinese in northern China and other Corridor populations (Supporting Information Figure S1).

We next calculated an outgroup *f3*-statistics of the form *f3* (Mbuti; X, Y) to quantify population differentiation across East Asia observed by PCA (Supporting Information Table S1 and Figure 3). The Han_Gansu cluster with other lowland populations, but Tibetan_Gansu group tightly together with other Tibeto-Burman speaking populations

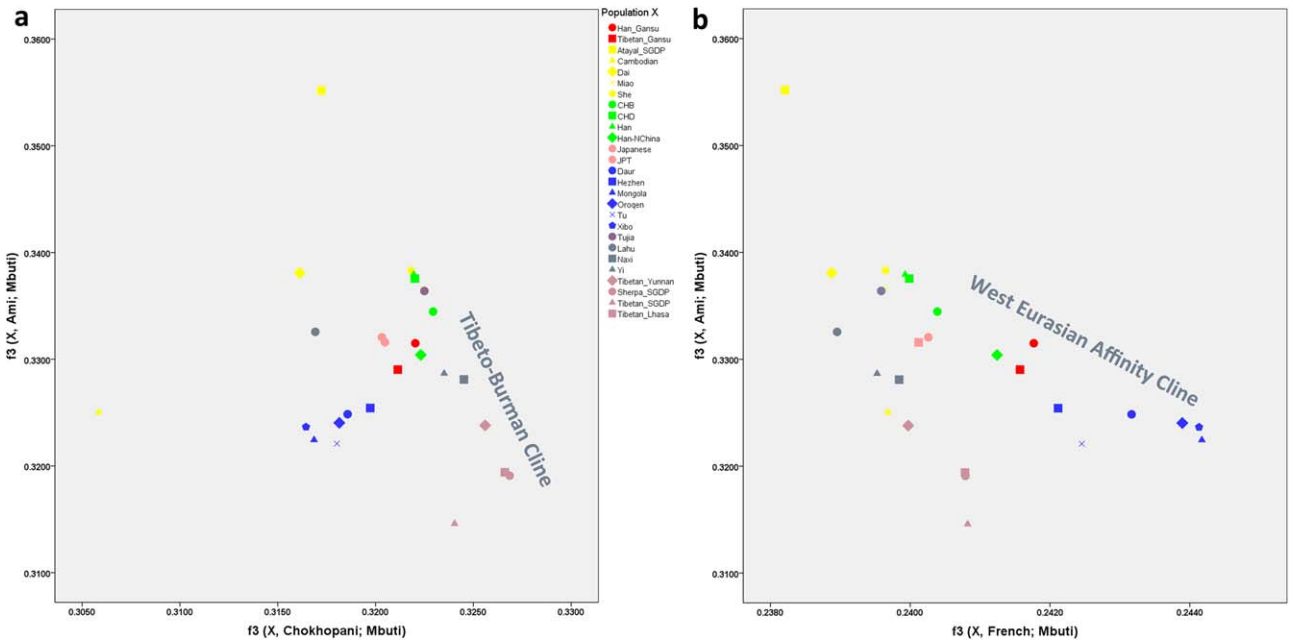

FIGURE 4 The Tibeto-Burman Cline (a) and West Eurasian Affinity Cline (b) inferred by Outgroup *f3* statistics

in the Corridor. The *f3*-statistics correlate well with the patterns observed via PCA that the populations in the Tibetan-Yi Corridor share an affinity with both Tibetans on the Plateau and Han Chinese and other East Asians in the lowland. We also plotted the outgroup *f3*-statistics in the form of *f3*(X, Chokhopani; Mbuti)/*f3*(X, Ami; Mbuti) to visualize the allele sharing of various East Asian populations with Ami and Chokhopani (an ancient sample tracing back to 3 kya in Nepal showing genetic continuity with present-day Tibetans) in Figure 4a. We observed a clear cline of differences in Tibetan related ancestry with Sherpa and Tibetan_Lhasa sharing the most genetic drift with

TABLE 1 The *f3*-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

X	Y	Z	*f3*	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

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.

TABLE 2 The f_4-ratio based estimates in the form of *f4*(Chokhopani, Mbuti; Test, Ami)/*f4* (Chokhopani, Mbuti; Tibetan_Lhasa, Ami) to estimate Tibetan related ancestry (α) in East Asian populations

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

Here, we only show the estimates with Z-score >3.

ancient Chokhopani and Cambodian sharing the least. Consistent with the PCA plot, the populations in the Tibetan-Yi Corridor are in the middle of this Tibetan ancestry cline.

We applied a formal admixture test using *f3*-statistics in the form of f_3 (Z; X, Y) where Z is our tested group and X and Y are worldwide populations that might be the genetic sources for modeling the admixture in population Z. We observed significant signals of admixture (Z < −5) in Han_Gansu, Yi, and Tibetan_Yunnan for Tibetan related and lowland East Asian related ancestry (Table 1).

We proceed to use *f4*-ratio based estimates in the form of *f4* (Chokhopani, Mbuti; Test, Ami)/*f4* (Chokhopani, Mbuti; Tibetan_Lhasa, Ami) to quantify the proportions of Tibetan related ancestry East Asian groups. We assigned the Tibetan_Lhasa and Ami as ancestral source populations based on the *f3*-statistics.We observed a consistent genetic cline with Figure 3a in term of Tibetan related ancestry (Table 2). The Tibetan_Yunnan, Naxi, Yi, and Tibetan_Gansu derive 90.1%, 80.1%, 69.6%, and 48.0% Tibetan related ancestry, respectively.

We also detected the evidence of West Eurasian admixture into our Han Chinese samples in Gansu as shown in the significant negative Z-score of *f3*-statistics when assuming West Eurasian populations as sources (Table 1 and Supporting Information Table S2). The outgroup *f3*-statistics we calculated in Figure 4b show Han Chinese and Tibetan populations in Gansu share more genetic drift with French compared with other Han Chinese and Tibetans. We confirmed the West Eurasian admixture using *f4*-statistics in the form of (West Eurasians, Mbuti; Han_Gansu/Tibetan_Gansu, Han) in which "Han" is the Han Chinese of HGDP in Table 3, where the significant positive statistics suggest West Eurasians share more allele with Han_Gansu and Tibetan_Gansu compared with Han Chinese of HGDP. We can get significant positive values when putting other Corridor populations (Tibetan_Yunnan, Naxi, and Yi) in place of "Han" in the above *f4*-statistics, which suggest there is substructure in those Corridor populations with the northern ones specially sharing an affinity with West Eurasians.

We estimated the admixture time and lower bounds on the admixture proportion using the linkage disequilibrium (LD)-based admixture inference implemented in ALDER (Loh et al., 2013) and showed the results in Table 4. For Tibetan highlander and lowland East Asian admixture in Corridor populations, we weighted LD curves with Corridor groups as test populations and Tibetan_Lhasa and CHB or CHD as possible source populations. We also took advantage of the one-

TABLE 3 The *f4*-statistics (Test, Outgroup; X, Y) are to detect if the Test population share more allele with population X or population Y

Test	Outgroup	X	Y	*f4*	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

The population "Han" we used here are the Han Chinese samples in HGDP.

TABLE 4 Admixture time and lower bound of proportion estimated by ALDER

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

We only show the results with Z-score >1.

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

reference inference capabilities of ALDER to only use CHB or CHD as the source to estimate the lower bounds of the admixture proportions. The average admixture times for Corridor populations range from 5 to 70 generations (about 150 to 2100 years ago assuming 30 years a generation). The Tibetans in Yunnan are suggested to derive 11%-18% ancestry from Han Chinese. For the West Eurasian admixture, we computed weighted LD curves with Han_Gansu and Tibetan_Gansu as the test populations and CEU and CHB as the sources. The average admixture times for Han Chinese and Tibetan in Gansu range from about 20 to 30 generations ago (about 600 to 900 years assuming 30 years a generation), suggesting relatively recent gene flow from West Eurasia to northwest China. We also estimated mixture fractions of at least 2.4% to 3.3% CEU-related ancestry for those populations. Changing the starting point of the LD fit does not qualitatively affect the results (Supporting Information Document S1). We caution that the date estimates might not reflect the initial admixture in present-day populations; instead, it is an average date of population mixture. If the admixture did not happen immediately when two populations met or occurred many times over an extended period, the true start of mixture would be more ancient.

The paternal Y chromosome also gave evidence for the West Eurasian admixture, as we identified West Eurasian characteristic lineages E and J in our Han Chinese and Tibetan samples (Table 5). The dominant lineage in Han Chinese of Gansu is O2, which is in consistent with the general paternal profile of other Han Chinese groups (Yan et al., 2011, 2014). However, the paternal history of Tibetans_Gansu is more complicated with haplogroup C, N, and Q. The maternal mtDNA lineages of our samples are consistent with the general profile of this region with high frequency of D4 (Qi et al., 2013; Qin et al., 2010; Wang et al., 2014).

4 | DISCUSSION

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

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

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

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

ACKNOWLEDGMENTS

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

ORCID

Hu-Qin Zhang http://orcid.org/0000-0001-6366-9228
Chuan-Chao Wang http://orcid.org/0000-0001-9628-0307

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *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. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 5523–5528.

Beall, C. M., Cavalleri, G. L., Deng, L., Elston, R. C., Gao, Y., Knight, J., … Zheng, Y. T. (2010). Natural selection on EPAS1 (HIF2a) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences of the United States of America*, *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., … Liu, J. (2009). Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *The American Journal of Human Genetics*, *85*, 775–785.

International HapMap Consortium. (2003). The International HapMap Project. *Nature*, *426*, 789–796.

Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D. B., Pritchard, J. K., … Di Rienzo, A. (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. *Nature Communications*, *5*, 3281.

Jeong, C., Ozga, A. T., Witonsky, D. B., Malmström, H., Edlund, H., Hofman, C. 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. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 7485–7490.

Kang, L., Lu, Y., Wang, C., Hu, K., Chen, F., Liu, K., … Li, H. Genographic Consortium. (2012). Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the eastern Himalayas. *Annals of Human Genetics*, *76*, 92–99.

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., … 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., … Xu, B. (2015). Allele frequency of 19 autosomal STR loci in the Bai population from the southwestern region of mainland China. *Electrophoresis*, *36*, 2498–2503.

Loh, P. R., Lipson, M., Patterson, N., Moorjani, P., Pickrell, J. K., Reich, D., & Berger, B. (2013). Inferring admixture histories of human populations using linkage disequilibrium. *Genetics*, *193*, 1233–1254.

Lu, D., Lou, H., Yuan, K., Wang, X., Wang, Y., Zhang, C., … Xu, S. (2016). Ancestral Origins and Genetic History of Tibetan Highlanders. *American Journal of Human Genetics*, *99*, 580–594.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., … Reich, D. (2016). The Simons Genome Diversity Project, 300 genomes from 142 diverse populations. *Nature*, *538*, 201–206.

Martisoff, J. A. (1991). Sino-Tibetan linguistics, present state and future prospects. *Annual Review of Anthropology*, *20*, 469–504.

Nothnagel, M., Fan, G., Guo, F., He, Y., Hou, Y., Hu, S., … Roewer, L. (2017). Revisiting the male genetic landscape of China, a multi-center study of almost 38,000 Y-STR haplotypes. *Human Genetics*, *136*, 485–497.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., … Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*, 1065–1093.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, *2*, e190.

Petousi, N., Croft, Q. P., Cavalleri, G. L., Cheng, H. Y., Formenti, F., Ishida, K., … Robbins, P. A. (2014). Tibetans living at sea level have a hyporesponsive hypoxia-inducible factor system and blunted physiological responses to hypoxia. *Journal of Applied Physiology*, *116*, 893–904.

Qi, X., Cui, C., Peng, Y., Zhang, X., Yang, Z., Zhong, H., … Su, B. (2013). Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the Tibetan plateau. *Molecular Biology and Evolution*, *30*, 1761–1778.

Qin, Z., Yang, Y., Kang, L., Yan, S., Cho, K., Cai, X., … Genographic, C. (2010). A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *American Journal of Physical Anthropology*, *143*, 555–569.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*, 489–494.

Shelach, G. (2000). The earliest Neolithic cultures of northeast China, recent discoveries and new perspectives on the beginning of agriculture. *Journal of World Prehistory*, *14*, 363–413.

Shi, S. (2005). *The tibetan-yi corridor, history and culture*. Chengdu: Sichuan People's Publishing House.

Simonson, T. S., Yang, Y., Huff, C. D., Yun, H., Qin, G., Witherspoon, D. J., … Ge, R. (2010). Genetic evidence for high-altitude adaptation in Tibet. *Science*, *329*, 72–75.

Su, B., Xiao, C., Deka, R., Seielstad, M. T., Kangwanpong, D., Xiao, J., … Jin, L. (2000). Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Human Genetics*, *107*, 582–590.

van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, *30*, E386–E394.

Wang, C. C., Wang, L. X., Shrestha, R., Zhang, M., Huang, X. Y., Hu, K., … Li, H. (2014). Genetic structure of Qiangic populations residing in the western Sichuan corridor. *PLoS One*, *9*, e103772.

Wang, B., Zhang, Y. B., Zhang, F., Lin, H., Wang, X., Wan, N., … Yu, J. (2011). On the origin of Tibetans and their genetic basis in adapting high-altitude environments. *PLoS One*, *6*, e17002.

Wang, C. C., Yan, S., Qin, Z. D., Lu, Y., Ding, Q. L., Wei, L. H., … Li, H. (2013). Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611. *Journal of Systematics and Evolution*, *51*, 280–286.

Wang, W. S. Y. (1998). *In the Bronze Age and early Iron Age peoples of Eastern Central Asia* (pp.508–534). Philadelphia: University of Pennsylvania Museum Publications.

Wen, B., Xie, X., Gao, S., Li, H., Shi, H., Song, X., … Jin, L. (2004). Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *American Journal of Human Genetics*, *74*, 856–865.

Wuren, T., Simonson, T. S., Qin, G., Xing, J., Huff, C. D., Witherspoon, D. J., … Ge, R. L. (2014). Shared and unique signals of high-altitude adaptation in geographically distinct Tibetan populations. *PLoS One*, *9*, e88252.

Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., … Jin, L. (2011). A genome-wide search for signals of high-altitude adaptation in Tibetans. *Molecular Biology and Evolution*, *28*, 1003–1011.

Xu, S., Yin, X., Li, S., Jin, W., Lou, H., Yang, L., … Jin, L. (2009). Genomic dissection of population substructure of Han Chinese and its implication in association studies. *American Journal of Human Genetics*, *85*, 762–774.

Yan, S., Wang, C. C., Li, H., Li, S. L., & Jin, L. (2011). An updated tree of Y chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *European Journal of Human Genetics*, *19*, 1013–1015.

Yan, S., Wang, C. C., Zheng, H. X., Wang, W., Qin, Z. D., Wei, L. 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., ... Ge, Q. (2012). Early millet use in northern China. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 3726–3730.

Yao, H. B., Wang, C. C., Wang, J., Tao, X., Shang, L., Wen, S. Q., ... Li, H. (2017). Genetic structure of Tibetan populations in Gansu revealed by forensic STR loci. *Scientific Reports*, *7*, 41195.

Yao, X., Tang, S., Bian, B., Wu, X., Chen, G., & Wang, C. C. (2017). Improved phylogenetic resolution for Y-chromosome Haplogroup O2a1c-002611. *Scientific Reports*, *7*, 1146.

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X., Pool, J. E., ... 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., ... Zhang, Y. P. (2009). Mitochondrial genome evidence reveals successful Late Paleolithic settlement on the Tibetan Plateau. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 21230–21235.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Yao H-B, Tang S, Yao X, et al. The genetic admixture in Tibetan-Yi Corridor. *Am J Phys Anthropol*. 2017;164:522-532. https://doi.org/10.1002/ajpa.23291